

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

**Modelos de transição para dados binários**

**Idemauro Antonio Rodrigues de Lara**

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

**Piracicaba**

**2007**

Idemauro Antonio Rodrigues de Lara  
Licenciado em Matemática

## Modelos de transição para dados binários

Orientadora:

Prof<sup>a</sup> Dr<sup>a</sup> **CLARICE G. B. DEMÉTRIO**

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**

Lara, Idemauro Antonio Rodrigues de  
Modelos de transição para dados binários / Idemauro Antonio Rodrigues de Lara.  
- - Piracicaba, 2007.  
128 p. : il.

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

1. Análise de dados longitudinais 2. Modelos lineares generalizados 3.  
Processos estocásticos 4. Verossimilhança I. Título

CDD 519.5

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

## Dedicatória

Dizem que **sempre** e **nunca** são tempos que podem não existir. Na verdade são expressões associadas ao tempo. Acho que posso valer da minha experiência para tentar dar significado a esses termos. Na vida há perdas, dificuldades e muitas dores ... mas há também muitas dádivas. Na minha história essa dicotomia existe ... por isso, há **sempre** tempo para agradecer por algo, para aprender, para recomeçar ... e **nunca** é tempo de desistir.

A DEUS,

por me guiar num caminho de luz e não me deixar esmorecer, mesmo diante de todas as adversidades.

Aos meus pais (*in memoriam*) e irmãos Cássia, Joel e Marcos.

## AGRADECIMENTOS

À professora Clarice Garcia Borges Demétrio, a orientação, a confiança em mim depositada e apoio.

Ao professor Dalton Francisco de Andrade, da Universidade Federal de Santa Catarina, a sugestão do tema e contribuições.

Aos professores João Maurício Araújo Mota e Juvêncio Santos Nobre, da Universidade Federal do Ceará, a constante ajuda na teoria dos processos estocásticos.

Aos professores do Departamento de Ciências Exatas da ESALQ/USP, os ensinamentos e a dedicação.

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

À UNIMEP, o auxílio na forma de redução de aulas na primeira etapa do curso.

À Maria Clara, o suporte na língua inglesa.

À Ana Maria Araújo, a amizade ímpar e o apoio incondicional.

Ao colega de doutorado Osmar, a ajuda com as configurações finais.

Aos demais colegas dos cursos de mestrado e doutorado do Departamento de Ciências Exatas da ESALQ/USP, a convivência.

A minha família por me apoiar e compreender os momentos de ausência.

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 . . . . .	10
1 INTRODUÇÃO . . . . .	12
2 DESENVOLVIMENTO . . . . .	17
2.1 Revisão de Literatura . . . . .	17
2.1.1 Modelos lineares generalizados . . . . .	18
2.1.1.1 Estimação dos parâmetros do MLG . . . . .	21
2.1.1.2 Função desvio, estatística generalizada de Pearson e seleção de modelos . . . . .	22
2.1.1.3 Quase-verossimilhança . . . . .	24
2.1.1.4 Análise de resíduos e diagnósticos . . . . .	26
2.1.2 Tópicos de processos estocásticos . . . . .	30
2.1.2.1 Definição de processo estocástico . . . . .	30
2.1.2.2 Cadeia de Markov . . . . .	31
2.1.2.3 Distribuição estacionária ou de equilíbrio . . . . .	33
2.1.2.4 Inferências relacionadas à cadeia de Markov . . . . .	34
2.1.2.4.1 Estimação das probabilidades . . . . .	35
2.1.2.4.2 Teoria assintótica . . . . .	36
2.1.2.4.3 Teste para a hipótese de homogeneidade . . . . .	38
2.1.2.4.4 Teste para a hipótese de estacionariedade . . . . .	39
2.1.2.4.5 Teste para a hipótese de uma matriz de transição específica . . . . .	40
2.1.2.4.6 Cadeias de Markov de alcance maior do que 1 . . . . .	41
2.1.3 Modelos para análise de dados longitudinais com resposta diferente da normal . . . . .	44
2.1.3.1 Modelos lineares generalizados marginais para dados binários . . . . .	44
2.1.3.2 Equações de estimação generalizadas e modelos marginais . . . . .	46
2.1.3.3 Modelos de transição de Markov . . . . .	53
2.1.3.4 Tabelas de contingência e matrizes de probabilidades de transição . . . . .	55
2.1.3.5 Modelos lineares generalizados de transição . . . . .	59

	6
2.1.3.6 Estimação por máxima verossimilhança para processos estacionários . . . . .	64
2.1.3.7 Modelos com efeitos aleatórios . . . . .	67
2.1.3.8 Estimação dos parâmetros de um modelo misto . . . . .	69
2.2 Metodologia . . . . .	74
2.2.1 Material . . . . .	74
2.2.2 Métodos . . . . .	75
2.2.2.1 Matriz das probabilidades de transição da cadeia de Markov para dados binários	76
2.2.2.2 Testes para comparação de tratamentos . . . . .	79
2.2.2.2.1 Teste da razão de chances . . . . .	82
2.2.2.2.2 Teste da diferença das probabilidades de transição . . . . .	85
2.2.2.3 Modelo de transição de Markov misto . . . . .	86
2.3 Resultados . . . . .	89
2.3.1 Modelo de transição de primeira ordem . . . . .	89
2.3.2 Modelo de transição de segunda ordem . . . . .	100
3 CONSIDERAÇÕES FINAIS . . . . .	105
REFERÊNCIAS . . . . .	107
APÊNDICES . . . . .	113

## RESUMO

### Modelos de transição para dados binários

Dados binários ou dicotômicos são comuns em muitas áreas das ciências, nas quais, muitas vezes, há interesse em registrar a ocorrência, ou não, de um evento particular. Por outro lado, quando cada unidade amostral é avaliada em mais de uma ocasião no tempo, tem-se dados longitudinais ou medidas repetidas no tempo. É comum também, nesses estudos, se ter uma ou mais variáveis explicativas associadas às variáveis respostas. As variáveis explicativas podem ser dependentes ou independentes do tempo. Na literatura, há técnicas disponíveis para a modelagem e análise desses dados, sendo os modelos disponíveis extensões dos modelos lineares generalizados. O enfoque do presente trabalho é dado aos modelos lineares generalizados de transição para a análise de dados longitudinais envolvendo uma resposta do tipo binária. Esses modelos são baseados em processos estocásticos e o interesse está em modelar as probabilidades de mudanças ou transições de categorias de respostas dos indivíduos no tempo. A suposição mais utilizada nesses processos é a da propriedade markoviana, a qual condiciona a resposta numa dada ocasião ao estado na ocasião anterior. Assim, são revistos os fundamentos para se especificar tais modelos, distinguindo-se os casos estacionário e não-estacionário. O método da máxima verossimilhança é utilizado para o ajuste dos modelos e estimação das probabilidades. Adicionalmente, apresentam-se testes assintóticos para comparar tratamentos, baseados na razão de chances e na diferença das probabilidades de transição. Outra questão explorada é a combinação do modelo de efeitos aleatórios com a do modelo de transição. Os métodos são ilustrados com um exemplo da área da saúde. Para esses dados, o processo é considerado estacionário de ordem dois e o teste proposto sinaliza diferença estatisticamente significativa a favor do tratamento ativo. Apesar de ser uma abordagem inicial dessa metodologia, verifica-se, que os modelos de transição têm notável aplicabilidade e são fontes para estudos e pesquisas futuras.

Palavras-chave: Dados longitudinais; Modelo linear generalizado; Processos estocásticos; Probabilidades de transição; Máxima verossimilhança



## ABSTRACT

### Transition models for binary data

Binary or dichotomous data are quite common in many fields of Sciences in which there is an interest in registering the occurrence of a particular event. On the other hand, when each sampled unit is evaluated in more than one occasion, we have longitudinal data or repeated measures over time. It is also common, in longitudinal studies, to have explanatory variables associated to response measures, which can be time dependent or independent. In the literature, there are many approaches to modeling and evaluating these data, where the models are extensions of generalized linear models. This work focus on generalized linear transition models suitable for analyzing longitudinal data with binary response. Such models are based on stochastic processes and we aim to model the probabilities of change or transitions of individual response categories in time. The most used assumption in these processes is the Markov property, in which the response in one occasion depends on the immediately preceding response. Thus we review the fundamentals to specify these models, showing the differences between stationary and non-stationary processes. The maximum likelihood approach is used in order to fit the models and estimate the probabilities. Furthermore, we show asymptotic tests to compare treatments based on odds ratio and on the differences of transition probabilities. We also present a combination of random-effects model with transition model. The methods are illustrated with health data. For these data, the process is stationary of order two and the suggested test points to a significant statistical difference in favor of the active treatment. This work is an initial approach to transition models, which have high applicability and are great sources for further studies and researches.

Keywords: Longitudinal data; Generalized linear model; Stochastic processes; Transition probabilities; Maximum likelihood

**LISTA DE FIGURAS**

- Figura 1 - Probabilidades marginais de chuva observadas para cada dia do mês de novembro, em Piracicaba-SP, no período de 1997 a 2006 e as correspondentes probabilidades estimadas pela cadeia de Markov . . . . . 58
- Figura 2 - Gráficos semi-normal de probabilidades com envelopes simulados, para os modelos de transição estacionários de 1<sup>a</sup> ordem, no estudo da condição respiratória . . . . . 99
- Figura 3 - Gráfico dos resíduos de Pearson versus a ordem das observações para o modelo (1) de transição estacionário de 1<sup>a</sup> ordem, no estudo da condição respiratória . . . . . 99
- Figura 4 - Gráfico semi-normal de probabilidades com envelopes simulados para o modelo de transição estacionário de 2<sup>a</sup> ordem, no estudo da condição respiratória . . . . . 103

## LISTA DE TABELAS

Tabela 1 - Tabela de contingência $T \times k$ para a apresentação das probabilidades de transição, para um estado fixo $a$ . . . . .	39
Tabela 2 - Estudo longitudinal com resposta binária supondo amostra homogênea . . .	55
Tabela 3 - Tabela de contingência $2 \times 2$ , na $t$ -ésima transição, de um estudo longitudinal com resposta binária e amostra homogênea . . . . .	56
Tabela 4 - Dias do mês de novembro com precipitação (1) ou ausênica (0) em Piracicaba-SP . . . . .	57
Tabela 5 - Total de frequências de transições de primeira ordem observadas nos dados sobre precipitação, em Piracicaba-SP, para os dias do mês de novembro no período de 1997 a 2006 . . . . .	57
Tabela 6 - Distribuição das probabilidades de transição, considerando os grupos placebo e ativo (tratamento), com estado prévio fixado $Y_{(t-1)} = a$ , num modelo de Markov . . . . .	80
Tabela 7 - Estrutura usual de uma tabela de contingência $2 \times 2$ para o teste de Mantel-Hanszel . . . . .	83
Tabela 8 - Tabela de contingência $2 \times 2$ , considerando os estratos de tratamento e as frequências de transições para os estados 0 e 1, dado $Y_{(t-1)} = a$ , $a \in \{0, 1\}$ .	84
Tabela 9 - Distribuição do número de transições da condição respiratória, no exemplo do estudo epidemiológico. Probabilidades estimadas para o processo estacionário . . . . .	89
Tabela 10 - Número de transições e probabilidades estimadas do estudo sobre a condição respiratória, para as quatro transições . . . . .	90
Tabela 11 - Distribuição das transições de acordo com o tipo de tratamento no estudo epidemiológico sobre a condição respiratória . . . . .	91
Tabela 12 - Número de transições e estimativas de probabilidades, de acordo com o tipo de tratamento e condição respiratória na visita anterior . . . . .	92
Tabela 13 - Total de transições e estimativas de probabilidades, de acordo com o tipo de tratamento e condição respiratória na visita anterior . . . . .	93

Tabela 14 -Estimativas dos parâmetros para os modelos de transição de 1 <sup>a</sup> ordem, considerando efeitos de tratamento e interação tratamento e resposta prévia, no estudo sobre a condição respiratória . . . . .	93
Tabela 15 -Estimativas dos parâmetros para o modelo de transição estacionário de 1 <sup>a</sup> ordem, considerando efeitos de tratamento e interação tratamento e resposta prévia, no estudo sobre a condição respiratória . . . . .	95
Tabela 16 -Estimativas dos parâmetros fixos, $\hat{\beta}$ , e respectivos erros-padrão, $s(\hat{\beta})$ , para três modelos de transição estacionários de primeira ordem, no estudo sobre a condição respiratória . . . . .	97
Tabela 17 -Frequências de transições de segunda ordem observadas no estudo sobre a condição respiratória . . . . .	100
Tabela 18 -Total de frequências de transições de segunda ordem observadas no estudo sobre a condição respiratória . . . . .	101
Tabela 19 -Total de frequências de transições de primeira ordem observadas no estudo sobre a condição respiratória, considerando 333 observações . . . . .	102
Tabela 20 -Estimativas dos parâmetros e respectivos erros-padrão para o modelo de transição estacionário de segunda ordem, no estudo sobre a condição respiratória . . . . .	102

## 1 INTRODUÇÃO

De um modo geral, o termo medidas repetidas é comumente empregado para designar um estudo, observacional ou experimental, em que cada unidade amostral é avaliada em mais de uma ocasião (tempo, distância, dosagem, etc). Quando a ocasião utilizada é o tempo, diz-se também estudo longitudinal. Ainda, os termos medidas repetidas no tempo ou estudo longitudinal são designações estatísticas muito associada às áreas da biologia, saúde e agronomia, embora estudos com essa característica, sejam realizados e aplicados a outras áreas do conhecimento. Assim, na literatura, não é raro, encontrar o termo “dados de painel”, para esses estudos nas áreas econômicas e sociológicas, como também, “dados em fila” na área das ciências políticas. Há ainda alguns autores que preferem o uso do termo “corte transversal dos dados”.

Embora sob diferentes designações, essas variações de termo de um assunto comum, nas diferentes áreas, em síntese, referem-se a uma pesquisa, em que são tomadas observações repetidas, em geral numa escala ordenada de tempo, de uma variável resposta e um conjunto de variáveis explicativas (covariáveis) para cada uma das unidades amostrais ou experimentais. A variável resposta assim como as explicativas podem ter natureza qualitativa ou quantitativa e os métodos de análise são comuns a todas as áreas. Na verdade, o que determina os métodos para a análise de dados longitudinais é o tipo de resposta avaliada e as questões de interesse do pesquisador. De acordo com Andrade e Singer (1986), Noletto (1991) e Diggle et al. (2002), entre os objetivos básicos de um estudo longitudinal, podem-se citar:

1. Comparar o efeito de tratamentos ou grupos quanto ao padrão de variação da distribuição das respostas ao longo das ocasiões, isto é, verificar se há interação entre tratamento (ou grupo) e o fator ocasião (tempo);
2. Comparar os diferentes tratamentos, fatores ou grupos quanto às distribuições médias de resposta, isto é, verificar se há efeito de tratamento (fator ou grupo);
3. Verificar se há efeito do fator ocasião (tempo) comparando as diferentes ocasiões em relação às distribuições médias de resposta;
4. Descrever a distribuição das mudanças de respostas no tempo (transições), bem como

avaliar os efeitos ou fatores envolvidos nessas mudanças.

Os estudos longitudinais, por sua vez, contrastam e diferem dos estudos denominados séries temporais e dos estudos transversais. Nas séries temporais, somente uma unidade amostral é avaliada, várias vezes, numa escala de tempo. Por outro lado, no estudo transversal, cada um dos elementos amostrais é avaliado em uma única ocasião. Como no estudo longitudinal, cada unidade pode fornecer várias observações, isso proporciona algumas vantagens em relação aos demais planejamentos. Sob este aspecto, Andrade e Singer (1986) destacam que um estudo longitudinal proporciona condições adequadas para um controle de fatores secundários que podem influenciar a variável resposta, permitem incorporar informações sobre a variação individual na análise, além de produzir estimadores mais eficientes para os parâmetros associados a estas medidas de variabilidade. Adicionalmente, Zeger e Liang (1992) complementam, enfatizando o maior poder e a robustez para seleção de modelos, condições para o controle da heterogeneidade presente entre os elementos, análises que são insensíveis à omissão de variáveis explicativas que não mudam com o tempo, bem como, vantagens teóricas em termos de dimensionamento amostral.

Quanto à estrutura, os dados longitudinais podem ser classificados como regulares em relação ao tempo, quando os intervalos entre duas medidas consecutivas quaisquer forem constantes; também são chamados de balanceados em relação ao tempo, se todas as observações das unidades amostrais são feitas nas mesmas ocasiões. A estrutura também é dita completa quando não há perda de observações.

As técnicas clássicas de análise de dados longitudinais, em geral, são destinadas à classe dos planejamentos completos e balanceados em relação ao tempo. Também, há uma abordagem amplamente difundida para os casos em que a variável resposta é contínua, mais especificamente, quando tem distribuição normal. Nesse contexto, a metodologia da análise está alicerçada nos modelos lineares clássicos pois além da normalidade, estabelece-se uma relação funcional linear entre a parte aleatória do modelo (variável resposta) e a parte sistemática, que engloba as variáveis explicativas. Porém, as pressuposições desses modelos devem atender a uma estrutura para a matriz de variâncias e covariâncias, para a qual se espera que medidas repetidas de modo sistemático, tenham uma correlação não nula e heterocedasticidade das variâncias nas diversas ocasiões. Assim, a escolha da estrutura

da matriz de variâncias e covariâncias que melhor representa os dados é importante para as inferências acerca dos efeitos que se deseja avaliar.

Entretanto, nem sempre, a variável resposta é contínua ou tem distribuição normal, por ser oriunda de um processo de contagem, dando origem a dados de contagem, por exemplo. Demétrio (2002) salienta, ainda, que a estrutura aditiva entre o componente sistemático e o componente aleatório de um modelo nem sempre é satisfeita, assim como não há razões para se restringir à função identidade entre o componente aleatório e o sistemático, nem à condição de normalidade e à suposição de homogeneidade de variâncias.

Nesses casos, uma metodologia adequada de análise pode ser implementada através dos modelos lineares generalizados (MLGs). Propostos originariamente por Nelder e Wedderburn (1972), os MLGs são uma extensão dos modelos lineares clássicos e permitiram a unificação de uma série de técnicas de modelagem estatística. Os autores mostraram que independentemente de a variável resposta ser discreta ou contínua, na maioria das áreas de conhecimento, é possível descrever o comportamento dessa variável em função de um conjunto de variáveis explicativas, como um modelo de regressão.

Certamente, na ótica da análise para dados longitudinais, merecem destaque, o trabalho de Wedderburn (1974), que apresenta a teoria de quase-verossimilhança como uma extensão dos MLGs, e os trabalhos de Zeger e Liang (1986) e Liang e Zeger (1986) que propõem as equações de estimação generalizadas (EEG) que possibilitaram a análise de dados correlacionados, inclusive com distribuição diferente da normal.

O enfoque do presente trabalho é dado aos modelos de transição para a análise de dados longitudinais envolvendo uma única resposta do tipo binária. Dados binários ou dicotômicos são comuns em muitas áreas das ciências, nas quais, muitas vezes, há interesse em registrar a ocorrência, ou não, de um evento particular (sucesso ou fracasso). Mesmo quando a resposta não é originariamente do tipo binária, muitas vezes, faz-se um agrupamento de categorias dicotomizando as respostas. Em geral, nessas situações, deseja-se modelar a probabilidade do sucesso como função das variáveis explicativas envolvidas. A regressão logística tem sido utilizada para tal propósito, devido às suas facilidades para interpretação das estimativas dos parâmetros e da razão de chances. Apesar de sua popularidade entre os usuários de Estatística, muitas técnicas padrões em diagnósticos de ajuste de modelos bem

como análise de resíduos não se aplicam para os casos em que a resposta é do tipo binária.

Além desse problema, em estudos longitudinais com resposta dessa natureza, existe uma dicotomia conceitual e metodológica para a análise desses dados. Noletto (1991) observa que a escolha do modelo a ser utilizado depende basicamente dos objetivos desejados por ocasião do planejamento da pesquisa. Nesse contexto, distinguem-se dois tipos de modelos: marginais e de transição. Assim, os modelos marginais dão conta dos objetivos (1), (2) e (3), mas não são aptos para modelar as mudanças ou transições entre os tempos de observação, isto é, não permitem avaliar o comportamento das mudanças de respostas em cada ocasião. Portanto, para atender ao objetivo (4) deve-se optar por um modelo de transição. Em particular, para dados binários, os modelos marginais caracterizam a probabilidade marginal do sucesso em cada ocasião, enquanto que os modelos de transição caracterizam as probabilidades condicionais do sucesso, dada a história da unidade amostral ou experimental. Devido a isso, os modelos de transição são também baseados em processos estocásticos, uma vez que o interesse é estimar as probabilidades de transição, isto é, as probabilidades de passar de uma categoria para a outra em ocasiões sucessivas. Na ótica dos processos estocásticos, o termo transição está relacionado ao tempo, isto é, o que ocorre de uma ocasião para a outra. Portanto, uma transição não implica necessariamente que a categoria de resposta numa dada ocasião seja distinta da categoria na ocasião anterior.

Há vários tipos de processos estocásticos que podem ser considerados em um modelo de transição, mas, em geral, a propriedade markoviana é a suposição estocástica mais freqüentemente adotada. Nesse sentido, a teoria que está por trás de um modelo de transição não é tão simples pois a inferência referente a tais modelos implica em trabalhar simultaneamente com métodos dos MLGs e com técnicas de processos estocásticos. Por exemplo, para ajustar modelos de transição pode-se usar o método da máxima verossimilhança, através de algum processo iterativo como o método score de Fisher e, como nos modelos marginais, testes do tipo Wald podem ser utilizados para testar hipóteses convenientes. Porém, como se está diante de um processo estocástico, algumas suposições têm que ser verificadas, como a suposição de Markov, a qual condiciona a probabilidade de estar numa dada categoria num tempo específico ao resultado no tempo imediatamente anterior e à condição de estacionariedade. Esse fato é de importância para dados categorizados e, em particular para dados



binários, em que se tem interesse em obter as matrizes de probabilidades de transição, que para o caso de não rejeição da hipótese de estacionariedade são consideradas estocasticamente iguais, e, portanto, o modelo de transição pode ser sintetizado por uma única matriz de probabilidades.

De certa forma, seja o processo estacionário ou não, uma questão de interesse prático relevante é poder comparar os diferentes tratamentos envolvidos a partir das matrizes de probabilidades de transição. Uma outra suposição que pode ser explorada é a de que há heterogeneidade entre os indivíduos, devido a fatores não controláveis, de tal forma que para se estimar a matriz de probabilidades de transição, deva-se incluir efeitos aleatórios convenientes. Esses problemas, em geral, aparecem como lacunas na literatura desses modelos, justificando estudos nessa área.

Nesse contexto, este trabalho tem como objetivo básico desenvolver, implementar e aplicar modelos de transição para análise de dados longitudinais com variável resposta binária. Como decorrência, constituem-se em objetivos específicos:

- i. Estudar os modelos existentes;
- ii. Aplicar as técnicas convencionais de análise de resíduos e diagnósticos dos MLGs para esses modelos;
- iii. Definir hipóteses apropriadas para comparação de tratamentos e desenvolver testes mais apropriados para esses casos;
- iv. Considerar no modelo de transição a inclusão de efeitos aleatórios.

## 2 DESENVOLVIMENTO

### 2.1 Revisão de Literatura

O desenvolvimento e a aplicação de técnicas para a análise de dados longitudinais têm um longo histórico. Entre os trabalhos clássicos da década de 50 estão os de Box (1950), Geisser e Greenhouse (1958), Rao (1959). Na década de 60, Potthoff e Roy (1964), Rao (1965) e Grizzle e Allen (1969) discutem as curvas de crescimento ou modelos polinomiais, acrescentando à análise de dados longitudinais a possibilidade de previsão de efeitos de tratamentos numa dada ocasião de tempo. Nos anos 70, destaque para o trabalho de Fearn (1975) que apresenta uma técnica de modelagem de curva de crescimento à luz da teoria Bayesiana e de Harville (1977) que desenvolve um modelo de efeitos aleatórios, no qual supõe que as observações repetidas têm um componente aleatório comum. Já nos anos 80, Laird e Ware (1982) apresentam uma visão mais geral sobre dados longitudinais, trabalhando com curvas de crescimento e com modelos de efeitos aleatórios. Azzalini (1984) apresenta uma estrutura de modelo com erro auto-regressivo, em que as correlações entre duas observações seguidas decrescem como uma função geométrica do tempo. Ware (1985) também apresenta uma visão geral de modelos lineares para dados longitudinais.

Esses trabalhos, em sua maioria, discutem e abordam métodos para os casos em que a variável resposta é normal. Essa tendência histórica, de acordo com Singer e Andrade (2000), é explicada pelo bom comportamento da distribuição normal multivariada. Como conseqüência, isso justifica o fato de que hoje, tem-se uma rica classe de modelos e procedimentos de análise para dados longitudinais “normais”, inclusive, implementado em muitos *softwares*.

A partir de meados da década de 70, entretanto, começam a aparecer na literatura outros trabalhos sobretudo quando a variável resposta não tem distribuição normal (COX, 1970; KOCH et al., 1977; KORN; WHITTEMORE, 1979; STIRATELLI; LAIRD; WARE, 1984; OCHI; PRENTICE, 1984; ANDERSON; AITKIN, 1985; ZEGER; LIANG; SELF 1985; LIANG; ZEGER, 1986; ZEGER; LIANG, 1986; BRESLOW; CLAYTON, 1993; MCGILCHRIST, 1994; LEE; NELDER, 1996; DIGGLE et al., 2002; MOLENBERGHS; VERBEKE, 2005; entre outros). Tal fato, certamente, é impulsionado, em parte, pela ne-

cessidade em se atenderem às mais variadas áreas de conhecimento, nas quais, muitas vezes, as variáveis respostas observadas repetidamente no tempo são oriundas de processos de contagem, ou são dicotômicas, ou estão em uma escala ordinal. Por exemplo, na área agrônômica, de saúde e biologia é comum o registro da ocorrência, ou não, de um evento particular ao longo do tempo. Assim, pode-se, em mais de uma ocasião, verificar se o indivíduo (uma planta, um animal ou uma pessoa), responde, ou não, a determinados tratamentos. Somado às necessidades de pesquisas e estudos, há de se considerar que estes avanços metodológicos foram possíveis graças à teoria dos modelos lineares generalizados e ao notável avanço computacional.

### 2.1.1 Modelos lineares generalizados

Nesse contexto, a fim de construir modelos para dados longitudinais não normais, faz-se necessária uma breve revisão da teoria dos modelos lineares generalizados, da quase-verossimilhança e extensões. Algumas referências são Nelder e Wedderburn (1972), Wedderburn (1974), Cordeiro (1986), McCullag e Nelder (1989), Collett (1991), Dobson (2001), Demétrio, (2002), Paula (2004).

De acordo com Nelder e Wedderburn (1972), os MLGs podem ser usados quando se tem uma única variável resposta  $Y$  associada a um conjunto de variáveis explicativas  $x_1, x_2, \dots, x_p$ . Para uma amostra de  $n$  observações  $(y_i, \mathbf{x}_i)$ , em que  $\mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{pi})'$  é o vetor coluna de variáveis explicativas, o modelo linear generalizado tem três componentes:

- i. Componente Aleatório - constituído de um conjunto de variáveis respostas, isto é, variáveis aleatórias independentes,  $Y_1, Y_2, \dots, Y_n$ , provenientes de uma mesma distribuição pertencente à família exponencial na forma canônica, com médias  $\mu_1, \mu_2, \dots, \mu_n$ , isto é,

$$f(y_i; \theta_i, \phi) = \exp \left\{ \frac{1}{a_i(\phi)} [y_i \theta_i - b(\theta_i)] + c(y_i, \phi) \right\}, \quad (1)$$

em que  $b(\cdot)$  e  $c(\cdot)$  são funções conhecidas e  $\theta_i$  é o parâmetro canônico. Em geral,  $a_i(\phi) = \frac{\phi}{w_i}$ , sendo  $w_i$  pesos a priori e  $\phi > 0$ , um parâmetro de escala conhecido. Além disso, demonstra-se que:

$$E(Y_i) = b'(\theta_i) = \mu_i \quad \text{e} \quad \text{Var}(Y_i) = a_i(\phi) b''(\theta_i) = a_i(\phi) V(\mu_i),$$

sendo  $V(\mu_i) = \frac{\partial \mu_i}{\partial \theta_i}$  a função de variância, uma função que depende unicamente da média. A função de variância desempenha papel fundamental nos MLGs, pois caracteriza a distribuição.

- ii. Componente sistemático - constituído por um conjunto de  $p$  variáveis explicativas, isto é,  $\mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{pi})'$ , para  $i = 1, 2, \dots, n$ , que entram no modelo na forma de uma soma linear de seus efeitos, constituindo o vetor dos preditores lineares,  $\boldsymbol{\eta}$ , de dimensão  $n \times 1$ , ou seja, com elementos:

$$\eta_i = \sum_{j=1}^p x_{ij}\beta_j,$$

ou sob a forma matricial,

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta},$$

em que  $\mathbf{X}$  é a matriz do modelo de dimensões  $n \times p$  e  $\boldsymbol{\beta}$ , o vetor de  $p$  parâmetros desconhecidos.

- iii. Função de Ligação - uma função que faz a ligação entre os componentes aleatório e sistemático do modelo, relacionando a média ao preditor linear, isto é,

$$\eta_i = g(\mu_i),$$

sendo  $g(\cdot)$  uma função monótona e derivável.

Demétrio (2002) salienta que uma decisão importante na escolha de um modelo linear generalizado é a escolha do trinômio - distribuição da variável resposta, matriz do modelo e função de ligação. As funções de ligação podem ser obtidas diretamente da forma da distribuição de probabilidade ou de algum modelo relacionado a  $\mu_i$ . Quando a função de ligação é escolhida tal que  $g(\mu_i) = \theta_i$ , ela é denominada canônica.

Como exemplo particular, considere um estudo no qual a variável resposta é dicotômica (sucesso ou fracasso) e associada a essa variável dependente têm-se  $p$  variáveis explicativas. Para uma amostra aleatória de  $n$  indivíduos, a variável aleatória  $Y_i$  denota a

ocorrência do sucesso (1) ou fracasso (0). Admita que  $Y_i \sim \text{Bernoulli}(\pi_i)$ , então a variável resposta tem distribuição pertencente à família exponencial em forma canônica (1):

$$f(y_i; \pi_i) = \exp \left\{ y_i \ln \left( \frac{\pi_i}{1 - \pi_i} \right) + \ln(1 - \pi_i) \right\},$$

obtendo-se:

$$\theta_i = \ln \left( \frac{\pi_i}{1 - \pi_i} \right) \quad (\text{parâmetro canônico}) \quad \Rightarrow \quad \pi_i = \frac{\exp(\theta_i)}{1 + \exp(\theta_i)},$$

sendo,

$$a_i(\phi) = 1, \quad b(\theta_i) = -\ln(1 - \pi_i) = \ln(1 + \exp(\theta_i)) \quad \text{e} \quad c(y_i, \phi) = 0.$$

Demonstra-se, ainda, que a média e a variância de  $Y_i$  são obtidas por:

$$E(Y_i) = b'(\theta_i) = \pi_i,$$

e

$$\text{Var}(Y_i) = a_i(\phi)b''(\theta_i) = \pi_i(1 - \pi_i),$$

e como  $\mu_i = \pi_i$ , segue que a função de variância que caracteriza o modelo é:

$$V(\mu_i) = \mu_i(1 - \mu_i).$$

O objetivo geral nesse tipo de estudo é modelar a probabilidade do sucesso  $\pi_i$ , como função das variáveis explicativas incluídas no modelo. Nessas condições, admitindo-se ligação canônica (logística), um modelo linear generalizado usual é o modelo de regressão logística:

$$\eta_i = \ln \left( \frac{\pi_i}{1 - \pi_i} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p.$$

Portanto, nesse exemplo, tem-se um modelo baseado na família exponencial com um parâmetro desconhecido, cujas médias são não lineares, num conjunto de parâmetros lineares, ou seja:

$$\mu_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}.$$

Apesar de ser mais usual a ligação logito, há outras funções que podem fazer este papel, como a função probito e a complemento *log-log*, isto é:

$$\eta_i = \phi^{-1}(\pi_i)$$

$$\eta_i = \ln[-\ln(1 - \pi_i)].$$

Há, ainda, a família de funções de ligação proposta por Aranda-Ordaz (1981) dada por:

$$\eta_i = \ln \left[ \frac{(1 - \pi_i)^{-\lambda} - 1}{\lambda} \right],$$

sendo  $\lambda$  uma constante conhecida e que tem como casos particulares a função logito, quando  $\lambda = 1$  e a função complemento *log-log* quando  $\lambda \rightarrow 0$ .

### 2.1.1.1 Estimação dos parâmetros do MLG

O método proposto por Nelder e Wedderburn (1972) para a estimação do vetor de parâmetros  $\beta$  foi o da máxima verossimilhança. Considerando uma amostra aleatória de  $n$  observações de uma distribuição pertencente à família exponencial (1), a função de verossimilhança é dada por:

$$L = L(\theta_i, \phi, y_i) = \prod_{i=1}^n f(y_i; \theta_i, \phi) = \exp \left\{ \sum_{i=1}^n \left[ \frac{1}{a_i(\phi)} [y_i \theta_i - b(\theta_i)] + c(y_i, \phi) \right] \right\},$$

cujos logaritmo é:

$$\ell = \ell(\theta_i, \phi, y_i) = \ln L(\theta_i, \phi, y_i) = \sum_{i=1}^n \left\{ \frac{1}{a_i(\phi)} [y_i \theta_i - b(\theta_i)] + c(y_i, \phi) \right\}. \quad (2)$$

Derivando-se a função (2) em relação a  $\beta_j$ , usando-se a regra da cadeia, obtém-se a função escore:

$$U_j = \frac{\partial \ell}{\partial \beta_j} = \frac{\partial \ell}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \sum_{i=1}^n \frac{1}{a_i(\phi)} (y_i - \mu_i) \left( \frac{\partial \theta_i}{\partial \eta_i} \right) x_{ij} \quad (3)$$

sendo que,

$$\frac{\partial \theta_i}{\partial \eta_i} = \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} = \frac{1}{V(\mu_i)} \left( \frac{\partial \mu_i}{\partial \eta_i} \right).$$

As equações (3) são não lineares e, em geral, não possuem soluções exatas. Soluções aproximadas podem ser encontradas por processo iterativo, por exemplo do tipo Newton-Raphson ou método escore de Fisher. Pelo método escore de Fisher, tem-se:

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}'\mathbf{W}^{(m)}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{(m)}\mathbf{z}^{(m)}, \quad (4)$$

em que  $\mathbf{X}$  é a matriz de especificação do modelo;  $\mathbf{W}$  é a matriz diagonal dos pesos, com elementos  $w_i^* = \frac{w_i}{V(\mu_i)} \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2$  e  $\mathbf{z}^{(m)} = \mathbf{X}\boldsymbol{\beta}^{(m)} + \Delta^{(m)}(\mathbf{y} - \boldsymbol{\mu})^{(m)} = \boldsymbol{\eta}^{(m)} + \Delta^{(m)}(\mathbf{y} - \boldsymbol{\mu})^{(m)}$  é o vetor da variável dependente ajustada no passo  $m$ , com  $\Delta = \text{diag} \left( \frac{\partial \eta_i}{\partial \mu_i} \right)$ .

O método usual para iniciar o algoritmo é especificar uma estimativa inicial  $\boldsymbol{\beta}^{(0)}$  e usando (4), iterá-la sucessivamente até obter convergência. Para efeito de passo inicial, pode-se especificar o vetor  $\hat{\boldsymbol{\beta}}^{(0)} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\eta}}$ , sendo que  $\hat{\eta}_i = g(\hat{\mu}_i) = g(y_i)$ .

### 2.1.1.2 Função desvio, estatística generalizada de Pearson e seleção de modelos

O ajuste de um modelo a um conjunto de dados pode ser considerado como uma maneira de se substituir o vetor de dados observados  $\mathbf{y}$  por um conjunto de valores estimados  $\hat{\boldsymbol{\mu}}$  para um modelo com número relativamente pequeno de parâmetros. A  $n$  observações de uma variável aleatória  $Y$ , pode-se ajustar do modelo mais simples, com um único parâmetro  $\mu$  (modelo nulo) até o modelo mais complexo possível com  $n$  parâmetros (modelo saturado). Esses dois extremos podem não ser informativos. O modelo nulo porque resume demais e o saturado porque atribui toda a variabilidade dos dados ao componente sistemático. A idéia central, então, é propor um modelo intermediário que seja razoável para explicar bem os dados (parcimonioso).

Uma medida de discrepância para avaliar o ajuste de um modelo aos dados, comparativamente ao modelo saturado, proposta por Nelder e Wedderburn (1972) é a função desvio (*deviance*), definida a partir de:

$$S_p = 2(\hat{\ell}_{(n)} - \hat{\ell}_{(p)}),$$

em que  $\hat{\ell}_{(n)}$  e  $\hat{\ell}_{(p)}$  denotam os máximos do logaritmo da função de verossimilhança para os modelos saturado e corrente, respectivamente. Sendo  $\hat{\theta} = \hat{\theta}(\hat{\mu})$  e  $\tilde{\theta} = \tilde{\theta}(y)$  as estimativas do

parâmetro canônico para os dois modelos, então:

$$S_p = \frac{1}{\phi} \sum_{i=1}^n 2w_i \left\{ y_i [\tilde{\theta}_i - \hat{\theta}_i] - b(\tilde{\theta}_i) + b(\hat{\theta}_i) \right\} = \frac{1}{\phi} D_p,$$

em que  $S_p$  é denominada *scaled deviance* e  $D_p$  é a função desvio.

A função desvio pode ser interpretada como uma medida de distância dos valores ajustados em relação aos valores observados e seu valor é sempre maior do que ou igual a zero. Assim quando se trabalha com modelos encaixados seu valor decresce, à medida que se incorporam variáveis no modelo. Quanto melhor for o ajuste do modelo, menor deverá ser o valor do desvio, e empiricamente, ele pode dar um indicativo da qualidade do ajuste. Para dados binários, conforme Demétrio (2002), a função desvio é apenas uma função dos dados e, portanto, não é informativa a respeito do ajuste do modelo. De acordo com McCullagh e Nelder (1989) para testar o ajuste de um modelo linear generalizado, sem muito rigor, no caso das distribuições binomial e Poisson, pode-se usar o percentil da distribuição qui-quadrado, com  $n-p$  graus de liberdade, pois sob determinadas condições de regularidade,  $S_p$  possui distribuição assintótica  $\chi_{n-p}^2$ . Portanto, se:

$$S_p \leq \chi_{n-p, \alpha\%}^2,$$

admite-se que há evidências, ao nível de significância  $100\alpha\%$ , de que o modelo proposto (corrente) está bem ajustado aos dados.

Outra medida de discrepância é a estatística  $X^2$  generalizada de Pearson, definida por:

$$X^2 = \sum_{i=1}^n w_i \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$$

que tem distribuição exata de  $\chi^2$  para o caso normal e assintótica para os demais casos.

Os métodos de seleção de modelos lineares generalizados baseiam-se em um método que pode ser visto como uma generalização da análise da variância. Dada uma seqüência de modelos encaixados, isto é, se  $M_p$  e  $M_q$  são dois modelos encaixados com  $p$  e  $q$  parâmetros ( $p < q$ ), a estatística  $D_p - D_q$  com  $(q-p)$  graus de liberdade é interpretada como uma medida de variação dos dados, explicada pelos termos que estão em  $M_q$  e não estão em  $M_p$ , incluídos os efeitos dos termos de  $M_p$  e ignorando quaisquer outros efeitos dos termos que não estejam em  $M_q$ .



Para  $\phi$  conhecido, o teste para seleção do modelo é dado por:

$$S_p - S_q = \frac{1}{\phi} (D_p - D_q) \sim \chi_{q-p}^2,$$

que coincide com o teste da razão de verossimilhanças. Para as distribuições binomial e Poisson, tem-se que  $\phi = 1$ . Para as distribuições em que  $\phi$  é desconhecido, Jorgensen (1987) mostra que:

$$F = \frac{(D_p - D_q)/(q - p)}{\hat{\phi}} \sim F_{q-p, n-m},$$

em que  $\hat{\phi}$  é estimado, preferencialmente a partir do modelo maximal (com  $m$  parâmetros), isto é:

$$\hat{\phi} = \frac{D_m}{n - m} \quad \text{ou} \quad \hat{\phi} = \frac{X_m^2}{n - m}.$$

Testes do tipo escore e Wald também podem ser usados para se testar um subconjunto de parâmetros. Em particular, se se deseja testar o vetor de parâmetros de dimensão  $p$ , isto é,  $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$  versus  $H_a : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ , o teste de Wald é dado por:

$$W = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' [Var(\hat{\boldsymbol{\beta}})]^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0),$$

com  $W \sim \chi_p^2$ , sob  $H_0$  quando  $n \rightarrow \infty$ .

O critério da informação de Akaike (AIC) também pode ser usado e sua expressão, segundo Paula (2004), como função do desvio é dada por:

$$AIC = D_p + 2p,$$

em que  $p$  é o número de parâmetros.

### 2.1.1.3 Quase-verossimilhança

Para a utilização do método de máxima verossimilhança é preciso especificar a distribuição da variável resposta, que é determinante para a construção da função de verossimilhança. Alternativamente, o método de quase-verossimilhança, proposto por Wedderburn (1974), requer apenas especificação de uma função de variância para a variável resposta e a relação funcional entre a resposta média e os parâmetros  $\boldsymbol{\beta}$ . Nesse contexto,

sejam  $n$  observações para a variável aleatória  $Y_i$ , tal que  $E(Y_i) = \mu_i$  e  $\text{Var}(Y_i) = \sigma^2 V(\mu_i)$ , em que  $V(\mu_i)$  é a função de variância e  $\sigma^2$  o parâmetro de dispersão. Suponha também que  $\mu_i$  é uma função do conjunto de parâmetros de interesse  $\beta_1, \beta_2, \dots, \beta_p$ . O logaritmo da função de quase-verossimilhança para  $Y_i$  é definido por:

$$Q(\mu_i, y_i) = \frac{1}{\sigma^2} \int_{y_i}^{\mu_i} \frac{y_i - t}{V(t)} dt,$$

e, portanto,

$$\frac{\partial Q(\mu_i, y_i)}{\partial \mu_i} = \frac{y_i - \mu_i}{\sigma^2 V(\mu_i)}.$$

O logaritmo da função de verossimilhança aparece como um caso particular do logaritmo da função de quase-verossimilhança. Tal como nos MLGs, o interesse central nos modelos de quase-verossimilhança está nos parâmetros  $\beta$ ,  $\sigma^2$  é tratado, portanto, como parâmetro de perturbação. O logaritmo da função de distribuição conjunta, para a totalidade das observações é dado por:

$$Q(\mu, y) = \sum_{i=1}^n Q(\mu_i, y_i) = \frac{1}{\sigma^2} \sum_{i=1}^n \int_{y_i}^{\mu_i} \frac{y_i - t}{V(t)} dt.$$

A estimativa do vetor de parâmetros é obtida maximizando-se a função de quase-verossimilhança total em relação aos parâmetros  $\beta$ , via processo iterativo de mínimos quadrados ponderados. McCullagh e Nelder (1983) mostram que a função escore para  $\beta$  é expressa por:

$$U(\beta) = \frac{1}{\sigma^2} \mathbf{D}' \mathbf{V}^{-1} (\mathbf{y} - \boldsymbol{\mu}),$$

em que  $\mathbf{V} = \text{diag}[V(\mu_i)]$ ,  $\mathbf{D} = \frac{\partial \boldsymbol{\mu}}{\partial \beta} = W^{1/2} V^{1/2} X$ ,  $\mathbf{W} = \text{diag}(w_i^*)$ ,  $w_i^* = \frac{w_i}{V(\mu_i)} \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2$ ,  $\mathbf{y} = (y_1, y_2, \dots, y_n)'$  e  $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)'$ .

Assim, a solução  $\hat{\beta}$  sai da resolução do sistema  $U(\hat{\beta}) = \mathbf{0}$ . Usando o método escore de Fisher, tem-se:

$$\beta^{(m+1)} = \beta^{(m)} + [\mathbf{D}^{(m)} \mathbf{V}^{-1(m)} \mathbf{D}^{(m)}]^{-1} \mathbf{D}'^{(m)} \mathbf{V}^{-1(m)} [\mathbf{y} - \boldsymbol{\mu}^{(m)}],$$

o qual independe do parâmetro de dispersão.

McCullagh e Nelder (1983) definem, por analogia aos MLGs, uma medida de discrepância, chamada de função quase-desvio:

$$D(y, \hat{\mu}) = 2\sigma^2 \{Q(y, y) - Q(\hat{\mu}, y)\} = 2 \sum_{i=1}^n \int_{\hat{\mu}_i}^{y_i} \frac{y_i - t}{V(t)} dt.$$

Dessa forma, se  $D(y, \hat{\mu}^{(0)})$  e  $D(y, \hat{\mu}^{(1)})$  são as funções de quase-desvio de dois modelos encaixados com  $p$  e  $q$  parâmetros ( $p < q$ ), para amostras grandes, tem-se:

$$\frac{1}{\sigma^2} [D(y, \hat{\mu}^{(0)}) - D(y, \hat{\mu}^{(1)})] \sim \chi_{q-p}^2,$$

para  $\sigma^2$  fixo que pode ser estimado como nos MLGs.

#### 2.1.1.4 Análise de resíduos e diagnósticos

As técnicas usadas nos MLGs para análise de resíduos e diagnósticos são análogas às dos modelos lineares clássicos, com resposta normal, tanto para as técnicas formais (baseadas em testes de hipóteses) como nas informais, baseadas em recursos gráficos. Pregibon (1981) foi um dos precursores na adaptação dessas técnicas para os MLGs. Williams (1984, 1987) e McCullagh (1987) também apresentam contribuições para a definição de um resíduo com melhores propriedades.

Sem perda de generalidade, os vetores  $\mathbf{y}$  e  $\hat{\boldsymbol{\mu}}$  dos modelos clássicos são, alternativamente, substituídos nos MLGs pela variável dependente ajustada  $\mathbf{z}$  e pelo preditor linear  $\hat{\boldsymbol{\eta}}$ . A variância residual, dada pelo quadrado médio do resíduo nos modelos clássicos é substituída por uma estimativa consistente para  $\phi$ . A matriz de projeção  $\mathbf{H}$ , é dada por:

$$\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}' \mathbf{W} \mathbf{X})^{-1} \mathbf{X}' \mathbf{W}^{1/2}.$$

Os resíduos mais usuais para análise e diagnósticos dos MLGs são:

- i. Resíduos de Pearson generalizados

$$r_i^p = \frac{y_i - \hat{\mu}_i}{\sqrt{\frac{\hat{\phi}}{w_i} V(\hat{\mu}_i)}}, \quad (5)$$

em que  $\hat{\phi}$  é uma estimativa consistente para o parâmetro de dispersão  $\phi$  e  $w_i$  são os pesos a priori.

ii. Resíduos de Pearson generalizados padronizados internamente

$$r_i^{P^i} = \frac{y_i - \hat{\mu}_i}{\sqrt{\frac{\hat{\phi}}{w_i} V(\hat{\mu}_i)(1 - h_i)}}, \quad (6)$$

em que  $h_i$  denota o  $i$ -ésimo elemento da diagonal da matriz de projeção  $\mathbf{H}$ . Williams (1984) mostrou através de estudos de simulação de Monte Carlo, que a distribuição dos resíduos dada por (6) não tem distribuição normal, mesmo para grandes amostras.

iii. Componentes do desvio

$$r_i^D = \pm(y_i - \hat{\mu}_i) \sqrt{\frac{2w_i}{\hat{\phi}} [y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)]}. \quad (7)$$

iv. Componentes do desvio padronizados internamente

$$r_i^{D^i} = \frac{r_i^D}{\sqrt{1 - h_i}},$$

que é a versão padronizada do resíduo (7), segundo McCullagh (1987), que usa as aproximações propostas por Cox e Snell (1968). É o mais utilizado, tendo em vista que, dentre todos, a distribuição de  $r_i^{D^i}$  é a que mais se aproxima da normal, conforme estudos de simulação realizados por Williams (1984).

Entre as principais técnicas exploratórias para diagnósticos dos MLGs, destacam-se:

1) Gráfico semi-normal de probabilidades

Uma alternativa para se verificar se um modelo está bem ajustado aos dados é utilizar a técnica exploratória do gráfico semi-normal de probabilidades. Esse gráfico é obtido plotando-se os valores absolutos ordenados de algum tipo de resíduo contra as estatísticas de ordem esperadas. Um modelo adequado apresenta um padrão, em que os pontos se distribuem em torno de uma reta. A fim de estabelecer um padrão de comparação, Atkinson (1985) propôs a adição ao gráfico de um envelope simulado, a partir de amostras simuladas através do modelo proposto e dos valores estimados para os parâmetros. Sob o modelo correto, os valores observados caem dentro do envelope.

## 2) Verificação de pontos discrepantes e/ou influentes

Na análise de diagnósticos é importante a identificação de pontos influentes, que são observações, cuja omissão do conjunto de dados, resulta em mudanças significativas no ajuste do modelo. Uma observação influente pode ser, ou não, um dado discrepante. O estudo da diagonal principal da matriz de projeção  $\mathbf{H}$  está relacionado com o conceito de ponto de alavanca e avalia a influência de  $y_i$  sobre seu próprio valor ajustado  $\hat{y}_i$ , expressa por  $\frac{\partial \hat{y}_i}{\partial y_i}$ , que coincide no caso do modelo linear clássico com o  $i$ -ésimo elemento da diagonal principal da matriz de projeção  $\mathbf{H}$ ,  $h_{ii}$ . Paula (2004) sugere examinar os pontos para os quais  $h_{ii} > \frac{2p}{n}$ , que são conhecidos como pontos de alavanca. A matriz  $\mathbf{H}$  para os MLGs depende das variáveis explanatórias, da função de ligação e da função de variância, o que certamente dificulta a interpretação dos elementos  $h_{ii}$  e, por isso, em geral, a técnica de deleção de pontos é usada para avaliar o impacto de uma observação nas estimativas do modelo. Alguns índices, como a distância de Cook (COOK; WEISBERG, 1982), foram adaptados para os MLGs e medem a alteração provocada no valor ajustado pela retirada da  $i$ -ésima observação. A distância de Cook, é dada por:

$$D_i = \left( \frac{\hat{h}_{ii}}{1 - \hat{h}_{ii}} \right) (r_i^{D_i})^2,$$

sendo suficiente para chamar a atenção sobre pontos discrepantes e influentes.

## 3) Verificação da função de ligação

Informalmente, a verificação da função de ligação pode ser feita através de dispositivos gráficos. Dois gráficos podem ser usados e o mais simples consiste no diagrama de dispersão da variável dependente ajustada  $\mathbf{z}$  contra o preditor linear,  $\hat{\boldsymbol{\eta}}$ . O padrão nulo, que é uma reta, indica que a função de ligação está adequada. Outro é o gráfico da variável adicionada,  $\mathbf{u} = \hat{\boldsymbol{\eta}} \odot \hat{\boldsymbol{\eta}}$ . Para obtê-lo, inicialmente ajusta-se o modelo linear generalizado, com preditor  $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$  e, a seguir, obtêm-se os resíduos de Pearson generalizados, o qual segundo Wang (1985) é dado por  $\mathbf{W}^{-1/2}\mathbf{s}$ , em que  $s_i = \frac{y_i - \hat{\mu}_i}{a_i(\phi)V(\hat{\mu}_i)} \frac{\partial \mu_i}{\partial \eta_i}$ . A seguir, ajusta-se a regressão ponderada pela matriz  $\mathbf{W}$  de  $\mathbf{u}$  contra as variáveis explicativas  $\mathbf{X}$  e, obtêm-se também seus resíduos de Pearson, que podem ser dados por  $(\mathbf{I} - \mathbf{H})\mathbf{W}^{1/2}\mathbf{u}$ . O diagrama de dispersão desses dois resíduos, com padrão nulo (reta),

indica adequação da função de ligação. Formalmente, pode-se usar o teste da razão de verossimilhança, que consiste em adicionar  $u$  como uma covariável extra no modelo e verificar se ocorre mudança significativa no desvio. Se isso ocorrer, a função de ligação é inadequada.

#### 4) Verificação da função de variância

A função de variância é definida pela distribuição da variável resposta, mas é afetada pela escolha errada da função de ligação, pela presença de pontos discrepantes e escala incorreta de variáveis explicativas. Uma forma exploratória de se diagnosticar a adequação da função de variância é usar o gráfico dos resíduos absolutos do modelo contra alguma função dos valores ajustados, transformados de tal forma a ter variância constante, de acordo com a distribuição. O padrão nulo para esse gráfico é uma distribuição aleatória dos resíduos com média zero e amplitude constante, indicando adequação da função de variância. Formalmente, pode-se usar o teste da razão de verossimilhança ou o teste score. Nesse caso, é necessário indexar a função de variância por um parâmetro  $\lambda$ ,  $V(\mu) = \mu^\lambda$ , e fazer o teste para  $H_0 : \lambda = \lambda_0$ .

#### 5) Verificação de escala de covariável

Para inspecionar se a escala de uma dada covariável está correta, pode-se usar o gráfico dos resíduos parciais. Inicialmente, ajusta-se o modelo com preditor linear  $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma}\mathbf{x}$ , obtendo-se a estimativa  $\hat{\boldsymbol{\gamma}}$  e os resíduos  $\mathbf{W}^{-1/2}\mathbf{s}$ . Constroem-se, então, os resíduos parciais (ou resíduos mais componentes),  $\mathbf{W}^{-1/2}\mathbf{s} + \hat{\boldsymbol{\gamma}}\mathbf{x}$ , que são plotados contra  $\mathbf{x}$ , e sinaliza se a escala de  $x$  é adequada. Formalmente, para o teste da escala de uma covariável, deve-se colocar  $x$  em uma família de funções indexada por  $\lambda$ , calcular a função desvio para um conjunto de valores de  $\lambda$  e estimar  $\hat{\lambda}$  como o valor que leva ao menor desvio. Esse método equivale ao teste do perfil de verossimilhança, pois o ajuste para  $\hat{\lambda}$  é comparado com o valor inicial 1.

#### 6) Adaptações para os modelos de quase-verossimilhança

O não conhecimento da verdadeira função de verossimilhança de  $\boldsymbol{\beta}$  dificulta algumas técnicas de diagnóstico, segundo Paula (2004). Em geral, tanto o estudo dos resíduos

como a análise das medidas de influência dependem do conhecimento de  $L(\boldsymbol{\beta})$ . O que tem sido feito para os modelos de quase-verossimilhança é o gráfico dos resíduos generalizados de Pearson (5) contra alguma função dos valores ajustados, isto é,  $g(\hat{\mu})$ , sendo sem sentido para dados binários. Um padrão aleatório com média zero desse gráfico, indica adequacidade da função de variância. Os valores dados por

$$r_i^{QD} = \frac{\pm d_i}{\hat{\phi} \sqrt{1 - \hat{h}_{ii}}}, \quad (8)$$

em que  $d_i$  é a raiz quadrada do componente de quase-desvio e  $h_{ii}$  é o  $i$ -ésimo elemento da diagonal da matriz de projeção  $\mathbf{H} = \mathbf{V}^{-1/2} \mathbf{D} (\mathbf{D}' \mathbf{V}^{-1} \mathbf{D})^{-1} \mathbf{D}' \mathbf{V}^{-1/2}$ , podem ser usados como resíduos. O gráfico dos resíduos, especialmente os obtidos por (8), contra a ordem das observações pode mostrar a existência de pontos influentes. Adicionalmente, gráficos dos resíduos parciais, da variável adicionada e semi-normal de probabilidades com envelopes simulados podem ser adaptados para auxiliar nas técnicas de diagnóstico.

### 2.1.2 Tópicos de processos estocásticos

A teoria dos processos estocásticos tem aplicação em diversos ramos da ciência. Em particular, em problemas relacionados a séries temporais e medidas repetidas no tempo, essa teoria pode ser empregada para estabelecer um modelo que possibilita a predição bem como para descrever a dependência serial das observações. Modelos auto-regressivos e modelos de transição são exemplos de aplicações. Como essa teoria é ampla, são apresentadas aqui, apenas algumas noções de fundamentos para compreensão do desenvolvimento deste trabalho. Uma descrição mais completa da teoria dos processos estocásticos é dada em Karlin e Taylor (1975), Kovács (1996), Ross (1996), Bhat e Miller (2002), Basu (2003), Lindsey (2004), entre outros.

#### 2.1.2.1 Definição de processo estocástico

Seja  $\tau$  um conjunto arbitrário. Um processo estocástico é uma coleção de variáveis aleatórias  $\{X(t), t \in \tau\}$  definidas num mesmo espaço de probabilidades  $(\Omega, A, P)$ . O conjunto  $\tau$  é um subconjunto não vazio dos reais,  $\tau \subseteq R$ , e refere-se ao tempo ou ao espaço em que o processo ocorre. Quando  $\tau$  assume um número finito ou enumerável de valores, isto

é,  $\{t_0, t_1, t_2, \dots\}$  diz-se que o processo é de parâmetro discreto, caso contrário, diz-se que o processo é de parâmetro contínuo. Por outro lado, o valor assumido pela variável aleatória  $X(t)$ , é denominado estado do processo em  $t$ . O conjunto de todos os possíveis valores dessas variáveis aleatórias  $\{X(t), t \in \tau\}$ , denotado por  $S$ , é chamado espaço de estados do processo, que também pode ser discreto ou contínuo.

No contexto dos objetivos desse trabalho,  $\tau$  refere-se ao tempo e  $S$  é um espaço discreto de estados, de tal forma que os processos considerados descrevem a evolução de algum fenômeno ou experimento no tempo.

### 2.1.2.2 Cadeia de Markov

Seja  $\{X(t), t \in \tau\}$  um processo estocástico, de tempo e espaço de estados discretos, isto é,  $S = \{1, 2, \dots, k\}$  e  $\tau = \{0, 1, \dots, T\}$ , satisfazendo a seguinte propriedade referente a probabilidade condicional:

$$\begin{aligned} P[X(t+1) = b \mid X(0) = x_0, X(1) = x_1, \dots, X(t-1) = x_{t-1}, X(t) = a] = \\ = P[X(t+1) = b \mid X(t) = a], \end{aligned} \quad (9)$$

para todo  $t \in \tau$  e  $a, b \in S$ . A propriedade (9) é chamada de propriedade markoviana de ordem 1 e, nesse caso, o processo estocástico é denominado cadeia de Markov. Portanto, em uma cadeia de Markov, tem-se que a probabilidade de ocorrência de um evento futuro  $X(t+1)$  condicionada ao passado (histórico) e ao presente do processo, é independente dos estados passados e depende somente do estado presente do processo,  $X(t)$ . Deve ser observado, que se o conhecimento do estado presente do processo é impreciso, então a probabilidade de um acontecimento futuro pode ser alterada pela informação adicional sobre o comportamento passado do processo.

As probabilidades condicionais, definidas pela propriedade de Markov (9), são também conhecidas como probabilidades de transição, pois na verdade, descrevem as mudanças ou transições dos estados do processo no tempo. Cumpre destacar que o termo transição está relacionado com o tempo, isto é, o que ocorre de uma ocasião para a outra, não implica necessariamente em mudança de estado. Nesse sentido, um indivíduo, avaliado no estado  $a$  na ocasião  $t$ , pode ter probabilidade não nula de permanecer nesse estado na



ocasião  $(t + 1)$  e, mesmo assim, considera-se que há transição, o que não há, é mudança de estado.

Essas probabilidades de transição, denotadas por  $P[X(t + 1) = b \mid X(t) = a] = \pi_{ab}(t, t + 1)$  satisfazem:

- i.  $\pi_{ab}(t, t + 1) \geq 0$
- ii.  $\sum_{b=1}^k \pi_{ab}(t, t + 1) = 1,$

para todo  $t \in T$  e  $a, b \in S$ . As probabilidades de transição podem ser representadas matricialmente:

$$\mathbf{P}(t, t + 1) = \begin{pmatrix} \pi_{11}(t, t + 1) & \pi_{12}(t, t + 1) & \dots & \pi_{1k}(t, t + 1) \\ \pi_{21}(t, t + 1) & \pi_{22}(t, t + 1) & \dots & \pi_{2k}(t, t + 1) \\ \vdots & \vdots & \dots & \vdots \\ \pi_{k1}(t, t + 1) & \pi_{k2}(t, t + 1) & \dots & \pi_{kk}(t, t + 1) \end{pmatrix}.$$

Quando as probabilidades de transição definidas pela matriz  $\mathbf{P}(t, t + 1)$  são independentes do tempo, diz-se que a cadeia de Markov tem probabilidades de transição estacionárias, ou ainda, que o processo é homogêneo em relação ao tempo. Nesse caso, a matriz de probabilidades de transição pode ser denotada por  $\mathbf{P}$ , com elementos  $\pi_{ab}$ , para todo  $a, b \in S$ .

Com relação aos estados de uma cadeia de Markov, eles podem ser classificados em absorventes, recorrentes ou transitórios. Quando para um estado  $a$  do processo, se verifica  $\pi_{ab}(t, t + 1) = 0 \forall a \neq b$  e  $\pi_{aa}(t, t + 1) = 1$ , então o estado  $a$  é denominado absorvente. Agora, considere  $\rho_{ab}$ , como a probabilidade de que uma cadeia iniciando no estado  $a$ , visite o estado  $b$  em algum instante positivo de tempo, um estado  $a$  é denominado recorrente se  $\rho_{aa} = 1$  e, transitório se  $\rho_{aa} < 1$ . Os estados absorventes são necessariamente recorrentes, a recíproca nem sempre é verdadeira. Um estado recorrente também pode ser classificado em positivo ou nulo. Um estado  $a$  é recorrente positivo, quando começando em  $a$ , espera-se que o tempo de retorno ao estado  $a$  seja finito. Se, porém, esse tempo de retorno for infinito o estado recorrente  $a$  é classificado como nulo. Nas cadeias finitas os estados recorrentes são positivos. Por outro lado, um estado  $a$  é dito aperiódico se seu período for igual a 1. Quando todos os estados de uma cadeia são recorrentes positivos aperiódicos, a cadeia é dita ergódica.

Adicionalmente, registra-se que, a caracterização estocástica completa de uma cadeia de Markov ocorre quando se conhece a distribuição inicial da cadeia e as probabilidades de transição. Desse modo, as probabilidades marginais num instante  $s$  são dadas por:

$$\boldsymbol{\pi}'_s = \begin{pmatrix} P[X(s) = 1] \\ P[X(s) = 2] \\ \vdots \\ P[X(s) = k] \end{pmatrix}' = \boldsymbol{\pi}'_0 P(0, s) = \boldsymbol{\pi}'_0 \prod_{t=1}^s \mathbf{P}(t-1, t), \quad (10)$$

em que  $\boldsymbol{\pi}'_0 = P[X(0) = a]_{a \in S}$  é o vetor de probabilidades iniciais e  $\boldsymbol{\pi}'_s = P[X(s) = a]_{a \in S}$  é a distribuição de  $X(s)$ . Se o processo é estacionário, então a equação (10) reduz-se a  $\boldsymbol{\pi}'_s = \boldsymbol{\pi}'_0 \mathbf{P}^s$ , em que  $\mathbf{P} = \mathbf{P}(t-1, t)$  para  $t \geq 1$ .

Represente, agora,  $P(s, t)$  a matriz de probabilidades de transição do tempo  $s$  para o tempo  $t$ , com  $t \geq s$ . Como decorrência da propriedade de Markov, essas probabilidades também satisfazem a equação de Chapman-Kolmogorov:

$$\pi_{ab}(s, u) = \sum_{k \in S} \pi_{ak}(s, t) \pi_{kb}(t, u),$$

para qualquer  $s \leq t \leq u$  e  $a, b \in S$ . Em termos matriciais, tem-se:

$$\mathbf{P}(s, u) = \mathbf{P}(s, t) \mathbf{P}(t, u).$$

Para um processo com probabilidades de transição estacionárias, a equação de Chapman-Kolmogorov permite expressar uma função de transição em  $(m+n)$  etapas:

$$\pi_{ab}^{(m+n)} = \sum_{k \in S} \pi_{ak}^{(m)} \pi_{kb}^{(n)}. \quad (11)$$

A equação (11) tem a seguinte interpretação:  $\pi_{ab}^{(m+n)}$  é a probabilidade de que uma cadeia que está inicialmente no estado  $a$  visite o estado  $b$  após  $(m+n)$  etapas. Para isso, é necessário que a partir do estado  $a$  essa cadeia visite um estado  $k$  após  $m$  etapas e a partir deste estado  $k$ , visite o estado  $b$  após  $n$  etapas adicionais. Essas probabilidades  $\pi_{ab}^{(m+n)}$  podem ser obtidas por meio de produto de matrizes, ou seja,  $\mathbf{P}^{(m+n)} = \mathbf{P}^{(m)} \mathbf{P}^{(n)}$ .

### 2.1.2.3 Distribuição estacionária ou de equilíbrio

Uma característica desejável em um processo estocástico é a de que ele admita uma distribuição estacionária, ou seja, de que ele se desenvolva no tempo em equilíbrio.

Intuitivamente, um processo é estacionário quando a escolha de uma origem dos tempos não é importante, isto é, as características de  $X(t + \Delta)$  são as mesmas de  $X(t)$  para todo  $\Delta \in \tau$ . Uma definição mais formal é dada a seguir.

Seja  $\{X(t), t \in \tau\}$  uma cadeia de Markov com espaço de estados  $S$  e matriz de transição  $\mathbf{P}$ . Se para cada  $a \in S$ , com  $\pi_a \geq 0$  e  $\sum_{a=1}^k \pi_a = 1$  e se:

$$\pi_b = \sum_{a=1}^k \pi_a \pi_{ab} \quad \forall b \in S, \quad (12)$$

então  $\boldsymbol{\pi} = (\pi_b)_{b \in S}$  é denominada distribuição estacionária da cadeia de Markov. Sob a forma matricial, tem-se  $\boldsymbol{\pi} \mathbf{P} = \boldsymbol{\pi}$ .

Admita que  $\boldsymbol{\pi}$  é a distribuição estacionária de uma cadeia de Markov e, suponha que  $\lim_{t \rightarrow \infty} \pi_{ab}(t) = \pi_b, \forall b \in S$ . Se  $\boldsymbol{\pi}_0$  é a distribuição inicial da cadeia, então  $\lim_{t \rightarrow \infty} P(X(t) = b) = \pi_b \forall b \in S$ . Isso significa que independente da distribuição inicial, a cadeia de Markov  $\{X(t), t \in \tau\}$  tem distribuição  $\boldsymbol{\pi}$ , quando  $t \rightarrow \infty$ . Desse modo, quando um sistema tem essa característica significa que ele tem comportamento estatisticamente estável.

#### 2.1.2.4 Inferências relacionadas à cadeia de Markov

Em muitas aplicações, sobretudo quando se pretende especificar um modelo matemático para o processo em estudo, questões relativas à estimação dos parâmetros, à condição de estacionariedade da cadeia e outras hipóteses envolvendo as probabilidades de transição são de importância singular. Há vários trabalhos que tratam de problemas de inferência relacionados a uma cadeia de Markov. Dentre eles, Bartlett (1951), Good (1955) e Anderson e Goodman (1957) são referências clássicas que concentram resultados de estimação e alguns testes de hipóteses necessários na prática. Registra-se que esses resultados são encontrados também em textos mais atuais, como em Crowder e Hand (1990), no trabalho de Noleto (1991) aplicado ao estudo das transições com dados categorizados ordinais, em Bhat e Miller (2002), Lindsey (2004), entre outros. Apresentam-se a seguir, os principais resultados desses trabalhos.

### 2.1.2.4.1 Estimação das probabilidades

Considere uma cadeia de Markov com espaço de estados,  $S = \{1, 2, \dots, k\}$  e sejam  $t = 0, 1, \dots, T$  os tempos de observação do processo. Assume-se que existam  $n_a(0)$  unidades no estado  $a$  em  $t = 0$ , com probabilidades de ocorrência  $\pi_a$ , para todo  $a \in S$ . Dada uma amostra aleatória,  $N = \sum_{a=1}^k n_a(0)$ , pode-se pensar no vetor  $\mathbf{n}(0) = (n_1(0), \dots, n_k(0))'$ , como um conjunto de variáveis aleatórias,  $n_a(0)$ , com distribuição multinomial, ou seja,  $\mathbf{n}(0) \sim M(N, \pi_1, \dots, \pi_k)$ , que tem função de probabilidade dada por:

$$f(\mathbf{n}(0)) = \frac{N!}{\prod_{a=1}^k n_a(0)!} \prod_{a=1}^k \pi_a^{n_a(0)}.$$

Seja, agora  $n_{ab}(t)$  o número de unidades no estado  $a$  em  $(t-1)$  e em  $b$  no tempo  $t$ , pode-se observar que:

$$n_{.b}(t) = \sum_{a=1}^k n_{ab}(t) \quad e \quad n_{a.}(t) = \sum_{b=1}^k n_{ab}(t) = n_a(t-1).$$

Novamente, fixado um estado  $a$ , pode-se mostrar que o vetor de variáveis aleatórias,  $\mathbf{n}_a(t) = (n_{a1}(t), \dots, n_{ak}(t))$ , com probabilidades  $\pi_{ab}(t-1, t)$ , tem distribuição multinomial com parâmetros  $n_a(t)$  e  $\pi_{a1}(t-1, t), \dots, \pi_{ak}(t-1, t)$ , ou seja:

$$f(\mathbf{n}_a(t)) = \frac{n_a(t-1)!}{\prod_{b=1}^k n_{ab}(t)!} \prod_{b=1}^k \pi_{ab}(t-1, t)^{n_{ab}(t)}.$$

Dessa forma, se todas as transições são independentes, então a função de verossimilhança de  $\pi_a$  e  $\pi_{ab}(t)$  pode ser expressa por:

$$L \propto \prod_{a=1}^k \pi_a^{n_a(0)} \prod_{t=1}^T \prod_{a=1}^k \prod_{b=1}^k \pi_{ab}(t-1, t)^{n_{ab}(t)},$$

a qual leva às estimativas de máxima verossimilhança para os parâmetros:

$$\hat{\pi}_a = \frac{n_a(0)}{N} \quad e \quad \hat{\pi}_{ab}(t-1, t) = \frac{n_{ab}(t)}{n_a(t-1)}.$$

Quando o processo é estacionário, tem-se que o conjunto formado pelos elementos  $n_{ab} = \sum_{t=1}^T n_{ab}(t)$  são estatísticas suficientes para as probabilidades de transição e, nesse caso, são estimadas por:

$$\hat{\pi}_{ab} = \frac{n_{ab}}{n_a} = \frac{\sum_{t=1}^T n_{ab}(t)}{\sum_{t=1}^T n_a(t-1)}.$$

### 2.1.2.4.2 Teoria assintótica

Anderson e Goodman (1957) apresentam alguns resultados sobre a distribuição assintótica do estimador de máxima verossimilhança  $\hat{\pi}_{ab}$ . Esses resultados são importantes para a construção de testes de interesse relacionados aos modelos de Markov.

Apresenta-se, a princípio, algumas considerações sobre as variáveis que descrevem as frequências de transições,  $n_{ab}(t)$ . Para cada estado no tempo  $t = 0$ ,  $a(0)$ , o conjunto  $n_{a(0)a(1), \dots, a(T)}$  é constituído de variáveis multinomiais com tamanho amostral  $n_{a(0)}(0)$  e parâmetros  $\pi_{a(0)a(1)}, \pi_{a(1)a(2)}, \dots, \pi_{a(T-1)a(T)}$ . Assim, quando o tamanho da amostra aumenta, essas variáveis tendem a uma distribuição normal. Como  $n_{ab}(t)$  são combinações lineares dessas variáveis, assintoticamente também têm distribuição normal. Seja agora,  $\pi_{ab}^{[t]}$  os elementos da matriz de transição  $P^{[t]}$ , isto é, a probabilidade do estado  $b$  no tempo  $t$  dado o estado  $a$  no tempo 0. Seja também,  $n_{j,ab}(t)$  o número de seqüências incluindo o estado  $j$  no tempo 0,  $a$  no tempo  $(t - 1)$  e  $b$  no tempo  $t$ . Então, deseja-se obter os momentos de:

$$n_{ab}(t) = \sum_{j=1}^k n_{j,ab}(t).$$

A probabilidade associada a  $n_{j,ab}(t)$  é  $\pi_{ja}^{[t-1]}\pi_{ab}$  com tamanho amostral  $n_j(0)$ . Portanto, se  $n_{j,ab}(t)$  tem distribuição multinomial, tem-se:

$$E[n_{j,ab}(t)] = n_j(0)\pi_{ja}^{[t-1]}\pi_{ab},$$

$$\text{Var}[n_{j,ab}(t)] = n_j(0)\pi_{ja}^{[t-1]}\pi_{ab}(1 - \pi_{ja}^{[t-1]}\pi_{ab}),$$

$$\text{Cov}[n_{j,ab}(t), n_{j,gh}(t)] = -n_j(0)\pi_{ja}^{[t-1]}\pi_{ab}\pi_{jg}^{[t-1]}\pi_{gh}, \quad (a, b) \neq (g, h).$$

Por outro lado, a distribuição condicional de  $n_{k,ab}(t)$  dado  $n_{j,a}(t - 1)$ , em que  $n_{j,a}(t - 1) = \sum_b n_{j,ab}(t)$  também é multinomial com probabilidades  $\pi_{ab}$ . Assim, podem-se especificar o primeiro e segundo momentos de  $n_{j,ab}(t) - n_{j,a}(t - 1)\pi_{ab}$ , haja vista que  $E[n_{j,ab}(t) | n_{j,a}(t - 1)] = \pi_{ab}n_{j,a}(t - 1)$ . Decorre:

$$E[n_{j,ab}(t) - n_{j,a}(t - 1)\pi_{ab}] = EE\{[n_{j,ab}(t) - n_{j,a}(t - 1)\pi_{ab}] | n_{j,a}(t - 1)\} = 0 \quad (13)$$

e

$$\begin{aligned}
\text{Var}[n_{j,ab}(t) - n_{j,a}(t-1)\pi_{ab}] &= \text{E}[n_{j,ab}(t) - n_{j,a}(t-1)\pi_{ab}]^2 \\
&= \text{EE}\{[n_{j,ab}(t) - n_{j,a}(t-1)\pi_{ab}]^2 \mid n_{j,a}(t-1)\} \\
&= \text{E}[n_{j,a}(t-1)\pi_{ab}(1 - \pi_{ab})] \\
&= n_j(0)\pi_{ja}^{[t-1]}\pi_{ab}(1 - \pi_{ab}). \tag{14}
\end{aligned}$$

Em síntese, as variáveis aleatórias  $n_{j,ab}(t) - n_{j,a}(t-1)\pi_{ab}$  para  $b = 1, 2, \dots, k$  têm média 0 e variância e covariância de variáveis multinomiais com probabilidades  $\pi_{ab}$  e tamanho amostral igual a  $n_j(0)\pi_{ab}^{[t-1]}$ . Com essas considerações preliminares, quando  $n \rightarrow \infty$ , pode-se mostrar que:

$$\begin{aligned}
\sqrt{n}(\hat{\pi}_{ab} - \pi_{ab}) &= \sqrt{n} \left[ \frac{\sum_{t=1}^T n_{ab}(t)}{\sum_{t=1}^T n_a(t-1)} - \pi_{ab} \right] \\
&= \sqrt{n} \left[ \frac{\sum_{t=1}^T [n_{ab}(t) - \pi_{ab}n_a(t-1)]}{\sum_{t=1}^T n_a(t-1)} \right] \\
&= \sqrt{n} \left[ \frac{\sum_{j=1}^k \sum_{t=1}^T [n_{j,ab}(t) - \pi_{ab}n_{j,a}(t-1)]}{\sum_{t=1}^T n_a(t-1)} \right],
\end{aligned}$$

tem uma distribuição normal como limite. Como  $n_{j,ab}(t)$  é uma variável aleatória multinomial, sabe-se que:

$$n_{j,ab}(t) \approx [n_{j,ab}(t)/n_j(0)]\pi_j$$

converge em probabilidade para seu valor esperado quando  $n_j(0)/n \rightarrow \pi_j$ . Desse modo:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^T n_a(t-1) = \lim_{n \rightarrow \infty} \frac{1}{n} \text{E} \left[ \sum_{t=1}^T n_a(t-1) \right] = \sum_{j=1}^k \pi_j \sum_{t=1}^T \pi_{ab}^{[t-1]}.$$

Portanto,  $\sqrt{n}(\hat{\pi}_{ab} - \pi_{ab})$  tem a distribuição limite como:

$$\frac{\sum_{t=1}^T [n_{ab}(t) - \pi_{ab}n_a(t-1)]/\sqrt{n}}{\sum_{j=1}^k \sum_{t=1}^T \pi_j \pi_{ab}^{[t-1]}}. \tag{15}$$

Usando os resultados das considerações iniciais dadas em (13) e (14), decorre que o numerador da expressão (15) tem média 0 e variância dada por:

$$\text{E} \left[ \sum_{t=1}^T n_{ab}(t) - \pi_{ab}n_a(t-1) \right]^2 / n = \sum_{j=1}^k \sum_{t=1}^T n_j(0)\pi_{ja}^{[t-1]}\pi_{ab}(1 - \pi_{ab})/n. \tag{16}$$

Considerando  $\sum_{j=1}^k \sum_{t=1}^T \pi_j \pi_{ab}^{[t-1]} = \varsigma_a$ , então a variância limite do numerador de (15) é  $\varsigma_a \pi_{ab}(1 - \pi_{ab})$ . Ademais, como essa expressão é combinação linear de variáveis aleatórias multinomias padronizadas com probabilidades fixas ela tem como limite uma distribuição normal. Portanto, para  $n \rightarrow \infty$ ,  $\sqrt{n}(\hat{\pi}_{ab} - \pi_{ab})$  tem distribuição normal, com média 0 e variância  $\pi_{ab}(1 - \pi_{ab})/\varsigma_a$ , ou ainda:

$$\sqrt{n\varsigma_a}(\hat{\pi}_{ab} - \pi_{ab}) \rightarrow N(0, \pi_{ab}(1 - \pi_{ab})). \quad (17)$$

### 2.1.2.4.3 Teste para a hipótese de homogeneidade

A hipótese de que as  $m$  matrizes de transição são iguais pode ser verificada através de um teste de homogeneidade. A idéia que está por trás desse teste é análoga àquela que se tem acerca das distribuições de freqüências em tabelas de contingência. Nesse contexto, considere  $m$  amostras de uma cadeia de Markov, com freqüências de transições  $n_{ab}(\ell)$  e, probabilidades de transição dadas por  $\pi_{ab}(\ell) = \frac{n_{ab}(\ell)}{n_{a.}(\ell)}$ ,  $\ell = 1, 2, \dots, m$ . A hipótese de interesse pode ser formulada por:

$$H_0 : \pi_{ab}(1) = \pi_{ab}(2) = \dots \pi_{ab}(m) = \pi_{ab}, \quad \forall a, b \in S.$$

Sob  $H_0$ , o estimador de máxima verossimilhança para as probabilidades de transição comum a todas as amostras é dado por:

$$\hat{\pi}_{ab} = \frac{\sum_{\ell=1}^m n_{ab}(\ell)}{\sum_{\ell=1}^m n_{a.}(\ell)}.$$

As seguintes estatísticas podem ser usadas para se testar  $H_0$ :

$$\xi = \sum_{\ell=1}^m \sum_{a=1}^k \sum_{b=1}^k \frac{[n_{ab}(\ell) - n_{a.}(\ell)\hat{\pi}_{ab}]^2}{n_{a.}(\ell)\hat{\pi}_{ab}} \quad (18)$$

e

$$\lambda = -2 \sum_{\ell=1}^m \sum_{a=1}^k \sum_{b=1}^k n_{ab}(\ell) \log \left( \frac{n_{a.}(\ell)\hat{\pi}_{ab}}{n_{ab}(\ell)} \right), \quad (19)$$

sendo, que sob  $H_0$ , são equivalentes e apresentam distribuição  $\chi^2$ , com  $(m-1)k(k-1)$  graus de liberdade.

#### 2.1.2.4.4 Teste para a hipótese de estacionariedade

A condição de estacionariedade é um dos aspectos desejados em um processo pois, na prática, implica em simplificação em termos de descrição das probabilidades de transição e, conseqüentemente do número de parâmetros a serem estimados. Por exemplo, suponha uma cadeia de Markov, com  $t$  transições (do tempo 1 para o tempo 2, do tempo 2 para o 3 e, assim, sucessivamente até a  $t$ -ésima transição). Se o processo é estacionário, a probabilidade de um indivíduo passar do estado  $a$  para o estado  $b$  é  $\pi_{ab}$ , independente da transição de tempo. No entanto, essa condição nem sempre é satisfeita, resultando na dependência das probabilidades de transição em relação ao tempo,  $\pi_{ab}(t)$ . Sendo assim, em um processo estocástico é de interesse testar se há, ou não, estacionariedade. A hipótese principal é,  $H_0 : \pi_{ab}(t) = \pi_{ab}$ , para todo  $t \in \tau$ . Esse teste, para um estado fixo  $a$  é equivalente ao teste de homogeneidade para uma tabela de contingência  $T \times k$ . De fato, fixado  $a$ , as estimativas  $\hat{\pi}_{ab}(t) = \frac{n_{ab}(t)}{n_a(t-1)}$ , para  $b = 1, 2, \dots, k$  e  $t = 1, 2, \dots, T$  podem ser representadas pela Tabela 1.

Tabela 1 - Tabela de contingência  $T \times k$  para a apresentação das probabilidades de transição, para um estado fixo  $a$

		b			
t	1	2	...	k	
1	$\hat{\pi}_{a1}(1)$	$\hat{\pi}_{a2}(1)$	...	$\hat{\pi}_{ak}(1)$	
2	$\hat{\pi}_{a1}(2)$	$\hat{\pi}_{a2}(2)$	...	$\hat{\pi}_{ak}(2)$	
⋮	⋮	⋮	⋮	⋮	
T	$\hat{\pi}_{a1}(T)$	$\hat{\pi}_{a2}(T)$	...	$\hat{\pi}_{ak}(T)$	

Desse modo, sob  $H_0$ , as variáveis aleatórias representadas pelas  $T$  linhas da Tabela 1 têm a mesma distribuição, isto é, a probabilidade de transição para cada uma das  $k$  categorias é igual em todas as linhas. A hipótese pode ser testada através da estatística:

$$\xi_a = \sum_{t=1}^T \sum_{b=1}^k \frac{n_a(t-1)[\hat{\pi}_{ab}(t) - \hat{\pi}_{ab}]^2}{\hat{\pi}_{ab}}, \quad (20)$$

sendo que  $\xi_a$  tem assintoticamente distribuição  $\chi_{(k-1)(T-1)}^2$ .



Note que para o processo ser estacionário, o teste descrito pela equação (20) deve ser não significativo para todo  $a = 1, 2, \dots, k$ . Como para dois estados distintos de  $a$ , as variáveis aleatórias  $\pi_{ab}(t)$  e  $\pi_{ab}$  são assintoticamente independentes, decorre que a estatística para a hipótese global,  $H_0 : \pi_{ab}(t) = \pi_{ab}, \forall a, b \in S$  e  $t \in T$  é dada por:

$$\xi = \sum_{a=1}^k \xi_a = \sum_{t=1}^T \sum_{a=1}^k \sum_{b=1}^k \frac{n_a(t-1)[\hat{\pi}_{ab}(t) - \hat{\pi}_{ab}]^2}{\hat{\pi}_{ab}}, \quad (21)$$

que tem assintoticamente distribuição  $\chi_{k(k-1)(T-1)}^2$ .

Pode-se, ainda, usar o teste da razão de verossimilhanças. As funções de verossimilhança maximizadas, sob as hipóteses nula e alterantiva, são:

$$L(\hat{\mathbf{P}}) = \prod_{t=1}^T \prod_{a,b} \hat{\pi}_{ab}^{n_{ab}(t)}$$

e

$$L(\hat{\mathbf{P}}(t)) = \prod_{t=1}^T \prod_{a,b} \hat{\pi}_{ab}(t)^{n_{ab}(t)},$$

respectivamente. Portanto:

$$\Lambda = \frac{L(\hat{\mathbf{P}})}{L(\hat{\mathbf{P}}(t))} = \prod_{t=1}^T \prod_{a,b} \left[ \frac{\hat{\pi}_{ab}}{\hat{\pi}_{ab}(t)} \right]^{n_{ab}(t)}.$$

Ademais, sob  $H_0$  verdadeira, como extensão do teorema de Neyman-Pearson, a estatística  $-2 \log \Lambda$  tem a seguinte distribuição:

$$\lambda = -2 \sum_{t=1}^T \sum_{a=1}^k \sum_{b=1}^k n_{ab}(t) \log \left( \frac{\hat{\pi}_{ab}}{\hat{\pi}_{ab}(t)} \right) \sim \chi_{k(k-1)(T-1)}^2, \quad (22)$$

sendo o teste (22) equivalente ao teste dado por (21).

#### 2.1.2.4.5 Teste para a hipótese de uma matriz de transição específica

Seja  $\{X(t), t \in \tau\}$  uma cadeia de Markov de probabilidades de transição estacionárias (processo homogêneo no tempo) e considere a hipótese de que as probabilidades de transição assumem valores específicos, ou seja,  $H_0 : \pi_{ab} = \pi_{ab}^0, a, b \in S$ . Na notação matricial:

$$H_0 : \mathbf{P} = \mathbf{P}^0. \quad (23)$$

Sob essa hipótese, a estatística  $(n_a)^{1/2}(\hat{\pi}_{ab} - \pi_{ab}^0)$  tem uma distribuição normal assintótica com média 0 e variância  $\pi_{ab}(1 - \pi_{ab})$ . Com base nessa informação, para um estado prévio fixo  $a$ , a estatística de ajuste:

$$\xi_a = \sum_{b=1}^k \frac{n_a(\hat{\pi}_{ab} - \pi_{ab}^0)^2}{\pi_{ab}^0},$$

tem distribuição  $\chi^2$  com  $k - 1$  graus de liberdade. Conseqüentemente, a estatística:

$$\xi = \sum_{a=1}^k \sum_{b=1}^k \frac{n_a(\hat{\pi}_{ab} - \pi_{ab}^0)^2}{\pi_{ab}^0} \sim \chi_{k(k-1)}^2,$$

corresponde a uma soma de  $k$  estatísticas qui-quadrado independentes e pode ser usada para testar a hipótese dada em (23). Outro teste assintótico equivalente é o da razão de verossimilhanças baseado no teorema de Neyman-Pearson:

$$\Lambda = \frac{L(\mathbf{P}^0)}{L(\hat{\mathbf{P}})},$$

em que  $L(\hat{\mathbf{P}})$  é o valor da função de verossimilhança maximizada obtida a partir da substituição pelas estimativas  $\hat{\pi}_{ab}$ . Como se sabe, quando  $H_0$  é verdadeira,  $-2 \log \Lambda$  tem distribuição  $\chi^2$  com  $k(k - 1)$  graus de liberdade. Aplicando esse critério:

$$\lambda = -2 \log \Lambda = 2 \log[L(\hat{\mathbf{P}}) - L(\mathbf{P}^0)] = 2 \sum_{a=1}^k \sum_{b=1}^k n_{ab} \log \left( \frac{n_{ab}}{n_a \pi_{ab}^0} \right).$$

#### 2.1.2.4.6 Cadeias de Markov de alcance maior do que 1

Há situações em que a ordem da cadeia pode ser maior do que 1, de tal forma que a probabilidade do estado do processo no tempo  $t$  depende dos  $q$  tempos anteriores e não somente de  $(t - 1)$ . Nesses casos, o número de estados prévios necessários ( $q$ ) para se estimarem as probabilidades de transição é referido como a ordem ou alcance da cadeia de Markov. A propriedade markoviana, para uma cadeia de ordem  $q$  é definida pela expressão:

$$\begin{aligned} & P[X(t) = a_{q+1} \mid X(t-1) = a_q, \dots, X(t-q) = a_1, \dots, X(1) = x_1, X(0) = x_0] = \\ & = P[X(t) = a_{q+1} \mid X(t-1) = a_q, \dots, X(t-q) = a_1]. \end{aligned} \quad (24)$$

Por exemplo, uma cadeia de Markov de ordem 2, com  $S = \{0, 1\}$ , a propriedade de Markov estabelece que a caracterização completa do estado do sistema num dado tempo

$t$  necessita da informação dos estados do sistema nos dois tempos precedentes, assim:

$$\pi_{abc}(t) = P(X_t = c \mid X_{(t-1)} = b, X_{(t-2)} = a) \quad (a, b, c \in S).$$

Logo, o processo é caracterizado por matrizes de probabilidades de transição de dimensões  $4 \times 2$ :

$$\mathbf{P}(t) = \begin{pmatrix} \pi_{000}(t) & \pi_{001}(t) \\ \pi_{010}(t) & \pi_{011}(t) \\ \pi_{100}(t) & \pi_{101}(t) \\ \pi_{110}(t) & \pi_{111}(t) \end{pmatrix}, \quad (25)$$

para  $t = 2, 3, 4, \dots, T$ .

Uma generalização para o caso de  $k$  categorias de resposta também é possível. Registra-se que a análise de cadeias de Markov de alcance superior a um é mais complexa, sobretudo com  $k$  categorias de resposta. Sendo assim, é importante verificar formalmente o alcance de uma cadeia. Pode-se usar o teste de qui-quadrado ou da razão de verossimilhanças, sendo assintoticamente equivalentes.

Nesse contexto, considere uma cadeia de Markov de segunda ordem, em que  $S = \{1, 2, \dots, k\}$  com probabilidades de transição:

$$\pi_{abc}(t) = P(X_t = c \mid X_{(t-1)} = b, X_{(t-2)} = a),$$

com  $a, b, c \in S$ , e  $t = 2, 3, \dots, T$ . Represente  $n_{abc}(t)$  o número de indivíduos no estado  $c$  na ocasião  $t$ , com estados prévios  $b$  em  $t - 1$  e  $a$  em  $t - 2$  e seja também  $n_{ab}(t - 1) = \sum_c n_{abc}(t)$ . Para o caso estacionário, Anderson e Goodman (1957) mostram que  $n_{abc} = \sum_{t=2}^T n_{abc}(t)$  formam um conjunto de estatísticas suficientes para as probabilidades de transição, cujo estimador de máxima verossimilhança é expresso por:

$$\hat{\pi}_{abc} = \frac{n_{abc}}{\sum_{\ell=1}^k n_{abl}} = \frac{\sum_{t=2}^T n_{abc}(t)}{\sum_{t=2}^T n_{ab}(t-1)}.$$

Então, quando a cadeia é estacionária,  $\pi_{abc}(t) = \pi_{abc}$ , uma cadeia de primeira ordem é um caso especial de uma cadeia de segunda ordem, em que as probabilidades de transição  $\pi_{abc}(t)$  independem do estado  $a$ . A hipótese nula de interesse, nesse caso, é:

$$H_0 : \pi_{abc} = \pi_{bc} \quad (a, b, c \in S), \quad (26)$$

contra a alternativa de que a cadeia é de ordem 2.

Para um estado fixo  $b$ , é possível construir uma tabela de contingência  $k \times k$ , que pode ser usada para representar as estimativas  $\hat{\pi}_{abc}$ ,  $a, c \in S$ . Assim, um teste  $\chi^2$  de homogeneidade para  $\pi_{abc} = \pi_{bc}$ ,  $a = 1, 2, \dots, k$  é apropriado. A estatística do teste é:

$$\xi_b = \frac{\sum_{a,c} n_{ab}^* (\hat{\pi}_{abc} - \hat{\pi}_{bc})^2}{\hat{\pi}_{bc}} \sim \chi_{(k-1)^2}^2,$$

em que  $n_{ab}^* = \sum_c n_{abc} = \sum_{t=2}^T n_{ab}(t-1)$  e  $\hat{\pi}_{bc} = \frac{\sum_{a=1}^k n_{abc}}{\sum_{a=1}^k \sum_{\ell=1}^k n_{abl}}$ . Logo, a estatística para a hipótese global (26) pode ser obtida por meio da soma:

$$\xi = \sum_b \xi_b = \frac{\sum_{a,b,c} n_{ab}^* (\hat{\pi}_{abc} - \hat{\pi}_{bc})^2}{\hat{\pi}_{bc}}, \quad (27)$$

que tem distribuição  $\chi^2$  com  $k(k-1)^2$  graus de liberdade.

Alternativamente, usando o critério da razão de verossimilhanças, tem-se:

$$\Lambda = \prod_{a,b,c} \left( \frac{\hat{\pi}_{bc}}{\hat{\pi}_{abc}} \right)^{n_{abc}},$$

ou, ainda:

$$\lambda = -2 \log \Lambda = 2 \sum_{abc} n_{abc} [\log \hat{\pi}_{abc} - \log \hat{\pi}_{bc}]. \quad (28)$$

Esse procedimento pode ser estendido para testar a ordem  $q$  de uma cadeia. Nesse caso, a hipótese nula é de que a cadeia de Markov é de  $(q-1)$ -ésima ordem contra a alternativa de que a cadeia de Markov é de  $q$ -ésima ordem. Particularmente, quando  $q = 1$ , testam-se as hipóteses de que as observações são independentes contra a alternativa de que o processo é uma cadeia de Markov de primeira ordem. Para o caso mais geral,  $q > 2$ , as expressões (27) e (28) podem ser generalizadas como se segue:

$$\xi = \frac{\sum_{a,b,\dots,\ell,m} n_{ab,\dots,\ell}^* (\hat{\pi}_{ab,\dots,\ell,m} - \hat{\pi}_{b,\dots,\ell,m})^2}{\hat{\pi}_{b,\dots,\ell,m}}$$

e

$$\lambda = -2 \log \Lambda = 2 \sum_{ab,\dots,\ell,m} n_{ab,\dots,\ell,m} [\log \hat{\pi}_{ab,\dots,\ell,m} - \log \hat{\pi}_{b,\dots,\ell,m}],$$

respectivamente, em que  $n_{ab,\dots,\ell}^* = \sum_m n_{ab,\dots,\ell,m} = \sum_{t=q-1}^{T-1} n_{ab,\dots,\ell}(t)$ ,  $\hat{\pi}_{ab,\dots,\ell,m} = \frac{n_{ab,\dots,\ell,m}}{n_{ab,\dots,\ell}^*}$  e

$\hat{\pi}_{b,\dots,\ell,m} = \frac{\sum_{a=1}^k n_{ab,\dots,\ell,m}}{\sum_{a=1}^k n_{ab,\dots,\ell}^*}$ . A correspondente estatística  $\chi^2$  tem  $k^{q-1}(k-1)^2$  graus de liberdade.

### 2.1.3 Modelos para análise de dados longitudinais com resposta diferente da normal

O modelo logístico para dados binários, observados ao longo do tempo, foi um dos primeiros modelos propostos para análise de dados longitudinais com resposta diferente da normal (COX, 1970; STIRATELLI; LAIRD; WARE, 1984; ANDERSON; AITKIN, 1985; LIANG; ZEGER; SELF, 1985, outros). Nesta seção são apresentados alguns procedimentos para o ajuste desses modelos. Embora seja dada uma ênfase para o caso binário, sempre que possível os resultados são estendidos aos demais tipos de variáveis respostas.

#### 2.1.3.1 Modelos lineares generalizados marginais para dados binários

Independentemente do tipo de variável resposta, modelos marginais podem ser ajustados, considerando uma amostra aleatória em cada uma das ocasiões. Assim, se  $\mathbf{y}_t = (y_{1t}, y_{2t}, \dots, y_{Nt})'$  representa uma amostra de  $N$  observações no tempo  $t$  em que a cada observação está associado o vetor de variáveis explicativas  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$ , em que  $i = 1, 2, \dots, N$ , um MLG pode ser ajustado para cada ocasião, seguindo os procedimentos usuais como descrito na seção (2.1.1). Para o caso de variáveis respostas binárias, pode-se usar a regressão logística múltipla:

$$\ln \left[ \frac{\pi_t(\mathbf{x})}{1 - \pi_t(\mathbf{x})} \right] = \beta_{0t} + \beta_{1t}x_{1t} + \beta_{2t}x_{2t} + \dots + \beta_{pt}x_{pt},$$

para avaliar a influência do efeito das variáveis explicativas nas probabilidades do sucesso em cada ocasião,  $\hat{\pi}_t(\mathbf{x})$ .

Esse processo pode ser visto como um caso particular da proposta de Stram, Wei e Ware (1988) para modelos marginais com respostas ordinais, que por sua vez, é uma adaptação da técnica descrita em McCullagh (1980), com a utilização do modelo de chances proporcionais (“proportional odds”). Esses modelos são baseados em probabilidades acumuladas e são extensões multivariadas dos MLGs.

Assim, considere uma amostra aleatória de  $N$  indivíduos, para os quais a variável resposta observada nas ocasiões ( $t = 0, 1, \dots, T$ ) é do tipo ordinal, isto é, pode assumir valores no conjunto  $S = \{1, 2, \dots, k\}$ , tais que,  $1 < 2 < \dots < k$ . A resposta para o  $i$ -ésimo indivíduo na  $t$ -ésima ocasião é denotada pelo vetor  $\mathbf{y}_{it} = (I_{i1t}, I_{i2t}, \dots, I_{ikt})'$ , em

que  $I_{ijt}$  representam variáveis indicadoras para as categorias de resposta, isto é  $I_{ijt} = 1$  se o indivíduo está na  $j$ -ésima categoria no tempo  $t$  e  $I_{ijt} = 0$ , caso contrário. Seja também  $\mathbf{x}_{it} = (x_{1it}, x_{2it}, \dots, x_{pit})'$  o vetor das  $p$  variáveis explicativas associadas ao indivíduo  $i$  no tempo  $t$ ,  $\pi_{jt}(\mathbf{x})$ ,  $j = 1, 2, \dots, k$ , a probabilidade marginal de ocorrência de  $I_{ijt}$  e, represente:

$$\gamma_{1t}(\mathbf{x}) = \pi_{1t}(\mathbf{x}), \gamma_{2t}(\mathbf{x}) = \pi_{1t}(\mathbf{x}) + \pi_{2t}(\mathbf{x}), \dots, \gamma_{k-1,t}(\mathbf{x}) = \pi_{1t}(\mathbf{x}) + \dots + \pi_{k-1,t}(\mathbf{x}), \gamma_{kt}(\mathbf{x}) = 1,$$

as probabilidades acumuladas.

A proposta dos autores é ajustar um modelo para cada ocasião, de tal forma a obter  $\hat{\boldsymbol{\beta}}_t = (\hat{\beta}_{1t}, \hat{\beta}_{2t}, \dots, \hat{\beta}_{pt})$ , o vetor de estimativas do vetor  $\boldsymbol{\beta}_t$  de parâmetros desconhecidos referentes aos efeitos das variáveis explicativas nas estimativas das probabilidades  $\hat{\pi}_{jt}(\mathbf{x})$ . Isso é possível maximizando a função de verossimilhança para cada ocasião em separado, utilizando o modelo de chances proporcionais:

$$\ln \left[ \frac{\gamma_{jt}(\mathbf{x})}{1 - \gamma_{jt}(\mathbf{x})} \right] = \lambda_{jt} - (\beta_{1t}x_{1t} + \beta_{2t}x_{2t} + \dots + \beta_{pt}x_{pt}) = \lambda_{jt} - \boldsymbol{\beta}'_t \mathbf{x}, \quad (29)$$

em que  $\lambda_{jt}$  é um parâmetro de perturbação do modelo, denominado intercepto e  $\mathbf{x}$  representa o vetor de valores das variáveis explicativas, o qual também pode variar no tempo. Quando a resposta é binária, o modelo (29) reduz-se ao modelo de regressão logística para a probabilidade da primeira resposta. Assim, para uma amostra aleatória de  $N$  elementos, o logaritmo da função de verossimilhança no tempo  $t$ , pode ser expresso por:

$$\ell_t \propto \sum_{i=1}^N \delta_{ti} \sum_{j=1}^k I_{ijt} \{ \ln[\phi(\lambda_{jt} - \mathbf{x}'_{it} \boldsymbol{\beta}_t) - \phi(\lambda_{j-1,t} - \mathbf{x}'_{it} \boldsymbol{\beta}_t)] \}, \quad (30)$$

em que  $\phi(\cdot) = \frac{e(\cdot)}{1+e(\cdot)}$  e  $\delta_{ti}$ , é uma variável indicadora incluída no modelo para o caso de observações omissas, isto é,  $\delta_{ti} = 1$  se  $\mathbf{y}_{it}$  e  $\mathbf{x}_{it}$  são observados e 0 em caso contrário.

Como é clássico nos MLGs o máximo global de (30) somente é possível ser obtido via processo iterativo. O método de Newton-Raphson pode ser implementado para obtenção das estimativas  $\hat{\boldsymbol{\beta}}_t$  e  $\hat{\lambda}_{jt}$ . Uma vez obtidas as estimativas dos parâmetros do modelo (29), podem-se estimar as probabilidades acumuladas por meio de:

$$\hat{\gamma}_{jt}(\mathbf{x}) = \frac{\exp(\hat{\lambda}_{jt} - \hat{\boldsymbol{\beta}}'_t \mathbf{x})}{1 + \exp(\hat{\lambda}_{jt} - \hat{\boldsymbol{\beta}}'_t \mathbf{x})} \quad j = 1, 2, \dots, k-1,$$

e, para se calcularem as probabilidades pontuais para cada um dos tempos de observação, basta trabalhar com a função de distribuição acumulada. Por exemplo, a probabilidade marginal de um indivíduo estar na categoria 3, no tempo  $t$  é calculada por:

$$\hat{\pi}_{3t}(\mathbf{x}) = \hat{\gamma}_{3t}(\mathbf{x}) - \hat{\gamma}_{2t}(\mathbf{x}).$$

Vale ressaltar que os modelos marginais ajustados em separado conduzem a estimativas de parâmetros que são específicos para cada ocasião. Somente a dependência de  $Y$  sobre  $\mathbf{x}$  é especificada, como num estudo transversal. Uma das críticas a esse tipo de abordagem está no fato, de que, em geral, o estudo longitudinal caracteriza-se pela presença de correlação entre as observações repetidas e essa correlação deve ser levada em conta no processo de ajuste. Diggle et al. (2002) apontam três conseqüências ao se ignorar a correlação quando ela existe em dados longitudinais:

1. inferências incorretas sobre os coeficientes de regressão,  $\beta$ , pois, em geral, os erros-padrão são os mais afetados;
2. estimativas de  $\beta$  que são ineficientes, isto é, menos precisas,
3. proteção sub-ótima contra vícios causados por dados perdidos.

Para contornar essas limitações, Liang e Zeger (1986) e Zeger e Liang (1986) propõem as equações de estimação generalizadas (EEG), necessitando apenas da especificação do primeiro e do segundo momentos da função de densidade marginal. A metodologia pode ser vista como uma extensão da teoria da quase-verossimilhança e permite a análise de dados longitudinais sejam eles discretos ou contínuos.

### 2.1.3.2 Equações de estimação generalizadas e modelos marginais

A fim de estabelecer a notação a ser usada nesta seção, considere  $\mathbf{y}_i$ , o vetor de variáveis respostas, de dimensões  $n_i \times 1$ , isto é,  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})'$ , e a matriz de variáveis explicativas  $\mathbf{X}_i = (x_{i1}, x_{i2}, \dots, x_{in_i})'$ , de dimensões  $n_i \times p$ , cujas linhas são os vetores  $\mathbf{x}_{it} = (x_{it1}, \dots, x_{itp})'$ , para a  $i$ -ésima unidade experimental ( $i = 1, 2, \dots, N$ ), nos  $t$ -ésimos instantes de observação (tempo),  $t = 1, 2, \dots, n_i$ . Assume-se, a princípio, que a

distribuição marginal da variável aleatória  $Y_{it}$  é conhecida e pertence à família definida em (1), ou seja:

$$f(y_{it}; \theta_{it}, \phi) = \exp \left\{ \frac{1}{a_i(\phi)} [y_{it}\theta_{it} - b(\theta_{it})] + c(y_{it}, \phi) \right\}, \quad (31)$$

na qual a inclusão do índice  $t$  é para considerar as observações repetidas. Assume-se também que  $E(Y_{it}) = \mu_{it} = b'(\theta_{it})$ ,  $\text{Var}(Y_{it}) = a_i(\phi)b''(\theta_{it}) = a_i(\phi)V(\mu_{it})$ , em que  $V(\mu_{it})$  é a função de variância. Também, em grande parte dos casos,  $a_i(\phi) = \phi$ . Pode-se, então, definir um modelo linear generalizado para cada instante de tempo  $t$ , acrescentando uma função que faz a ligação entre a parte aleatória (31) e a parte sistemática:

$$\eta_{it} = g(\mu_{it}),$$

em que  $\boldsymbol{\eta}_{it} = \mathbf{x}'_{it}\boldsymbol{\beta}$  é o preditor linear,  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$  é o vetor de parâmetros desconhecidos e de interesse e  $g(\cdot)$  é a função de ligação.

Uma primeira alternativa para modelagem e estimação dos parâmetros  $\boldsymbol{\beta}$ , seria ignorar a estrutura de correlação intra-unidade experimental, isto é, supor que as observações de um mesmo indivíduo são independentes. Sob esta suposição, a função escore é expressa por:

$$\mathbf{U}_I(\boldsymbol{\beta}) = \frac{1}{\phi} \sum_{i=1}^N \mathbf{D}'_i \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i), \quad (32)$$

em que  $\mathbf{D}_i = \mathbf{W}_i^{1/2} \mathbf{V}_i^{-1/2} \mathbf{X}_i$ ,  $\mathbf{X}_i$  é a matriz de variáveis explicativas;  $\mathbf{W}_i = \text{diag}[w_{it}]$  é a matriz dos pesos,  $w_{it} = \left( \frac{\partial \mu_{it}}{\partial \eta_{it}} \right)^2 / V(\mu_{it})$ ;  $\mathbf{V}_i = \text{diag}[V(\mu_{it})]$  e  $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \dots, \mu_{ini})'$ . Em caso de função de ligação canônica, a função escore se reduz a  $\mathbf{U}_I(\boldsymbol{\beta}) = \frac{1}{\phi} \sum_{i=1}^N \mathbf{X}'_i (\mathbf{y}_i - \boldsymbol{\mu}_i)$ . A estimativa  $\hat{\boldsymbol{\beta}}_I$  é obtida da solução do sistema de equações escore, isto é,

$$\mathbf{U}_I(\boldsymbol{\beta}) = \sum_{i=1}^N \mathbf{D}'_i \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}. \quad (33)$$

Independente do conhecimento da distribuição marginal de  $Y_{it}$  é possível especificar apenas a função de variância dos dados, o que caracteriza o modelo de quase-verossimilhança, e a função escore será igual à apresentada na equação (32).

Liang e Zeger (1986) mostram que  $\hat{\boldsymbol{\beta}}_I$  é um estimador consistente para  $\boldsymbol{\beta}$  e,  $\sqrt{N}(\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta})$  tem assintoticamente distribuição normal multivariada. A desvantagem desse



estimador é que ele pode não ter eficiência nos casos em que existe uma correlação grande entre as observações repetidas. Para solucionar esse problema, eles propõem as EEG, que levam em conta a correlação de tal forma a melhorar a eficiência dos estimadores, sem perder a consistência dos mesmos.

A idéia é introduzir uma matriz  $\mathbf{R}(\boldsymbol{\alpha})$ , simétrica e de dimensão  $n_i$ , que satisfaça as condições de uma matriz de correlação e, em que,  $\boldsymbol{\alpha}$  denota um vetor de dimensões  $s \times 1$ , que caracteriza completamente a matriz  $\mathbf{R}(\boldsymbol{\alpha})$ , chamada de “matriz de correlação de trabalho”. Assim, define-se:

$$\boldsymbol{\Omega}_i = \text{Var}(Y_i) = \phi \mathbf{V}_i^{1/2} \mathbf{R}(\boldsymbol{\alpha}) \mathbf{V}_i^{1/2}, \quad (34)$$

como a matriz de variâncias e covariâncias de  $Y_i$  se a verdadeira correlação entre as observações repetidas for dada por  $\mathbf{R}(\boldsymbol{\alpha})$ . Note que, sob a suposição de independência entre as observações, a matriz dada em (34) reduz-se a:

$$\boldsymbol{\Omega}_i = \text{Var}(Y_i) = \phi \mathbf{V}_i, \text{ pois } \mathbf{R}(\boldsymbol{\alpha}) = \mathbf{I}_{n_i}.$$

Admitindo-se uma outra estrutura para  $\boldsymbol{\Omega}_i$ , para estimar  $\boldsymbol{\beta}$  deve-se resolver o sistema:

$$\mathbf{U}_G(\boldsymbol{\beta}) = \mathbf{0},$$

ou seja,

$$\sum_{i=1}^N \mathbf{D}'_i \boldsymbol{\Omega}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \quad (35)$$

cujas equações são denominadas por Liang e Zeger (1986) de equações de estimação generalizadas. Em particular, sob suposição de independência a equação (35) reduz-se a (33), e para cada unidade amostral  $i$ , é similar a função score do modelo de quase-verossimilhança.

No entanto, as equações de estimação, dependem também dos parâmetros  $\phi$  e  $\boldsymbol{\alpha}$ , que devem ser estimados separadamente. Para implementar o processo iterativo, é preciso, portanto, dar estimativas iniciais para  $\phi$  e  $\boldsymbol{\alpha}$ , que são atualizadas separadamente a cada passo. O processo iterativo consiste em substituir  $\boldsymbol{\alpha}$  em (35) por  $\hat{\boldsymbol{\alpha}}$ , quando  $\boldsymbol{\beta}$  e  $\phi$  são conhecidos. A seguir, troca-se  $\phi$ , por  $\hat{\phi}$ , quando  $\boldsymbol{\beta}$  é conhecido. Portanto, as EEG (35)

podem ter a forma:

$$\sum_{i=1}^N \mathbf{U}_i \left[ \boldsymbol{\beta}, \hat{\boldsymbol{\alpha}} \left( \hat{\boldsymbol{\beta}}, \hat{\phi}(\boldsymbol{\beta}) \right) \right] = \mathbf{0},$$

cuja solução  $\hat{\boldsymbol{\beta}}_G$ , obtida a partir do método score de Fisher modificado, é dada por:

$$\boldsymbol{\beta}_G^{(m+1)} = \boldsymbol{\beta}_G^{(m)} + \left\{ \sum_{i=1}^N (\mathbf{D}'_i)^{(m)} (\boldsymbol{\Omega}_i^{-1})^{(m)} (\mathbf{D}_i)^{(m)} \right\}^{-1} \left\{ \sum_{i=1}^N (\mathbf{D}'_i)^{(m)} (\boldsymbol{\Omega}_i^{-1})^{(m)} (\mathbf{y}_i - \boldsymbol{\mu}_i^{(m)}) \right\}.$$

Sob condições moderadas de regularidade, isto é, se  $\hat{\boldsymbol{\alpha}}$  é um estimador  $N^{1/2}$  consistente para  $\boldsymbol{\alpha}$ , quando  $\boldsymbol{\beta}$  e  $\phi$  são conhecidos; se  $\hat{\phi}$  é um estimador  $N^{1/2}$  consistente para  $\phi$ , quando  $\boldsymbol{\beta}$  é conhecido, segundo Liang e Zeger (1986),  $\sqrt{N}(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})$  tem assintoticamente distribuição normal multivariada, com vetor de médias igual a zero e matriz de variâncias e covariâncias,  $\mathbf{V}_G$ , dada por:

$$\mathbf{V}_G = \lim_{N \rightarrow \infty} N \left[ \sum_{i=1}^N \mathbf{D}'_i \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i \right]^{-1} \left\{ \sum_{i=1}^N \mathbf{D}'_i \boldsymbol{\Omega}_i^{-1} \text{Var}(Y_i) \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i \right\} \left[ \sum_{i=1}^N \mathbf{D}'_i \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i \right]^{-1}.$$

Quando a matriz de correlação de trabalho é definida corretamente, um estimador consistente para  $\mathbf{V}_G$  é expresso por:

$$\hat{\mathbf{H}}_1 = \sum_{i=1}^N \hat{\mathbf{D}}'_i \hat{\boldsymbol{\Omega}}_i^{-1} \hat{\mathbf{D}}_i, \quad (36)$$

com  $\hat{\mathbf{D}}_i$  avaliada em  $\hat{\boldsymbol{\beta}}_G$  e  $\hat{\boldsymbol{\Omega}}_i$  avaliada em  $(\hat{\phi}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_G)$ . No entanto, se a matriz  $\mathbf{R}(\boldsymbol{\alpha})$  for especificada incorretamente, o estimador dado em (36) pode ser inconsistente e os autores sugerem a utilização de um estimador robusto, que será sempre consistente mesmo que a matriz de correlação de trabalho seja mal especificada, dado por:

$$\hat{\mathbf{V}}_G = \hat{\mathbf{H}}_1^{-1} \hat{\mathbf{H}}_2^{-1} \hat{\mathbf{H}}_1^{-1}, \quad (37)$$

em que  $\hat{\mathbf{H}}_2 = \left\{ \sum_{i=1}^N \mathbf{D}'_i \boldsymbol{\Omega}_i^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i) (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i) \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i \right\}$ .

Por outro lado, o estimador dado por (37), para um número pequeno de unidades experimentais ( $N \leq 30$ ), produz viés e, neste caso, Paik (1988) recomenda o uso do estimador de variância *jackknife*:

$$\frac{N-p}{N} \sum_{i=1}^N (\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}}) (\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}})',$$

em que  $p$  é o número de parâmetros e  $\hat{\beta}_{(i)}$  são as estimativas de  $\beta$  omitindo-se o  $i$ -ésimo *cluster* (indivíduo, unidade amostral).

As estimativas para os parâmetros de correlação  $\alpha$  e de escala  $\phi$  podem ser obtidas a partir do resíduo corrente de Pearson, cuja definição é :

$$\hat{r}_{it} = \frac{y_{it} - b'(\theta_{it})}{\sqrt{b''(\theta_{it})}}$$

em que  $\theta_{it}$  depende do valor corrente para  $\beta$ . Assim, tem-se que a estimativa do parâmetro de dispersão:

$$\hat{\phi} = \sum_{i=1}^N \sum_{t=1}^{n_i} \frac{\hat{r}_{it}^2}{(M - p)},$$

em que  $M = \sum n_i$ . A estimativa de  $\alpha$ , por sua vez, depende da escolha da matriz  $\mathbf{R}(\alpha)$ . Uma forma de se especificar  $\alpha$ , é a partir de uma função simples dos resíduos, como por exemplo:

$$\hat{R}_{tt'} = \frac{\sum_{i=1}^N \hat{r}_{it} \hat{r}_{it'}}{(M - p)}.$$

As estruturas mais comuns para a matriz de correlação de trabalho apresentadas pelos autores são:

i. Independência

Nesse caso, assume-se como matriz de correlação a matriz identidade,  $\mathbf{R}(\alpha) = \mathbf{I}_{n_i}$ , o que leva às equações de estimação definida em (33).

ii. M-dependente

Essa estrutura de correlação é usada, quando se assume que as observações repetidas seguem processo estacionário de dependência  $m$ , isto é:

$$R(\alpha) = \begin{cases} 0 & \text{se } |t_j - t'_k| > m; \\ \alpha^{|t_j - t'_k|} & \text{se } |t_j - t'_k| \leq m. \end{cases}$$

em que,  $t_j$  e  $t'_k$  denotam os  $j$ -ésimos e  $k$ -ésimos tempos de observação para a unidade amostral  $i$ . Nota-se, neste caso, que a correlação depende da distância entre as medidas repetidas, podendo tender a zero, para espaçamentos maiores do que  $m$ . Como um

caso particular, assumindo,  $\rho(y_{it}, y_{i(t+1)}) = \alpha_t$ , tem-se um modelo de dependência um. Neste caso, há  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_{n_i-1})'$  parâmetros a serem estimados, para os tempos  $t = 1, 2, \dots, n_i - 1$ , e:

$$R(\boldsymbol{\alpha}) = \begin{cases} 1 & \text{se } t = t'; \\ \alpha_t & \text{se } |t - t'| = 1; \\ 0 & \text{se } |t - t'| > 1. \end{cases}$$

com  $\alpha_t$  podendo ser estimado pela expressão:

$$\hat{\alpha}_t = \hat{\phi} \sum_{i=1}^N \frac{\hat{r}_{it}\hat{r}_{i,t+1}}{N-p}.$$

### iii. Simétrica ou Permutável

Nessa estrutura, assume-se uma matriz de correlação do tipo:

$$R(\boldsymbol{\alpha}) = \begin{cases} 1 & \text{se } t = t'; \\ \alpha & \text{se } t \neq t'. \end{cases}$$

Deve ser usada quando se assume um modelo de efeitos aleatórios com intercepto aleatório para cada indivíduo, como exemplificado em Laird e Ware (1982). A estimativa para  $\alpha$  é dada por:

$$\hat{\alpha} = \hat{\phi} \sum_{i=1}^N \sum_{t>t'} \frac{\hat{r}_{it}\hat{r}_{it'}}{\sum_{i=1}^N \frac{1}{2}n_i(n_i-1) - p}.$$

### iv. Auto-Regressiva (AR-1)

Nesse caso, assume-se uma estrutura de correlação auto-regressiva, sendo que as observações mais próximas tendem a ser mais correlacionadas, ou seja:

$$R(\boldsymbol{\alpha}) = \begin{cases} 1 & \text{se } t = t'; \\ \alpha^{|t-t'|} & \text{se } t \neq t'. \end{cases}$$

que, em particular, para  $n_i = 3$ , pode ser expressa por:

$$R(\boldsymbol{\alpha}) = \begin{bmatrix} 1 & \alpha & \alpha^2 \\ \alpha & 1 & \alpha \\ \alpha^2 & \alpha & 1 \end{bmatrix}$$

## v. Não-estruturada

A matriz não estruturada pode ser definida pela seguinte expressão:

$$R(\alpha) = \begin{cases} 1 & \text{se } t = t'; \\ \alpha_{tt'} & \text{se } t \neq t'. \end{cases}$$

e terá  $n_i(n_i - 1)/2$  parâmetros. Em particular, para  $n_i = 3$ , tem-se:

$$R(\alpha) = \begin{bmatrix} 1 & \alpha_{12} & \alpha_{13} \\ \alpha_{12} & 1 & \alpha_{23} \\ \alpha_{13} & \alpha_{23} & 1 \end{bmatrix}$$

havendo, portanto, 3 parâmetros a serem estimados. Essa estrutura é interessante quando não há muitos instantes de observação. Ademais, neste caso,  $\mathbf{R}(\alpha)$  pode ser estimada por:

$$\hat{\mathbf{R}}(\alpha) = \frac{\hat{\phi}}{N} \sum_{i=1}^N \hat{\mathbf{V}}_i^{-1/2} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)' \hat{\mathbf{V}}_i^{-1/2}.$$

O procedimento básico para se ajustarem modelos para respostas correlacionadas, através das EEG, atualmente, encontra-se implementado em vários *softwares*, como no procedimento **genmod** do SAS, no pacote **oswald** do S-PLUS (SMITH, 1998), nos pacotes **gee** e **geepack** do R e no *software* STATA. Em todos esses *softwares* há opções para se definir uma estrutura de correlação (decorre daí o nome correlação de “trabalho”). Há de se ressaltar, que ao se adotar esse procedimento metodológico, o foco do pesquisador está na distribuição marginal, isto é, a resposta média populacional é modelada como uma função das variáveis explicativas, sem levar em conta a heterogeneidade entre os indivíduos. Os coeficientes da regressão têm interpretação para a população, isto é, servem para descrever uma mudança na resposta média populacional para uma mudança na covariável  $x$ . Desse modo, esses modelos são exemplos específicos da classe *population-averaged* (ZEGGER; LIANG; ALBERT, 1988).

Com dados binários, os modelos marginais têm importância singular em estudos epidemiológicos, em que se deseja diferenciar grupos populacionais com diferentes fatores de risco. Suponha, a título de ilustração, um estudo longitudinal, no qual o estado de saúde dos

indivíduos é qualificado pelo sintoma de alguma doença (1) ou ausência desse sintoma (0). Entre as variáveis explicativas, está um fator de risco, como por exemplo, ser, ou não ser, fumante para o caso de avaliar doença respiratória. Adotando-se o modelo:

$$\text{logito}(\mu_{it}) = \ln\left(\frac{\mu_{it}}{1 - \mu_{it}}\right) = \ln\frac{P(Y_{it} = 1)}{P(Y_{it} = 0)} = \beta_0 + \beta_1 x_{1it} + \sum_{k=2}^p \beta_k x_{kit},$$

em que  $x_{1it}$  representa a variável binária para o fator de risco (1 presença e 0 ausência). Uma vez ajustado o modelo, sabe-se que  $\exp(\hat{\beta}_k)$  tem interpretação em termos de razão de chances. Nesse caso, o coeficiente  $\exp(\hat{\beta}_0)$  é a razão de frequências de indivíduos doentes e não doentes entre aqueles que não apresentam o fator de risco. Já o termo  $\exp(\hat{\beta}_1)$  é a razão de chances de doença entre os que apresentam e não apresentam o fator de risco. Quando a prevalência da resposta é baixa, essa razão de chances pode ser interpretada aproximadamente como a razão de prevalências da doença entre duas populações de indivíduos com e sem o fator de risco, numa dada visita (ZEGGER; LIANG, 1992).

### 2.1.3.3 Modelos de transição de Markov

Os modelos marginais, apesar de permitirem incorporar uma estrutura de correlação para os dados através da matriz  $\mathbf{R}(\boldsymbol{\alpha})$ , ainda não são suficientes para descrever toda a informação relacionada a um estudo longitudinal. Assim, quando a questão da dependência do tempo é o objeto central, ou seja, se se deseja descrever uma relação de dependência da distribuição condicional da resposta corrente no tempo  $t$ , isto é,  $y_{it}$ , sob as respostas anteriores  $(y_{i1}, y_{i2}, \dots, y_{i,(t-1)})$  e as variáveis explicativas  $\mathbf{x}_{it}$ , os modelos de transição são os mais apropriados. De fato, uma das grandes virtudes do modelo de transição é modelar as mudanças individuais (transições) no tempo e, avaliar, como essas mudanças são influenciadas pelas variáveis explicativas consideradas no estudo.

O estudo dos modelos de transição está também alicerçado na teoria dos processos estocásticos e, freqüentemente a propriedade markoviana é a mais usada. Neste trabalho, o enfoque é direcionado para a cadeia de Markov, uma vez que a variável resposta tem espaço discreto de estados,  $S = \{0, 1\}$ , e, o tempo, também é considerado discreto  $\tau = \{0, 1, 2, \dots, T\}$ . A respeito do tempo, cabe considerar que o que determina o fato de ele ser discreto é a forma como a resposta é obtida, através de ocasiões especificadas. Por exem-

plo, uma pesquisa de opinião feita quinzenalmente, um diagnóstico realizado em pacientes mensalmente, etc. Em grande parte dos processos, a rigor, o tempo é contínuo, caracterizando um processo de saltos (se o espaço de estados é discreto) ou processo de difusão (se o espaço de estados é contínuo). Em um processo de saltos, por exemplo, a modelagem probabilística dos intervalos de tempo em que o estado de processo se encontra é importante para obtenção dos parâmetros infinitesimais, que juntamente com as probabilidades de transição, caracterizam o processo. Maiores detalhes sobre processos em tempo contínuo, podem ser encontrados nos trabalhos de Singer e Spilerman (1976), Tuma, Hannan e Groeneveld (1979), Kalbfleisch e Lawless (1985), Ware, Lipsitz e Speizer (1988), Noletto (1991), Lindsey (2004), entre outros.

Assim, direcionando as cadeias de Markov para o caso de dados binários, a descrição das mudanças ou transições das observações entre uma ocasião e outra é caracterizada estocasticamente por matrizes de probabilidades de transição, de dimensões  $2 \times 2$ , pois a priori, assume-se que a propriedade markoviana é satisfeita. Para estabelecer a notação, considere:

i) Matriz de probabilidades de transição - processo não estacionário

$$P(t) = \begin{pmatrix} \pi_{00}(t) & \pi_{01}(t) \\ \pi_{10}(t) & \pi_{11}(t) \end{pmatrix},$$

em que o subscrito ( $t$ ) denota a dependência em relação ao tempo, havendo nesse caso,  $T$  matrizes de transição.

ii) Matriz de probabilidades de transição - processo estacionário

$$P = \begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix}.$$

No tocante à estimação, há na literatura, vários métodos disponíveis para se estimarem essas probabilidades condicionais e caracterizarem essas matrizes. No contexto deste estudo, faz-se uma distinção entre dois procedimentos básicos. O primeiro procedimento, baseado na análise de tabelas de contingência, é aplicado aos casos em que a amostra é homogênea. Entretanto, considerando a presença de variáveis explicativas, apresenta-se

um segundo procedimento, obtido a partir da definição de MLGs, incorporando a estrutura estocástica de uma cadeia de Markov.

#### 2.1.3.4 Tabelas de contingência e matrizes de probabilidades de transição

Para esse procedimento de estimação das probabilidades condicionais, considera-se que a amostra aleatória de indivíduos, avaliada em diversas ocasiões pré-determinadas, é homogênea. Alguns trabalhos como os de Anderson e Goodman (1957), Goodman (1962), Bishop, Fienberg e Holland (1975), Agresti (1990), Fitzmaurice e Laird (1993), Lindsey (1995, 2004) descrevem que as mudanças ou transições de categorias de resposta no tempo, nesse caso, podem ser estudadas através de técnicas aplicadas a tabelas de contingência, com as hipóteses de interesse sendo testadas por meio de modelos log-lineares. Uma situação hipotética é apresentada na Tabela 2, considerando respostas binárias. Em cada um dos tempos, apenas a resposta da unidade amostral (indivíduo) é considerada.

Tabela 2 - Estudo longitudinal com resposta binária supondo amostra homogênea

Unidade amostral	Tempo			
	0	1	...	T
1	0	1	...	1
2	1	1	...	0
3	0	0	...	0
4	1	0	...	1
⋮	⋮	⋮	⋮	⋮
$N$	0	0	...	1

Assim, para a situação em que as probabilidades de transição não são estacionárias, as transições observadas são representadas por um conjunto de estatísticas suficientes,  $n_{ab}(t), \forall a, b \in S = \{0, 1\}$  e, podem ser vistas como frequências observadas em  $T$  tabelas de contingência  $2 \times 2$ , como mostra a Tabela 3.

As estimativas pontuais das probabilidades de transição para as  $T$  matrizes de



Tabela 3 - Tabela de contingência  $2 \times 2$ , na  $t$ -ésima transição, de um estudo longitudinal com resposta binária e amostra homogênea

$Y_{(t-1)}$	$Y_t$		Total marginal ( $t - 1$ )
	0	1	
0	$n_{00}(t)$	$n_{01}(t)$	$n_{0.}(t - 1)$
1	$n_{10}(t)$	$n_{11}(t)$	$n_{1.}(t - 1)$
Total marginal ( $t$ )	$n_{.0}(t)$	$n_{.1}(t)$	$N$

transições são dadas pelas freqüências relativas dessas tabelas, fixando-se os totais marginais ( $t - 1$ ), coincidindo com as estimativas por máxima verossimilhança  $\hat{\pi}_{ab}(t) = \frac{n_{ab}(t)}{n_{a.}(t - 1)}$ , apresentadas na seção (2.1.2.4.1). Caso o teste para estacionariedade, baseado na estatística de qui-quadrado ou na razão de verossimilhanças, não seja significativo, é possível reduzir as  $T$  tabelas a uma só, com freqüências dadas por  $n_{ab} = \sum_{t=1}^T n_{ab}(t)$ . Decorre que  $\hat{\pi}_{ab} = \frac{n_{ab}}{n_{a.}}$ , freqüências relativas condicionais observadas, constituem-se em estimativas de probabilidades para a matriz de transição estacionária.

Para ilustrar as idéias descritas, considere os dados referentes às observações sobre precipitação pluviométrica diária (em  $mm$ ), em Piracicaba-SP, no mês de novembro, durante o período de 1997 a 2006. Os dados são do Departamento de Física e Meteorologia da ESALQ-USP, obtidos por instrumento automático. Nesse exemplo, pode-se tomar o ano como unidade amostral e os dias como as ocasiões de tempo em que são feitas as mensurações. A variável precipitação é categorizada, em ocorrência de chuva (1) ou não (0), conforme mostra a Tabela 4. Nesse período, há 159 dias sem chuva e 141 com incidência de precipitação.

Para cada ano há 29 transições, cujas freqüências são obtidas por meio de tabelas de contingência  $2 \times 2$ . Assumindo um processo estocástico estacionário, as 290 transições do período são sintetizadas na Tabela 5, que permite estimar as probabilidades de transição dada a condição no dia anterior, através de suas freqüências relativas. Observa-se que há uma certa dependência em relação ao estado no dia anterior. A probabilidade de chuva, no mês de novembro, dado que não chove no dia anterior é estimada em 0,3494, ao passo que se chove no dia anterior, essa probabilidade passa a ser 0,5162.

Tabela 4 - Dias do mês de novembro com precipitação (1) ou ausênica (0) em Piracicaba-SP

Anos	Dias									
1997	100	001	100	001	111	111	101	111	111	100
1998	000	001	010	110	100	000	000	000	011	000
1999	000	010	011	101	000	000	010	111	010	000
2000	001	001	000	111	111	011	011	101	110	011
2001	000	100	000	011	111	110	100	110	000	111
2002	000	110	000	111	111	011	100	001	111	111
2003	110	010	110	000	100	111	000	011	011	101
2004	000	111	000	011	001	111	100	000	000	110
2005	000	100	100	000	011	111	110	100	100	001
2006	110	001	001	100	111	011	110	000	110	110

Fonte: Estação meteorológica automática, ESALQ-USP

Tabela 5 - Total de freqüências de transições de primeira ordem observadas nos dados sobre precipitação, em Piracicaba-SP, para os dias do mês de novembro no período de 1997 a 2006

Dias <sub>(t-1)</sub>	Dias <sub>(t)</sub>		Total
	0	1	
0	108 (0,6506)	58 (0,3494)	166
1	60 (0,4838)	64 (0,5162)	124

Para caracterizar completamente o processo, as probabilidades iniciais da cadeia são estimadas marginalmente por  $\hat{\pi}'_0 = (0,7; 0,3)$ , referindo-se à ocorrência (0,3), ou não, de chuva (0,7) para o primeiro dia do mês. Assim, o modelo estocástico prediz uma probabilidade marginal de chuva para um dado dia do mês de novembro em 0,4194.

$$\hat{P}^{(6)} = \begin{pmatrix} 0,6506 & 0,3494 \\ 0,4838 & 0,5162 \end{pmatrix}^{(6)} = \begin{pmatrix} 0,5806 & 0,4194 \\ 0,5806 & 0,4194 \end{pmatrix}.$$

Essa probabilidade é obtida multiplicando-se o vetor de probabilidades iniciais pela matriz de transição sucessivamente. Verifica-se que a matriz de transição se estabiliza no passo 6. A Figura 1 ilustra esse processo.

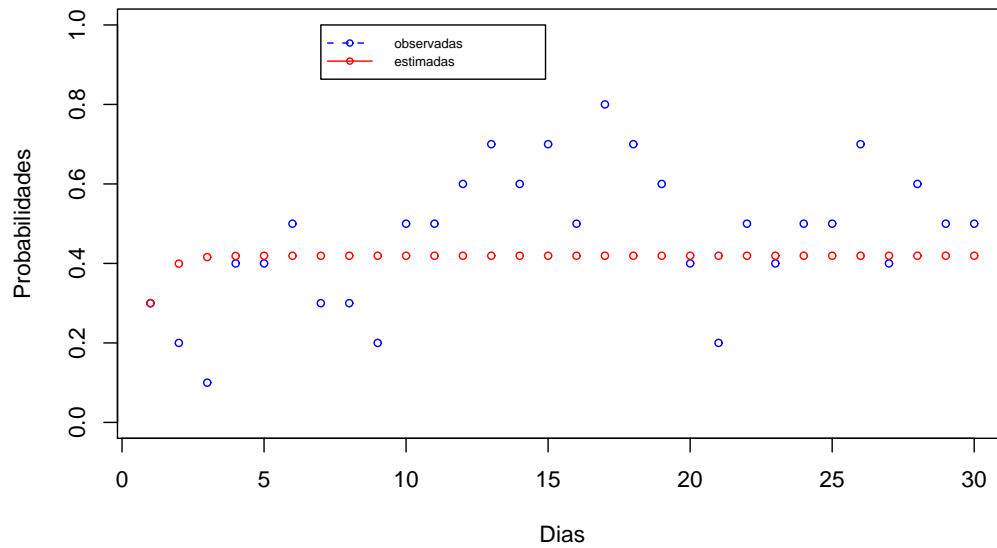


Figura 1 - Probabilidades marginais de chuva observadas para cada dia do mês de novembro, em Piracicaba-SP, no período de 1997 a 2006 e as correspondentes probabilidades estimadas pela cadeia de Markov

Na prática, porém, conforme destaca Noleto (1991), dificilmente a amostra é considerada homogênea. É fato também, que em estudos longitudinais, além do fator tempo, um dos objetivos é justamente avaliar a influência das variáveis explicativas na variável resposta. Se há um fator apenas, como por exemplo num estudo epidemiológico, fator tratamento qualificado em ativo e placebo, pode-se estratificar a amostra em duas, de tal forma a obter dentro de cada grupo (suposto homogêneo), as probabilidades de transição e através de testes apropriados, tirar as conclusões desejadas. No entanto, quando o número de variáveis explicativas envolvidas na análise aumenta, a estratificação sucessiva pode levar a uma diminuição nas frequências de caselas em cada um dos grupos, e, conseqüentemente,

o tamanho insuficiente da amostra inviabiliza a obtenção de estimativas consistentes para as probabilidades de transição, além de comprometer as estatísticas para os testes de interesse.

### 2.1.3.5 Modelos lineares generalizados de transição

A alternativa mais apropriada, quando o interesse é estudar o efeito das variáveis explicativas sob as probabilidades da matriz de transição, é através da especificação de modelos para descrever a relação funcional de dependência dessa probabilidade em relação ao conjunto de variáveis explicativas, consideradas no estudo. Nessa ótica, ao invés de se estimarem todas as probabilidades de transição em separado, como no procedimento anterior, busca-se modelar essas probabilidades de tal forma que as estimativas dos parâmetros obtidas podem ser interpretadas como os pesos que cada uma das variáveis explicativas exercem na estimação da probabilidade de transição.

Nesse sentido, Cox (1970) sugere o uso da regressão logística para modelar a esperança condicional de dados binários, em que as probabilidades de transição dependem das variáveis explicativas, correspondendo a uma cadeia de Markov. Esse método é aplicado no estudo do efeito da poluição do ar sobre a saúde (ataque de asma), por Korn e Whittemore (1979), que consideram o estado de saúde do indivíduo no tempo imediatamente anterior como covariável.

Nesses dois trabalhos, os dados são obtidos como séries temporais e assume-se que as respostas binárias para o  $i$ -ésimo indivíduo, dado um conjunto de variáveis explicativas, formam um processo estocástico de tempo discreto, isto é,  $t = 1, 2, \dots, n_i$ , com  $Y_{it}$  assumindo valores no espaço de estados  $S = \{0, 1\}$ . Seja  $\pi_t$  a probabilidade de um evento numa ocasião  $t$ , dadas as variáveis explicativas e a história do processo de  $Y_{it}$  do início do estudo até a ocasião  $(t - 1)$ , assumindo o modelo de regressão logística,  $\pi_t$  é especificada por:

$$\pi(t) = \frac{\exp(\alpha + \boldsymbol{\beta}'\mathbf{x})}{1 + \exp(\alpha + \boldsymbol{\beta}'\mathbf{x})}, \quad (38)$$

em que  $\mathbf{x}$  é um vetor de variáveis explicativas associadas à ocasião  $t$ ,  $\boldsymbol{\beta}$  o vetor de parâmetros desconhecidos e  $\alpha$  uma função de parâmetros desconhecidos definida por  $\alpha = \alpha_0 + y_{i(t-1)}\alpha_1$ . A função de verossimilhança para os dados  $y_{i1}, y_{i2}, \dots, y_{in}$  de um indivíduo é:

$$\prod_{t=1}^n \pi(t)^{y_{it}} [1 - \pi(t)]^{(1-y_{it})}. \quad (39)$$

A fim de maximizar (39), os autores consideram duas situações. A situação mais simples ocorre quando  $\alpha = \text{constante}$ , tal que o estado do processo numa dada ocasião é independente do estado processo em outra ocasião. Nesse caso, a fim de se obterem as estimativas dos parâmetros, Cox(1970) sugere maximizar a função de verossimilhança usual, ou seja, procedimento de estimação de parâmetros quando se assume uma regressão logística para dados independentes. Como as respostas são independentes, ocasiões com observações omissas podem ser simplesmente retiradas do estudo, assim,  $n$  em (39) representa o número de ocasiões sem dados perdidos. Korn e Whittemore (1979) mostram que o parâmetro  $\alpha$  pode ser eliminado da estimação de  $\beta$ , maximizando a função de verossimilhança condicional ao número total de eventos. A função de verossimilhança é expressa, então como:

$$L(\beta, \mathbf{y}) \propto \frac{\exp(\beta' \sum y_{it} \mathbf{x})}{\sum_A \exp(\beta' \sum_A \mathbf{x})}, \quad (40)$$

sendo a soma no denominador sob todos os conjuntos  $A$  de ocasiões no período.

Num segundo caso, assume-se que as respostas estão correlacionadas serialmente. Assim,  $\alpha$  pode assumir dois valores,  $\alpha_0$  ou  $\alpha_0 + \alpha_1$ , dependendo de  $y_{i(t-1)} = 0$  ou  $y_{i(t-1)} = 1$ . A função de verossimilhança (39) pode ser escrita como:

$$L(\beta, \alpha_0, \alpha_1, \mathbf{y}) \propto \frac{\exp(\alpha_0 y + \alpha_1 \sum y_{it} y_{i(t-1)} + \beta' \sum y_{it} \mathbf{x})}{\prod_{t=1}^n [1 + \exp(\alpha_0 + \alpha_1 y_{i(t-1)} + \beta' \mathbf{x})]}, \quad (41)$$

resultando nas estimativas de máxima verossimilhança, apresentadas por Cox (1970). Nesse segundo caso, se a resposta na ocasião  $(t - 1)$  é perdida, então a resposta na ocasião  $t$  deve ser omitida, a fim de que  $\pi(t)$  possa ser calculada.

Cabe ressaltar que, diferentemente do caso normal, modelos para esperança condicional, como os de Cox (1970), de uma variável binária não implicam igualmente em expressões simples para as esperanças marginais. Assim, de forma diferente, Zeger, Liang e Self (1985) consideram dois modelos de trabalho, nos quais as probabilidades marginais ao invés das condicionais são expressas como funções logísticas das variáveis explicativas, a fim de aproximar a verdadeira função de verossimilhança por função de verossimilhança de trabalho que levam a estimativas consistentes de  $\beta$  sob suposições moderadas sobre a dependência do tempo. Como nas propostas anteriores, os autores usam séries temporais ao invés de modelos de medidas repetidas para explicar essa dependência.

No primeiro modelo, o estimador,  $\hat{\beta}_0$ , é obtido, supondo-se que observações repetidas para um indivíduo são independentes entre si. Esse estimador é consistente dado qualquer conjunto de processos estacionários binários tais que  $\text{logito}(\pi_i) = \mathbf{x}'_i\beta$ . No segundo modelo proposto, o estimador,  $\hat{\beta}_1$ , é obtido a partir do modelo de trabalho no qual cada série é uma cadeia estacionária de Markov de ordem um. Ele é consistente quando a série satisfaz a suposição de função de ligação logística e tem uma autocorrelação comum de *lag* 1, ou seja  $\rho = \text{cor}(Y_{it}, Y_{i,(t-1)})$ . Sob o modelo estacionário de primeira ordem, o logaritmo da função de verossimilhança é:

$$\ell(\beta, \rho) \propto \sum_{i=1}^N \left[ y_{i1} \mathbf{x}'_i \beta - \ln(1 + \exp(\mathbf{x}'_i \beta)) + \sum_{t=2}^{n_i} \{y_{it} \ln \pi_{it} + (1 - y_{it}) \ln(1 - \pi_{it})\} \right],$$

em que  $\pi_{it} = \mathbf{E}(Y_{it} | Y_{i(t-1)}, \mathbf{x}_i) = \pi_i + \rho(Y_{i(t-1)} - \pi_i)$ .

Os autores apresentam também as equações score bem como as condições de consistência e normalidade assintótica associadas aos estimadores propostos. Cabe ressaltar que a proposta de Zeger, Liang e Self (1985) pode ser entendida como um caso particular do trabalho de Azzalini (1983), que apresenta questões sobre a estimação por máxima verossimilhança de processos estocásticos de séries estacionárias de ordem  $q$ . De acordo com a proposta de Azzalini (1983), há situações em modelos de processos estocásticos em que a maximização da função de verossimilhança exata não é viável, restando como alternativa, uma aproximação para essa função. Assim, se  $Y_{it}$  representa um processo estocástico estacionário indexado por um vetor de parâmetros  $\beta$ , a função de verossimilhança de  $q$ -ésima ordem para  $i$ -ésimo indivíduo é aproximada pela expressão:

$$\ell_i^{(n)}(\beta) \propto \sum_{t=1}^{n_i} \ln f(y_{it} | y_{i(t-1)}, \dots, y_{i(t-q)}; \beta), \quad (42)$$

em que  $f(y_{it} | y_{i(t-1)}, \dots, y_{i(t-q)})$  é a função de densidade ou distribuição de probabilidade de  $Y_{it}$  dado  $(y_{i(t-1)}, \dots, y_{i(t-q)})$ . Se  $q = 0$ , então não existe condicional e  $f$  é simplesmente uma função marginal. Azzalini (1983) mostra também que sob condições de estacionariedade e ergodicidade do processo,  $\hat{\beta}$ , o estimador de máxima verossimilhança de (42), tende ao valor verdadeiro quando  $n_i \rightarrow \infty$ . O autor também apresenta exemplos de eficiência assintótica desse estimador para os casos de um processo de Markov normal e para uma cadeia de Markov de segunda ordem de uma série binária.

Esses trabalhos são exemplos pioneiros daqueles que são tratados na literatura, como modelos baseados na cadeia de Markov, sendo que a dimensão das respostas anteriores, define a ordem dessa cadeia. A desvantagem das estruturas propostas tanto por Cox (1970), Korn e Whittemore (1979), Zeger, Liang e Self (1985) está no fato de que os modelos limitam-se a variáveis binárias e dados perdidos ou uma especificação incorreta da cadeia de Markov, podem gerar estimativas inconsistentes. Além disso, o trabalho de Zeger, Liang e Self (1985) só permite variáveis explicativas independentes do tempo e a metodologia proposta é aplicável somente quando o processo é estacionário.

Uma proposta alternativa, que permite trabalhar com duas ou mais categorias de resposta, variáveis explicativas dependentes do tempo e com inclusão de variável indicadora para possibilidade de observações omissas, é o trabalho de Ware, Lipsitz e Speizer (1988). Na verdade, essa proposta é uma adaptação do modelo marginal de Stram, Wei e Ware (1988), em que a cadeia de Markov é caracterizada no modelo pela inclusão da categoria de resposta no tempo anterior como uma covariável independente em um modelo de regressão. Assim, seguindo os procedimentos e a notação definidos para o modelo de Stram, Wei e Ware (1988), as probabilidades de transição são dadas por:

$$\pi_{ab}(t-1, t) = P(I_{ibt} \mid I_{ia(t-1)}, \mathbf{x}_{it}), \quad a, b = 1, 2, \dots, k; \quad t = 0, 1, \dots, T,$$

em que  $\mathbf{x}_{it} = (x_{1it}, x_{2it}, \dots, x_{pit}, x_{(p+1),it})'$  é o vetor das  $(p+1)$  variáveis explicativas associadas ao indivíduo  $i$  na  $t$ -ésima transição. Aqui,  $x_{1it}$  representa o valor assumido por  $I_{ia(t-1)}$ . Essas probabilidades de transição podem ser estimadas como nos modelos marginais, ou seja, a partir do modelo de chances proporcionais. O logaritmo da função de verossimilhança desse modelo, no tempo  $t$  é:

$$\ell_t \propto \sum_{i=1}^N \delta_{ti} \sum_{j=1}^k I_{ijt} \{ \ln[\phi(\lambda_{\ell mt} - \mathbf{x}'_{it} \boldsymbol{\beta}_t) - \phi(\lambda_{\ell, m-1, t} - \mathbf{x}'_{it} \boldsymbol{\beta}_t)] \}, \quad (43)$$

em que  $\delta_{ti}$  é a variável indicadora para observações omissas, ou seja,  $\delta_{ti} = 1$ , se  $I_{it}$  e  $\mathbf{x}_{it}$  são observadas, e assume 0 em caso contrário. As probabilidades de transição acumuladas são especificadas por:

$$\gamma_{abt}(\mathbf{x}) = \frac{\exp(\lambda_{abt} - \boldsymbol{\beta}'_t \mathbf{x})}{1 + \exp(\lambda_{abt} - \boldsymbol{\beta}'_t \mathbf{x})} \quad b = 1, 2, \dots, k-1, \quad (44)$$

em que  $\lambda_{abt}$  é um parâmetro de perturbação do modelo, denominado intercepto e o vetor  $\mathbf{x}$  representa o conjunto de valores das variáveis explicativas, o qual também pode variar no tempo e  $\boldsymbol{\beta}'_t = (\beta_{1t}, \beta_{2t}, \dots, \beta_{(p+1),t})$  é o vetor de parâmetros desconhecidos que representam os pesos que as variáveis explicativas exercem nas probabilidades de transição.

Para dados binários, o modelo de chances proporcionais (44) reduz-se ao modelo de regressão logística para modelar as probabilidades condicionais. É importante notar que nesse método, modelam-se as  $T$  transições possíveis, que permitem estimar e estudar no decorrer das transições as matrizes de probabilidades condicionais. Constitui-se, portanto, em uma técnica apropriada aos processos não estacionários.

Há outras extensões de modelos lineares generalizados para modelos de transição. Um exemplo é o trabalho proposto por Zeger e Qaqish (1988), em que a teoria da quase-verossimilhança é usada para definir o modelo de Markov. Posteriormente, Zeger e Liang (1992) e, mais recentemente, Diggle et al. (2002) discutem condições mais gerais para definir um modelo de transição de Markov. De acordo com esses autores, um modelo de transição de Markov especifica um modelo linear generalizado para a distribuição condicional de  $Y_{it}$  dadas as respostas passadas. Assim, seja  $\mathbf{h}_{it} = (y_{i1}, y_{i2}, \dots, y_{i,(t-1)})$  o vetor de dimensão  $q$  das respostas prévias, então a distribuição condicional de  $Y_{it}$  é dada por:

$$f(y_{it} | \mathbf{h}_{it}) = \exp \left\{ \frac{w_i}{\phi} \left[ y_{it} \theta_{it} - b(\theta_{it}) \right] + c(y_{it}, \phi) \right\}, \quad (45)$$

em que

$$\mu_{it}^C = E(Y_{it} | \mathbf{h}_{it}) = b'(\theta_{it}) \quad \text{e} \quad v_{it}^C = \text{Var}(Y_{it} | \mathbf{h}_{it}) = b''(\theta_{it})\phi,$$

denotam a média e variância condicionais a  $\mathbf{h}_{it}$ .

Considera-se, ainda, que a média e a variância condicional satisfazem as equações:

$$g(\mu_{it}^C) = \eta_{it} = \mathbf{x}'_{it} \boldsymbol{\beta} + \sum_{r=1}^s f_r(\mathbf{h}_{it}; \boldsymbol{\alpha}) \quad \text{e} \quad v_{it}^C = v(\mu_{it}^C)\phi, \quad (46)$$

em que  $g(\mu_{it}^C)$  e  $v(\mu_{it}^C)$  denotam a função de ligação e a função de variância, respectivamente. Nota-se pela expressão (46), que a média condicional depende tanto das variáveis explicativas como das respostas nos instantes anteriores. As respostas prévias ou funções delas mesmas são tratadas como variáveis explicativas adicionais.



### 2.1.3.6 Estimação por máxima verossimilhança para processos estacionários

Um processo para se ajustarem modelos de transição de Markov é apresentado por Diggle et al. (2002). Esse procedimento difere da proposta de Ware, Lipsitz e Speizer (1988), pois ajusta-se um único modelo, a partir de uma soma de contribuições individuais (perfis de respostas) para a função de verossimilhança. Nesse contexto, em analogia, as estimativas obtidas a partir das tabelas de contingência, pode-se pensar como um procedimento para modelos estocásticos estacionários.

Assumindo um modelo de transição de Markov de ordem  $q$ , a distribuição condicional de  $Y_{it}$  é expressa por:

$$f(y_{it} | \mathbf{h}_{it}) = f(y_{it} | y_{i(t-1)}, y_{i(t-2)}, \dots, y_{i(t-q)}),$$

tal que a contribuição na função de verossimilhança do  $i$ -ésimo indivíduo é dada por:

$$f(y_{i1}, y_{i2}, \dots, y_{iq}) \prod_{t=q+1}^{n_i} f(y_{it} | y_{i(t-1)}, y_{i(t-2)}, \dots, y_{i(t-q)}).$$

Na prática, o modelo linear generalizado (45) define apenas a distribuição condicional, portanto, a função de verossimilhança das  $q$  primeiras observações  $f(y_{i1}, y_{i2}, \dots, y_{iq})$  não é especificada diretamente, exceto para o caso normal, em que sob determinadas suposições é possível especificar a distribuição marginal de  $Y_{i1}, \dots, Y_{iq}$  e, conseqüentemente maximizar a função de verossimilhança para o ajuste do modelo auto-regressivo normal. Por exemplo, se para dados contínuos, assume-se que:

$$Y_{i1} = \mathbf{x}'_{i1}\beta + \varepsilon_{i1} \quad e \quad Y_{it} = \mathbf{x}'_{it}\beta + \alpha Y_{i(t-1)} + \varepsilon_{it},$$

e,  $\varepsilon_{i1} \sim N(0, \sigma^2)$  e  $\varepsilon_{it} \sim N(0, \sigma^2(1 - \alpha^2))$ , decorre imediatamente que  $\text{Var}(Y_{it}) = \sigma^2$  e  $\text{Cov}(Y_{it}Y_{it'}) = \alpha^{|t'-t|}\sigma^2$ . Essas suposições levam a um modelo marginal normal multivariado com matriz de variâncias e covariâncias  $AR(1)$ .

Para os demais casos, a função  $f(y_{i1}, y_{i2}, \dots, y_{iq})$  não é determinada e, por conseguinte, a função de verossimilhança total não pode ser especificada. Uma alternativa, é estimar  $\beta$  e  $\alpha$  maximizando a função de verossimilhança condicional:

$$\prod_{i=1}^N f(y_{i(q+1)}, \dots, y_{in_i} | y_{i1}, \dots, y_{iq}) = \prod_{i=1}^N \prod_{t=q+1}^{n_i} f(y_{it} | \mathbf{h}_{it}). \quad (47)$$

Há duas possibilidades distintas para se maximizar a função (47). Numa primeira situação, se  $f_r(h_{it}; \boldsymbol{\alpha}, \boldsymbol{\beta}) = \alpha_r f_r(h_{it})$ , tal que  $g(\mu_{it}^C) = \mathbf{x}'_{it} \boldsymbol{\beta} + \sum_{r=1}^s f_r(h_{it})$  é uma função linear de ambos os parâmetros  $\boldsymbol{\alpha}$  e  $\boldsymbol{\beta}$ , então é possível proceder à estimação dos parâmetros como se fosse um MLG para dados independentes, isto é, fazendo a regressão de  $Y_{it}$  contra  $(p + s)$  variáveis explicativas  $(\mathbf{x}_{it}, f_1(h_{it}), \dots, f_s(h_{it}))$ . Desse modo, usando a suposição estabelecida em (45), segue que o logaritmo da função de verossimilhança condicional (47) é:

$$\ell(\boldsymbol{\theta}_{it}) = \sum_{i=1}^N \sum_{t=q+1}^{n_i} \frac{1}{\phi} \left[ y_{it} \theta_{it} - b(\theta_{it}) \right] + c(y_{it}, \phi). \quad (48)$$

Para obter a função escore, tem-se que obter as derivadas em relação ao conjunto de parâmetros, aqui denotado por  $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$ . Mas, observa-se em (48) que:

$$\ell = f(\boldsymbol{\theta}_{it}); \quad \boldsymbol{\theta}_{it} = q(\boldsymbol{\mu}_{it}^C); \quad \boldsymbol{\mu}_{it}^C = g^{-1}(\boldsymbol{\eta}_{it}) \text{ e } \boldsymbol{\eta}_{it} = \mathbf{x}'_{it} \boldsymbol{\beta} + \sum_{r=1}^q \alpha_r y_{it-r}.$$

Portanto, usando a regra da cadeia obtém-se a função escore:

$$\mathbf{U}^C(\boldsymbol{\delta}) = \sum_{i=1}^N \sum_{t=q+1}^{n_i} \frac{1}{\phi} (y_{it} - \mu_{it}^C) \frac{\partial \theta_{it}}{\partial \mu_{it}^C} \frac{\partial \mu_{it}^C}{\partial \boldsymbol{\eta}_{it}} \frac{\partial \boldsymbol{\eta}_{it}}{\partial \boldsymbol{\delta}}$$

e, igualando a zero, obtém-se:

$$\mathbf{U}^C(\boldsymbol{\delta}) = \sum_{i=1}^N \sum_{t=q+1}^{n_i} \frac{\partial \mu_{it}^C}{\partial \boldsymbol{\delta}} (v_{it}^C)^{-1} (y_{it} - \mu_{it}^C) = 0. \quad (49)$$

A equação de estimação (49) para esse caso pode ser vista como uma analogia condicional da equação escore definida para os MLGs. A derivada  $\frac{\partial \mu_{it}^C}{\partial \boldsymbol{\delta}}$ , faz o papel da matriz  $\mathbf{X}$ , mas ela depende tanto de  $\boldsymbol{\alpha}$  como de  $\boldsymbol{\beta}$ .

Adicionalmente, pode-se pensar nesse procedimento de estimação como um processo iterativo de mínimos quadrados ponderados. Para tanto, considere  $\mathbf{Y}_i$  um vetor de respostas, com dimensão  $(n_i - q)$ , para  $t = (q+1), \dots, n_i$  e  $\boldsymbol{\mu}_i^C$  seu valor esperado condicionado a  $\mathbf{h}_{it}$ . Seja, também,  $\mathbf{X}_i^*$  uma matriz, com dimensões  $(n_i - q) \times (p + s)$ , cuja  $k$ -ésima linha é  $\frac{\partial \mu_{i(q+k)}}{\partial \boldsymbol{\delta}}$  e  $\mathbf{W}_i = \text{diag}(1/v_{i(k+q)}^C)$ ; com  $k = (1, \dots, (n_i - q))$ , a matriz de ponderação, de dimensões  $(n_i - q) \times (n_i - q)$ . Por fim, seja  $\mathbf{Z}_i = \mathbf{X}_i^* \hat{\boldsymbol{\delta}} + (\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i^C)$ . Então, o valor atualizado  $\hat{\boldsymbol{\delta}}$  pode ser obtido iterativamente pela regressão de  $\mathbf{Z}$  contra  $\mathbf{X}^*$  usando a matriz de pesos

$\mathbf{W}$ , isto é:

$$\hat{\boldsymbol{\delta}}^{(m+1)} = \left( \sum_{i=1}^N \mathbf{X}_i^{*'} \mathbf{W}_i^{(m)} \mathbf{X}_i^* \right)^{-1} \left( \sum_{i=1}^N \mathbf{X}_i^{*'} \mathbf{W}_i^{(m)} \mathbf{Z}_i^{(m)} \right). \quad (50)$$

Nessas condições, quando o modelo assumido para a média e a variância condicionais é correto, a solução  $\hat{\boldsymbol{\delta}}$  de (50), assintoticamente, tem distribuição normal com média igual ao valor esperado,  $\boldsymbol{\delta}$ , e matriz de variâncias e covariâncias, de dimensão  $(p + s)$  dada por

$$\mathbf{V}_{\hat{\boldsymbol{\delta}}} = \left( \sum_{i=1}^N \mathbf{X}_i^* \mathbf{W}_i \mathbf{X}_i^* \right)^{-1}.$$

Entretanto, para os casos em que não se garante que o modelo esteja corretamente especificado, pode ser usado um estimador robusto:

$$\mathbf{V}_{\hat{\boldsymbol{\delta}}_R} = \left( \sum_{i=1}^N \mathbf{X}_i^* \mathbf{W}_i \mathbf{X}_i^* \right)^{-1} \left( \sum_{i=1}^N \mathbf{X}_i^* \mathbf{W}_i \mathbf{V}_i \mathbf{W}_i \mathbf{X}_i^* \right) \left( \sum_{i=1}^N \mathbf{X}_i^* \mathbf{W}_i \mathbf{X}_i^* \right)^{-1}. \quad (51)$$

Uma estimativa consistente para  $\mathbf{V}_{\hat{\boldsymbol{\delta}}}$  pode ser obtida, substituindo-se na equação (51),  $\mathbf{V}_i = \text{Var}(\mathbf{Y}_i \mid \mathbf{h}_{it})$ , por sua estimativa dada por  $(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i^C)(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i^C)'$ . Um segundo caso para a estimação dos parâmetros de um modelo de transição pode ocorrer quando as funções das respostas prévias,  $f_r$ , incluem ambos os parâmetros  $\boldsymbol{\alpha}$  e  $\boldsymbol{\beta}$ . Um exemplo dessa situação é dado em Zeger e Qaqish (1988). Os autores trabalham com dados de contagem e assumem um modelo de transição de Markov estacionário de primeira ordem, em que  $f = \alpha[\log(y_{i(t-1)} - \mathbf{x}'_{i(t-1)}\boldsymbol{\beta})]$ , caracterizando, portanto, um preditor não linear. Nesse caso, pode ser necessário usar métodos para modelos não lineares. Situações como essa, não são consideradas neste trabalho.

Para exemplificar as matrizes envolvidas no algoritmo de estimação, considere uma situação com duas variáveis explicativas e quatro tempos de observação. Admita processo estacionário de dependência 1. Seja, a princípio,  $\mathbf{y}_i$ , o vetor de variáveis respostas, de dimensões  $4 \times 1$ , isto é,  $\mathbf{y}_i = (y_{i1}, y_{i2}, y_{i3}, y_{i4})'$ , e seja  $\mathbf{X}_i = (1, x_{i1}, x_{i2})'$  a matriz de variáveis explicativas, de dimensões  $4 \times 3$ , cujas linhas são os vetores  $\mathbf{x}_{it} = (1, x_{it1}, x_{it2})'$ . Represente,  $Y_{it}$  a presença ou ausência de uma determinada doença, enquanto que  $X_{it1}$  denota a presença ou ausência de um determinado fator, para a  $i$ -ésima unidade experimental ( $i = 1, 2, \dots, N$ ), na ocasião  $t$ . Assumindo função de ligação canônica, tem-se:

$$\eta_{it} = \text{logito } P(Y_{it} = 1 \mid Y_{i(t-1)} = a) = \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \alpha y_{i(t-1)}, \quad a \in \{0, 1\}, \quad (52)$$

sendo  $\text{Var}(Y_{it} | Y_{i(t-1)}) = \mu_{it}^C(1 - \mu_{it}^C)$ .

Dado que a contribuição individual para a função de verossimilhança é expressa por:

$$f(y_{it}, y_{i2}, \dots, y_{i4}) = f(y_{i1}) \prod_{t=2}^4 f(y_{it} | y_{i(t-1)}),$$

e, a dificuldade está em determinar a distribuição marginal de  $f(y_{i1})$ , uma alternativa é trabalhar somente com a distribuição condicional. Assim, para maximizar a função de verossimilhança condicional de  $Y_{i2}, Y_{i3}, Y_{i4}$  dado  $Y_{i1}$  deve-se omitir  $f(y_{i1})$  dessa equação. Operacionalmente, isso equivale a excluir as observações das variáveis resposta e explicativas associadas à primeira ocasião e, incorporar uma “nova” covariável, a resposta prévia. Para o exemplo em questão, as matrizes  $\mathbf{Y}_i$ ,  $\mathbf{X}_i^*$  e  $\mathbf{W}_i$ , são, respectivamente, expressas por:

$$\mathbf{Y}_i = \begin{bmatrix} y_{i2} \\ y_{i3} \\ y_{i4} \end{bmatrix}, \quad \mathbf{X}_i^* = \begin{bmatrix} 1 & x_{i21} & x_{i22} & y_{i1} \\ 1 & x_{i31} & x_{i32} & y_{i2} \\ 1 & x_{i41} & x_{i42} & y_{i3} \end{bmatrix}, \quad \mathbf{W}_i = \begin{bmatrix} \frac{1}{\mu_{i2}^C(1-\mu_{i2}^C)} & 0 & 0 \\ 0 & \frac{1}{\mu_{i3}^C(1-\mu_{i3}^C)} & 0 \\ 0 & 0 & \frac{1}{\mu_{i4}^C(1-\mu_{i4}^C)} \end{bmatrix}.$$

e o vetor de parâmetros por  $\boldsymbol{\delta}' = (\beta_0, \beta_1, \beta_2, \alpha)'$ . Assim, por mínimos quadrados ponderados iterativamente (50), é possível obter a solução  $\hat{\boldsymbol{\delta}}' = [\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\alpha}]'$ .

Quando se trabalha com dados binários em modelos de transição, Zeger e Liang (1992) explicam que as estimativas dadas em  $\boldsymbol{\delta}$ , assim como nos modelos marginais, também têm interpretações em termos de razão de chances, através de  $\exp(\delta)$ . Em particular, para a situação dada em (52), após o ajuste do modelo,  $\exp(\hat{\alpha})$  pode ser interpretada como a estimativa da razão de chances de doença entre os indivíduos que tinham e não tinham a doença na ocasião anterior. Por outro lado,  $\exp(\hat{\beta}_1)$  é aproximadamente a razão de incidência da doença entre dois grupos: com e sem a presença do fator  $x_1$ .

### 2.1.3.7 Modelos com efeitos aleatórios

Num modelo de efeitos aleatórios assume-se que a resposta é uma função de uma ou mais variáveis explicativas com coeficientes de regressão que variam de um indivíduo para o outro. Essa variabilidade entre os coeficientes de regressão surge de uma heterogeneidade natural entre os indivíduos devido a fatores não mensuráveis. Esses modelos, segundo Zeger,

Liang e Albert (1988), usam tanto a informação contida na resposta média populacional como uma suposição acerca da heterogeneidade entre os indivíduos para estimar os coeficientes específicos. Portanto, são modelos da classe *subject-specific* (modelos SS).

Na definição do modelo linear generalizado misto (MLGM), assume-se que a correlação entre as respostas para um indivíduo seja resultante de uma heterogeneidade natural nos coeficientes de regressão entre os indivíduos e, a distribuição condicional das respostas dado um coeficiente individual, são variáveis aleatórias condicionalmente independentes, seguindo um modelo linear generalizado (DIGGLE et al., 2002; MOLENBERGHS; VERBEKE, 2005). Assim, formalmente, tem-se:

$$Y_{it}|\mathbf{d}_i \sim f_{Y_{it}|\mathbf{d}_i}(y_{it}|\mathbf{d}_i),$$

$$f_{Y_{it}|\mathbf{d}_i}(y_{it}|\mathbf{d}_i) = \exp \left\{ \frac{w_i}{\phi} \left[ y_{it}\theta_{it} - b(\theta_{it}) \right] + c(y_{it}, \phi) \right\}, \quad (53)$$

visto que  $\mu_{it} = b'(\theta_{it})$ , decorre:

$$\mathbf{E}(Y_{it}|\mathbf{d}_i) = \mu_{it}$$

é a média condicional da variável aleatória  $Y_{it}$ , que é funcionalmente ligada à parte sistemática por uma função de ligação:

$$g(\mathbf{E}(Y_{it}|\mathbf{d}_i)) = g(\mu_{it}) = \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{it}\mathbf{d}_i, \quad (54)$$

em que  $\mathbf{x}_{it}$  é a  $i$ -ésima linha da matriz de delineamento associada aos efeitos fixos,  $\boldsymbol{\beta}$  é o vetor de parâmetros dos efeitos fixos,  $\mathbf{z}_{it}$  é a  $i$ -ésima linha da matriz do modelo associada aos efeitos aleatórios e  $\mathbf{d}_i$  é o vetor dos parâmetros dos efeitos aleatórios. A parte sistemática do modelo (54) inclui tanto os efeitos fixos como os aleatórios. Assume-se que:

$$\mathbf{d}_i \sim f_d(\mathbf{d}),$$

sendo usual a escolha da distribuição normal multivariada para esses efeitos aleatórios, ou seja,  $\mathbf{d}_i \sim N_q(\mathbf{0}, \mathbf{G})$ . O objetivo da análise é estimar os coeficientes dos efeitos fixos,  $\boldsymbol{\beta}$ , os parâmetros de  $f_d$  e, possivelmente o parâmetro de escala,  $\phi$ .

### 2.1.3.8 Estimação dos parâmetros de um modelo misto

A função de verossimilhança para o vetor de parâmetros desconhecidos  $\Delta$ , que inclui ambos os elementos de  $\beta$  e  $\mathbf{G}$  é:

$$L(\Delta, \phi, \mathbf{y}) = \prod_{i=1}^N \int \prod_{t=1}^{n_i} f(y_{it} | \mathbf{d}_i, \beta, \phi) f(\mathbf{d}_i, \mathbf{G}) d(\mathbf{d}_i), \quad (55)$$

que é a distribuição marginal de  $\mathbf{Y}$  obtida integrando a distribuição conjunta  $(\mathbf{Y}, \mathbf{d})$  em relação a  $\mathbf{d}$ .

O problema para maximizar a função (55) é a presença de  $N$  integrais sob os efeitos aleatórios,  $\mathbf{d}_i$ . No caso do modelo linear misto, com função de ligação identidade, essas integrais podem ser resolvidas analiticamente. De fato, conforme Laird e Ware (1982), sob um modelo linear misto:

$$f_i(\mathbf{y}_i | \Delta, \phi) = \int \prod_{t=1}^{n_i} f(y_{it} | \mathbf{d}_i, \beta, \phi) f(\mathbf{d}_i, \mathbf{G}) d(\mathbf{d}_i),$$

é a função de densidade de uma distribuição normal multivariada, com parâmetros:

$$\mathbf{E}(\mathbf{Y}_i) = \mathbf{X}_i \beta \quad \text{e} \quad \text{Cov}(\mathbf{Y}_i) = \mathbf{Z}_i \mathbf{G} \mathbf{Z}_i' + \Sigma_i. \quad (56)$$

Assim, de (56), nota-se que os efeitos aleatórios num modelo linear misto não alteram a esperança marginal de  $\mathbf{Y}$ , somente a matriz de covariâncias. Conseqüentemente, os parâmetros  $\beta$ , têm interpretações tanto como modelos PA como SS (ZEGGER; LIANG; ALBERT, 1988). Para os casos em que a resposta não tem distribuição normal, a literatura apresenta vários procedimentos para se maximizar a função de verossimilhança, com o auxílio de métodos numéricos para solução aproximada das integrais dessa função (55). O algoritmo EM é também o processo iterativo mais usual para obtenção das estimativas dos parâmetros do MLGM (STIRATELLI; LAIRD; WARE, 1984; BRESLOW; CLAYTON, 1993, PINHEIRO; BATES, 2000, DIGGLE et al., 2002). Molenberghs e Verbeke (2005) salientam que, em geral, as aproximações numéricas exigidas podem ser divididas em três grupos: a aproximação dos integrandos, a aproximação dos dados e a aproximação da integral. A seguir, apresenta-se uma breve descrição desses métodos, de acordo com os referidos autores.

i. Aproximação dos integrandos

Essa técnica é aplicada para obter uma integral tratável e de forma fechada, quando os integrandos são aproximados. Entre os métodos existentes, o mais usual é o método baseado na aproximação de Laplace da função a ser integrada. Esse método permite aproximar integrais na forma:

$$I = \int e^{Q(\mathbf{d})} d(\mathbf{d}), \quad (57)$$

quando  $Q(\mathbf{d})$  é uma função conhecida e unimodal de uma variável  $q$ -dimensional,  $\mathbf{d}$ . Considere  $\hat{\mathbf{d}}$  o valor de  $\mathbf{d}$  para o qual  $Q$  é maximizada. Fazendo a expansão de Taylor de segunda ordem, obtém-se:

$$Q(\mathbf{d}) \approx Q(\hat{\mathbf{d}}) + \frac{1}{2}(\mathbf{d} - \hat{\mathbf{d}})' Q''(\hat{\mathbf{d}})(\mathbf{d} - \hat{\mathbf{d}}), \quad (58)$$

em que  $Q''(\hat{\mathbf{d}})$  é matriz hessiana de  $Q$ , avaliada em  $\hat{\mathbf{d}}$ . Assim, trocando  $Q(\mathbf{d})$  em (57) por sua aproximação em (58), obtém-se:

$$I = (2\pi)^{q/2} \left| -Q''(\hat{\mathbf{d}}) \right|^{-1/2} e^{Q(\hat{\mathbf{d}})}.$$

Esse procedimento pode ser aplicado para permitir maximizar a função de verossimilhança do MLGM, pois cada integral na função (55) é proporcional a uma integral da forma (57) para funções  $Q(\hat{\mathbf{d}})$  dadas por:

$$Q(\hat{\mathbf{d}}) = \phi^{-1} \sum_{t=1}^{n_i} [y_{it}(\mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{it}\mathbf{d})] - \frac{1}{2}\mathbf{d}'\mathbf{G}^{-1}\mathbf{d},$$

as quais permitem a aplicação do método de Laplace. Como a moda  $\hat{\mathbf{d}}$  de  $Q$  depende dos parâmetros desconhecidos,  $\boldsymbol{\beta}$ ,  $\phi$  e  $\mathbf{G}$ , a cada iteração da maximização numérica da função de verossimilhança,  $\hat{\mathbf{d}}$  é calculado condicionalmente aos valores correntes para estimar esses parâmetros.

ii. Aproximação dos dados

O procedimento da aproximação dos dados baseia-se na sua decomposição nos termos média e erro, com uma expansão em série de Taylor da média, que é uma função não linear do preditor linear. Sem perda de generalidade, considere a decomposição:

$$Y_{it} = \mu_{it} + \varepsilon_{it} = g^{-1}(\mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{it}\mathbf{d}_i) + \varepsilon_{it}, \quad (59)$$

em que  $g^{-1}(\cdot)$  é a inversa da função de ligação e o termo do erro tem distribuição apropriada com variância igual a  $\text{Var}(Y_{it} \mid \mathbf{d}_i) = \phi V(\mu_{it})$ . As técnicas conhecidas como quase-verossimilhança penalizada (QVP) e quase-verossimilhança marginal (QVM) podem ser implementadas para a aproximação de  $\mu_{it}$  na expressão (59). No procedimento da QVP há uma expansão em série de Taylor de  $g^{-1}(\mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{it}\mathbf{d}_i)$  ao redor das estimativas correntes dos efeitos fixos ( $\hat{\boldsymbol{\beta}}$ ) e aleatórios ( $\hat{\mathbf{d}}_i$ ). Esse desenvolvimento leva a uma aproximação dos dados na forma:

$$\mathbf{Y}_i^* \equiv \hat{\mathbf{V}}_i^{-1}(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i) + \mathbf{X}_i\hat{\boldsymbol{\beta}} + \mathbf{Z}_i\hat{\mathbf{d}}_i \approx \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{d}_i + \boldsymbol{\varepsilon}_i^*, \quad (60)$$

em que  $\boldsymbol{\varepsilon}_i^* = \hat{\mathbf{V}}_i^{-1}\boldsymbol{\varepsilon}_i$ , sendo  $\hat{\mathbf{V}}_i$  a matriz diagonal com elementos  $[V(\hat{\mu}_{it})]$  e  $\mathbf{X}_i$ ,  $\mathbf{Z}_i$  matrizes conhecidas do delineamento.

A aproximação dos dados (60) permite um algoritmo para ajustar o MLGM. A partir de valores iniciais para  $\boldsymbol{\beta}$ ,  $\phi$  e  $\mathbf{G}$  na função de verossimilhança marginal, estimativas empíricas de Bayes são calculadas para  $\mathbf{d}_i$  e para os pseudo-dados  $\mathbf{Y}_i^*$ . Então o modelo misto linear (60) é ajustado, levando a valores atualizados para  $\boldsymbol{\beta}$ ,  $\phi$  e  $\mathbf{G}$ . O procedimento é repetido até convergência.

Alternativamente, no procedimento da QVM, a expansão em série de Taylor da média em (59) é feita ao redor das estimativas de  $\hat{\boldsymbol{\beta}}$  para os efeitos fixos e ao redor de  $\mathbf{d}_i = \mathbf{0}$  para os efeitos aleatórios, ou seja, usa-se a forma  $g^{-1}(\mathbf{x}'_{it}\boldsymbol{\beta})$  ao invés de  $g^{-1}(\mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{it}\mathbf{d}_i)$ . Os pseudo-dados têm a forma  $\mathbf{Y}_i^* \equiv \hat{\mathbf{V}}_i^{-1}(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i) + \mathbf{X}_i\hat{\boldsymbol{\beta}}$  e satisfazem a aproximação do modelo linear misto:

$$\mathbf{Y}_i^* \approx \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{d}_i + \boldsymbol{\varepsilon}_i^*, \quad (61)$$

assim, o ajuste do MLGM é feito pela iteração entre o cálculo dos pseudo-dados e o ajuste do modelo (61) para esses dados.

As estimativas dadas por QVP e QVM, podem ser obtidas otimizando a função de quase-verossimilhança, que envolve os momentos condicionais aos efeitos aleatórios de primeira e segunda ordens. Para o ajuste do modelo linear misto aos pseudo-dados, pode-se usar tanto a máxima verossimilhança quanto a máxima verossimilhança restrita. Maiores detalhes sobre esses procedimentos podem ser encontrados em Breslow e Clayton (1993).



Registra-se que esses procedimentos encontram-se implementados no *software* SAS, através do procedimento GLIMMIX.

### iii. Aproximação da Integral

A aproximação da integral, isto é, a aplicação de técnicas de integração numéricas, também pode ser utilizada, especialmente quando as técnicas anteriores falham. No contexto do MLGM, Pinheiro e Bates (2000) destacam que as regras da quadratura gaussiana e quadratura gaussiana adaptativa são as mais usuais. A idéia básica dessas técnicas é aproximar uma integral da forma:

$$\int f(z)\phi(z)dz, \quad (62)$$

em que  $f(z)$  é uma função conhecida e  $\phi(z)$  a função de densidade da distribuição normal padrão, a partir da seleção de um conjunto de pontos e, sob eles, a construção de uma soma ponderada de funções, ou seja:

$$\int f(z)\phi(z)dz \approx \sum_{q=1}^Q w_q f(z_q), \quad (63)$$

sendo  $z_q$  os pontos de quadratura,  $w_q$  os pesos e  $Q$  a ordem da aproximação. Em geral, quanto maior for  $Q$ , melhor é a aproximação. Basicamente, o método da quadratura gaussiana adaptativa requer menos pontos que a quadratura gaussiana clássica, porém demanda mais tempo, pois requer o cálculo de  $\hat{z}$ , a moda de  $\ln[f(z)\phi(z)]$ .

De qualquer forma, essas duas técnicas quando aplicadas no ajuste do MLGM, permitem uma aproximação para a contribuição individual na função de verossimilhança. Para isso, a princípio padronizam-se os efeitos aleatórios a fim que de que tenham matriz de covariância identidade. Assim, se  $\boldsymbol{\gamma}_i = \mathbf{G}^{-1}\mathbf{d}_i$ , então  $\boldsymbol{\gamma}_i$  tem distribuição normal com média  $\mathbf{0}$  e covariância  $\mathbf{I}$ , e o preditor linear torna-se  $\eta_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{z}_{it}\mathbf{G}^{1/2}\boldsymbol{\gamma}_i$ . A contribuição para o  $i$ -ésimo indivíduo é:

$$\begin{aligned} f_i(\mathbf{y}_i | \boldsymbol{\beta}, \mathbf{G}, \phi) &= \int \prod_{t=1}^{n_i} f_{it}(y_{it} | \mathbf{d}_i, \boldsymbol{\beta}, \phi) f(\mathbf{d}_i | \mathbf{G}) d\mathbf{d}_i \\ &= \int \prod_{t=1}^{n_i} f_{it}(y_{it} | \mathbf{d}_i, \boldsymbol{\beta}, \phi) f(\boldsymbol{\gamma}_i) d\boldsymbol{\gamma}_i. \end{aligned} \quad (64)$$

Note que a integral dada em (64) tem forma análoga a (62) e, pode ser aproximada pela quadratura gaussiana adaptativa, como na expressão (63). Os métodos de aproximação da integral podem ser implementados no SAS, usando o procedimento NLMIXED e no R, usando o pacote glmML.

A literatura freqüentemente apresenta os modelos marginais, de transição e de efeitos aleatórios como três propostas independentes e de objetivos específicos. Embora a interpretação dos parâmetros e inferências decorrentes difiram de um modelo para outro, na prática, essa segmentação estrutural dos três modelos não é tão restritiva quanto parece. Conforme já exposto na seção (2.1.3.5), o método de ajuste dos modelos de transição é uma extensão do processo dos modelos marginais, com uma sensível mudança no preditor linear. Da mesma forma, as estruturas do MLGM e a do modelo de transição podem ser combinadas, num único modelo. Apesar da idéia de modelos de transição com efeitos aleatórios não ser usual (em geral são modelos de efeitos fixos), essa técnica é considerada nos trabalhos de Korn e Whittemore (1979) e Stiratelli, Laird e Ware (1984).

No trabalho de Korn de Whittemore (1979), as observações para o  $i$ -ésimo indivíduo provêm de uma série temporal longa (observações diárias durante um período de oito meses). Com isso, os autores propõem uma análise em duas etapas. Na etapa 1 é modelada a variabilidade intra-unidade experimental ( $\sigma_i^2$ ) e estimados os parâmetros individuais ( $\beta_i$ ), ajustando-se, para cada indivíduo, um modelo logístico de transição:

$$\eta_{it} = \beta_{0i} + \beta_{1i}y_{i(t-1)} + \sum_{j=2}^p \beta_{ji}x_{jti},$$

em que  $x_{jti}$  representa a  $j$ -ésima covariável na  $t$ -ésima ocasião para o  $i$ -ésimo indivíduo. Na etapa 2, modela-se a variação entre os indivíduos, isto é, entre os coeficientes das curvas individuais, adotando-se a suposição de que os  $\beta_i$  são uma amostra de uma distribuição normal com média  $\beta$  e variância  $\sigma^2$ . Os autores usam o método de máxima verossimilhança para obter as estimativas dos parâmetros e obtêm a variância assintótica de  $\hat{\beta}$ , invertendo-se a matriz da informação observada de Fisher. Se se supõe que  $\beta_i \equiv \beta$  para todos os indivíduos, então  $\hat{\beta} = \sum w_i \hat{\beta}_i$ , com pesos  $w_i = (\sigma_i^2 + \sigma^2)^{-1} [\sum (\sigma_i^2 + \sigma^2)^{-1}]^{-1}$ , assumindo  $\hat{\sigma}^2 = 0$ . Quando essa suposição não é rejeitada, os autores consideram o modelo de efeitos fixos.

Stiratelli, Laird e Ware (1984) usam os mesmos dados mas propõem uma análise

única, na qual assume-se que os coeficientes de regressão têm uma distribuição normal multivariada na população e usam um método baseado no algoritmo EM para aproximar as estimativas de máxima verossimilhança dos parâmetros do modelo logístico misto. A proposta desses autores se aproxima da linha metodológica usual dos MLGM, a menos pelo fato de que os autores fazem uma aproximação das equações escores de tal forma que as integrais possam ser evitadas. Diggle et al. (2002) comentam que a idéia central dessa aproximação é usar modas condicionais ao invés de médias condicionais nas equações escores para  $\beta$ , o que equivale a aproximar a distribuição condicional de  $\mathbf{d}_i \mid \mathbf{y}_i$  por uma distribuição normal com a mesma moda e curvatura. Usando moda no lugar da média, troca-se a integração com uma otimização que pode ser incorporada no passo M do algoritmo EM.

Esses dois trabalhos são exemplos de como a estrutura dos modelos de transição pode ser adaptada para considerar a inclusão de efeitos aleatórios. Apesar de que, conforme salientam Ware, Lipsitz e Speizer (1988), os resultados dessas propostas podem não ter boa performance quando o número de observações é pequeno. O trabalho de Korn e Whittemore (1979), por exemplo, é inviável, pois, em geral, não se dispõe de um grande número de observações por indivíduo de tal forma que se possa ajustar um modelo para cada um.

Assim, as abordagens mais contemporâneas sobre o MLGM podem ser usadas para permitir modelos de transição mistos, não sendo tão restritivas quanto ao número de observações e, também, levando-se em conta as facilidades dos recursos computacionais.

## 2.2 Metodologia

### 2.2.1 Material

Para ilustrar a aplicação dos métodos descritos, usa-se um conjunto de dados da área epidemiológica. Os dados desse exemplo referem-se a um ensaio clínico, envolvendo doença respiratória, extraído do trabalho de Koch et al. (1990). O estudo envolve 111 pacientes com problemas respiratórios de dois centros médicos (56 do Centro 1 e 55 do Centro 2), nos quais os indivíduos são aleatorizados para receber um tratamento ativo ou placebo. O objetivo do estudo é avaliar o efeito da droga ativa sob a condição respiratória.

No trabalho original de Koch et al. (1990), o status respiratório de cada paciente

é qualificado de acordo com um conjunto de cinco categorias ordinais, que expressam uma classificação da resposta menos favorável a mais favorável (0=terrível, 1=ruim, 2=razoável, 3=bom, 4=excelente). Essa escala é usada para avaliar a saúde (respiração) dos indivíduos no início do estudo (*baseline*) e em mais quatro visitas igualmente espaçadas, durante o período no qual os tratamentos são administrados. As demais variáveis explicativas, sexo e idade (em anos) são obtidas na ocasião de entrada dos indivíduos no estudo.

A simplificação das classificações observadas para medidas dicotômicas são 0 (terrível ou ruim ou razoável) e 1 (bom ou excelente). Com essa classificação, esses dados têm sido usados na literatura como exemplo para abordagem de modelos marginais por meio das EEGs, seguindo a proposta de Liang e Zeger (1986). Portanto, para atender aos objetivos deste trabalho, utiliza-se a classificação dicotômica. Também, por simplicidade, faz-se referência ao estado de saúde (respiração do paciente) como bom (1) ou ruim (0). Maiores detalhes sobre esses dados e respectiva documentação também encontram-se disponíveis nos pacotes *gee* e *geepack* do *software* R.

### 2.2.2 Métodos

Para a situação descrita, as respostas resultantes da condição respiratória nos tempos de observação são variáveis binárias. Para caracterizar as transições nos tempos, assume-se que  $Y_{it} \mid \mathbf{h}_{it}, \mathbf{x}_{it}$  segue um modelo linear generalizado de transição, que permite modelar as probabilidades condicionais do estado de saúde do indivíduo,  $P[Y_{it} = y_{it} \mid \mathbf{h}_{it}, \mathbf{x}_{it}]$ , dada sua história ( $\mathbf{h}_{it}$ ) e os efeitos das variáveis explicativas ( $\mathbf{x}_{it}$ ). A estimação dessas probabilidades pode ser feita através de  $\eta_{it} = g(\mu_{it}^C)$ , sendo  $g(\cdot)$  uma função adequada que faz a ligação entre a parte aleatória e a parte sistemática (respostas prévias, efeitos de tratamentos e demais variáveis explicativas). Além disso, há diversas formas de se especificar,  $\eta_{it}$ , de tal forma a incorporar uma função das respostas prévias (ordem da cadeia), bem como os efeitos das variáveis explicativas. Nesse contexto, adotam-se os seguintes procedimentos metodológicos:

1. Ajustar modelos de transição de Markov a partir da proposta de Ware, Lipsitz e Speizer (1988), especificando as matrizes de probabilidades para cada transição;
2. Verificar se o processo é estacionário e em caso de não rejeição dessa hipótese, ajustar

o modelo de transição estacionário pela proposta de Diggle et al. (2002);

3. Efetuar diagnósticos desses modelos e estudar formas adequadas para comparar tratamentos;
4. Estimar as probabilidades de transição considerando um modelo estocástico com intercepto aleatório.

### 2.2.2.1 Matriz das probabilidades de transição da cadeia de Markov para dados binários

A distribuição condicional de  $Y_{it}$ , para  $t = 2, 3, \dots, n_i$  segue um modelo linear generalizado de transição, ou seja,

$$f(y_{it} | \mathbf{h}_{it}) = \exp \left\{ y_{it} \ln \left( \frac{\mu_{it}^C}{1 - \mu_{it}^C} \right) + \ln(1 - \mu_{it}^C) \right\}, \quad (i = 1, 2, \dots, N)$$

para o qual:

$$g(\mu_{it}^c) = \text{logito}(\mu_{it}^c) = \ln \left( \frac{\mu_{it}^c}{1 - \mu_{it}^c} \right) \quad \text{e} \quad v(\mu_{it}^c) = \mu_{it}^c(1 - \mu_{it}^c),$$

são, respectivamente, as funções de ligação e de variância do modelo. Para o caso em que a ordem da cadeia é 1, tem-se como preditor linear:

$$\eta_{it} = \text{logito} [P(Y_{it} = b | Y_{i(t-1)} = a)] = \ln \left( \frac{\pi_{ab}(t)}{1 - \pi_{ab}(t)} \right) = \mathbf{x}'_{it} \boldsymbol{\beta} + \alpha y_{i(t-1)}. \quad (65)$$

A caracterização das matrizes de probabilidades de transição,  $2 \times 2$ , com elementos  $\pi_{ab}(t) = P(Y_{it} = b | Y_{i(t-1)} = a)$ , em que  $a$  e  $b$  assumem valores em  $S = \{0, 1\}$ , pode ser feita a partir da equação definida em (65), por meio de:

$$\pi_{ab}(t) = \frac{\exp(\eta_{it})}{1 + \exp(\eta_{it})},$$

avaliando-se  $\eta_{it}$  primeiramente com relação a  $Y_{i(t-1)} = 0$  e, posteriormente a  $Y_{i(t-1)} = 1$ .

Logo:

$$\mathbf{P}(t) = \begin{pmatrix} \pi_{00}(t) = \frac{1}{1 + \exp(\mathbf{x}'_{it} \boldsymbol{\beta})} & \pi_{01}(t) = \frac{\exp(\mathbf{x}'_{it} \boldsymbol{\beta})}{1 + \exp(\mathbf{x}'_{it} \boldsymbol{\beta})} \\ \pi_{10}(t) = \frac{1}{1 + \exp(\mathbf{x}'_{it} \boldsymbol{\beta} + \alpha)} & \pi_{11}(t) = \frac{\exp(\mathbf{x}'_{it} \boldsymbol{\beta} + \alpha)}{1 + \exp(\mathbf{x}'_{it} \boldsymbol{\beta} + \alpha)} \end{pmatrix},$$

em que  $\pi_{00}(t) = 1 - \pi_{01}(t)$  e  $\pi_{10}(t) = 1 - \pi_{11}(t)$ .

Note que para especificar a matriz de probabilidades de transição, avaliam-se duas regressões separadamente. Diggle et al. (2002) mostram que uma forma mais concisa é dada por:

$$\eta_{it} = \text{logito} [P(Y_{it} | Y_{i(t-1)})] = \ln \left( \frac{\pi_{ab}(t)}{1 - \pi_{ab}(t)} \right) = \mathbf{x}'_{it}\boldsymbol{\beta} + y_{i(t-1)}\mathbf{x}'_{it}\boldsymbol{\alpha}, \quad (66)$$

sendo a equação (65) um caso particular da equação (66), que inclui todas as interações entre a resposta prévia e as variáveis explicativas, resultando em vantagens em termos de inferência.

Os resultados apresentados partem da pressuposição de que a função de ligação adequada é a logística. Para os casos em que essa função não seja satisfatória, pode-se optar por outras funções de ligação, como a probito e complemento *log-log*. Para a função probito, tem-se:

$$\eta_{it} = \text{probito}(\pi_{ab}(t)) = \Phi^{-1}(\pi_{ab}(t)) = \mathbf{x}'_{it}\boldsymbol{\beta} + y_{i(t-1)}\mathbf{x}'_{it}\boldsymbol{\alpha},$$

com probabilidades de transição especificadas por:

$$\pi_{ab}(t) = \Phi(\mathbf{x}'_{it}\boldsymbol{\beta} + y_{i(t-1)}\mathbf{x}'_{it}\boldsymbol{\alpha}),$$

em que  $\Phi(\cdot)$  representa a função de distribuição normal padrão. Assim, a matriz estocástica pode ser representada por:

$$\mathbf{P}(t) = \begin{pmatrix} \pi_{00}(t) = 1 - \Phi(\mathbf{x}'_{it}\boldsymbol{\beta}) & \pi_{01}(t) = \Phi(\mathbf{x}'_{it}\boldsymbol{\beta}) \\ \pi_{10}(t) = 1 - \Phi(\mathbf{x}'_{it}\boldsymbol{\beta} + y_{i(t-1)}\mathbf{x}'_{it}\boldsymbol{\alpha}) & \pi_{11}(t) = \Phi(\mathbf{x}'_{it}\boldsymbol{\beta} + y_{i(t-1)}\mathbf{x}'_{it}\boldsymbol{\alpha}) \end{pmatrix}.$$

E, para um modelo com função de ligação complemento *log-log*, especifica-se:

$$\eta_{it} = \ln(-\ln(1 - \pi_{ab}(t))) = \mathbf{x}'_{it}\boldsymbol{\beta} + y_{i(t-1)}\mathbf{x}'_{it}\boldsymbol{\alpha},$$

levando a:

$$\mathbf{P}(t) = \begin{pmatrix} \pi_{00}(t) = \exp(-\exp(\mathbf{x}'_{it}\boldsymbol{\beta})) & \pi_{01}(t) = 1 - \exp(-\exp(\mathbf{x}'_{it}\boldsymbol{\beta})) \\ \pi_{10}(t) = \exp(-\exp(\mathbf{x}'_{it}\boldsymbol{\beta} + y_{i(t-1)}\mathbf{x}'_{it}\boldsymbol{\alpha})) & \pi_{11}(t) = 1 - \exp(-\exp(\mathbf{x}'_{it}\boldsymbol{\beta} + y_{i(t-1)}\mathbf{x}'_{it}\boldsymbol{\alpha})) \end{pmatrix}.$$

Uma extensão para um modelo de ordem  $q$  tem como preditor linear:

$$\eta_{it} = g[P(Y_{it} = b \mid \mathbf{h}_{it})] = \mathbf{x}'_{it}\boldsymbol{\beta} + \sum_{r=1}^q \alpha_r y_{i(t-r)} \mathbf{x}_{it},$$

sendo a cadeia de ordem 1, isto é  $f_r(\mathbf{h}_{it}, \boldsymbol{\alpha}) = \alpha_r y_{i(t-r)} \mathbf{x}_{it}$  com  $s = q = 1$ , um caso específico. Em particular, uma cadeia de Markov de alcance 2, com função de ligação canônica tem como preditor linear:

$$\begin{aligned} \eta_{it} &= \text{logito} [P(Y_{it} = c \mid Y_{i(t-2)} = b, Y_{i(t-1)} = a)] \\ &= \mathbf{x}'_{it}\boldsymbol{\beta} + y_{i(t-1)}\mathbf{x}_{it}\boldsymbol{\alpha}_1 + y_{i(t-2)}\mathbf{x}_{it}\boldsymbol{\alpha}_2 + y_{i(t-1)}y_{i(t-2)}\mathbf{x}_{it}\boldsymbol{\alpha}_3, \end{aligned} \quad (67)$$

com  $a, b, c \in S = \{0, 1\}$ . Assim, considerando o preditor linear (67) e todas as combinações para  $\mathbf{h}_{it} = (y_{i(t-2)}, y_{i(t-1)})$ , é possível especificar a seguinte matriz de transição de segunda ordem:

$$\mathbf{P}(t) = \begin{pmatrix} \pi_{000}(t) = \frac{1}{1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta})} & \pi_{001}(t) = \frac{\exp(\mathbf{x}'_{it}\boldsymbol{\beta})}{1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta})} \\ \pi_{010}(t) = \frac{1}{1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{x}_{it}\boldsymbol{\alpha}_1)} & \pi_{011}(t) = \frac{\exp(\mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{x}_{it}\boldsymbol{\alpha}_1)}{1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{x}_{it}\boldsymbol{\alpha}_1)} \\ \pi_{100}(t) = \frac{1}{1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{x}_{it}\boldsymbol{\alpha}_2)} & \pi_{101}(t) = \frac{\exp(\mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{x}_{it}\boldsymbol{\alpha}_2)}{1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{x}_{it}\boldsymbol{\alpha}_2)} \\ \pi_{110}(t) = \frac{1}{1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{x}_{it}\boldsymbol{\alpha}_1 + \mathbf{x}_{it}\boldsymbol{\alpha}_2 + \mathbf{x}_{it}\boldsymbol{\alpha}_3)} & \pi_{111}(t) = \frac{\exp(\mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{x}_{it}\boldsymbol{\alpha}_1 + \mathbf{x}_{it}\boldsymbol{\alpha}_2 + \mathbf{x}_{it}\boldsymbol{\alpha}_3)}{1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{x}_{it}\boldsymbol{\alpha}_1 + \mathbf{x}_{it}\boldsymbol{\alpha}_2 + \mathbf{x}_{it}\boldsymbol{\alpha}_3)} \end{pmatrix}.$$

De modo análogo pode-se proceder para outras funções de ligação, inclusive para cadeias de alcance maior do que 2. Deve ser observado que como a matriz de probabilidades de transição depende das variáveis explicativas,  $\mathbf{P}(t)$  pode variar entre os indivíduos e, cadeias de alcance maior podem se tornar extremamente complexas, em virtude do número de parâmetros envolvidos no processo de estimação. Para o ajuste dos modelos usa-se o método da máxima verossimilhança, via processo iterativo, seguindo procedimentos usuais

dos MLGs. O teste da razão de verossimilhanças é usado para verificar o alcance da cadeia.

A hipótese nula de interesse é:

$$H_0: \text{A cadeia é de } (q - 1)\text{-ésima ordem;}$$

contra a alternativa de que a cadeia é de  $q$ -ésima ordem. Denotem  $\ell(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \mathbf{y})_{(q-1)}$  e  $\ell(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \mathbf{y})_{(q)}$  os logaritmos das funções de verossimilhança maximizadas sob as condições de ordem  $(q - 1)$  e  $q$ , respectivamente. Tem-se, assintoticamente que:

$$\lambda = 2(\ell(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \mathbf{y})_{(q)} - \ell(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \mathbf{y})_{(q-1)}) \sim \chi_1^2,$$

sendo necessário o mesmo número de observações para os ajustes nos dois casos. Assim, para o modelo de menor alcance, as  $q$  primeiras observações devem ser omitidas.

Uma vez verificada a ordem da cadeia, estimam-se as probabilidades de transição a partir das estimativas dos parâmetros do modelo especificado. Quanto aos procedimentos computacionais, *softwares* padrões podem ser utilizados para o ajuste desses modelos, como os procedimentos **logistic** e **genmod** do SAS ou os pacotes, **glm**, **gee** e **geepack** do R. Para o caso estacionário, há necessidade de organizar o conjunto de dados de tal forma a incorporar as respostas prévias como covariáveis extras, ignorando as observações referentes à ordem da cadeia. Para esses casos, usa-se a macro **DROPOUT**, proposta por Molenberghs e Verbeke (2005), desenvolvida no programa SAS. Essa macro gera a partir do conjunto de dados original, um arquivo novo de observações, com colunas adicionais para as respostas prévias. As observações referentes à ordem da cadeia, não são lidas e são consideradas como dados perdidos.

### 2.2.2.2 Testes para comparação de tratamentos

As estruturas da matriz de probabilidades de transição descritas na seção anterior, permitem descrever as mudanças ou transições dos indivíduos no tempo. Quando se tem um experimento ou estudo envolvendo tratamentos é objetivo básico avaliar qual tratamento é o mais eficiente do ponto vista prático. As probabilidades de transição podem ser usadas na construção de testes que auxiliem a tomada de decisão. Os testes propostos são assintóticos e, portanto, podem não ser válidos, para pequenas amostras.



Considere a situação do estudo epidemiológico, em que os estratos placebo e ativo são amostras independentes. Para um estado prévio  $a$ , com  $a \in S = \{0, 1\}$ , as probabilidades de transição podem ser descritas pela Tabela 6.

Tabela 6 - Distribuição das probabilidades de transição, considerando os grupos placebo e ativo (tratamento), com estado prévio fixado  $Y_{(t-1)} = a$ , num modelo de Markov

Tratamentos	$Y_{it}$	
	0	1
Placebo (P)	$\pi_{a0}(P)$	$\pi_{a1}(P)$
Ativo (A)	$\pi_{a0}(A)$	$\pi_{a1}(A)$

Para  $Y_{(t-1)} = 0$ , as hipóteses de interesse podem ser formuladas em termos de transições, considerando permanência no estado 0 ( $\pi_{00}$ ) ou mudanças para o estado 1 ( $\pi_{01}$ ). Se “0” denota estado de saúde ruim e “1” estado de saúde bom, do ponto de vista prático, parece sensato formular as hipóteses:

$$\text{Hipóteses}^{(1)} : \text{mudanças para o estado 1} \begin{cases} H_0 : \pi_{01}(A) - \pi_{01}(P) = 0; \\ H_a : \pi_{01}(A) - \pi_{01}(P) > 0. \end{cases}$$

No entanto, há outras formulações possíveis e equivalentes. Tendo em vista que  $\pi_{01} = 1 - \pi_{00}$ , as hipóteses <sup>(1)</sup> são equivalentes a:

$$\text{Hipóteses}^{(2)} : \text{permanência no estado 0} \begin{cases} H_0 : \pi_{00}(P) - \pi_{00}(A) = 0; \\ H_a : \pi_{00}(P) - \pi_{00}(A) > 0. \end{cases}$$

Do mesmo modo, podem-se formular hipóteses considerando  $Y_{(t-1)} = 1$ . Por outro lado, quando se trabalha com respostas binárias, um dos anseios do pesquisador é fazer inferências com relação à razão de chances, principalmente se o estudo é retrospectivo. Com base na Tabela 6, a razão de chances associada a  $Y_{(t-1)} = a$  é dada por:

$$\psi_{Y_{(t-1)}=a} = \frac{\pi_{a0}(P) \times \pi_{a1}(A)}{\pi_{a0}(A) \times \pi_{a1}(P)},$$

sendo  $\psi_{Y_{(t-1)}=a}$  estimada pelo princípio da invariância dos estimadores de máxima verossimilhança, ou seja  $\hat{\psi}_{Y_{(t-1)}=a} = \frac{\hat{\pi}_{00}(P) \times \hat{\pi}_{01}(A)}{\hat{\pi}_{00}(A) \times \hat{\pi}_{01}(P)}$ . Desse modo, para um estado fixo  $a$ , tem-se que:

- i.  $\psi = 1$  indica que há independência das probabilidades de transição em relação aos tratamentos, ou seja, isso equivale à hipótese  $\pi_{ab}(P) = \pi_{ab}(A)$ ;
- ii.  $\psi > 1$  indica que os indivíduos do grupo placebo estão mais propensos à transição para o estado 0 do que os indivíduos do grupo ativo ( $\pi_{a0}(P) > \pi_{a0}(A)$ ) ou equivalentemente que os indivíduos do grupo ativo estão mais propensos à transição para o estado 1 do que os indivíduos do grupo placebo ( $\pi_{a1}(A) > \pi_{a1}(P)$ );
- iii.  $\psi \in (0, 1)$  indica que os indivíduos do grupo placebo estão menos propensos à transição para o estado 0 que os indivíduos do grupo ativo ( $\pi_{a0}(P) < \pi_{a0}(A)$ ) ou equivalentemente que os indivíduos do grupo ativo estão menos propensos à transição para o estado 1 do que os indivíduos do grupo placebo ( $\pi_{a1}(A) < \pi_{a1}(P)$ ).

Diante dessa explicação, é razoável também formular as seguintes hipóteses:

$$\text{Hipóteses}^{(3)} : \text{em termos de razão de chances} \begin{cases} H_0 : \psi = 1; \\ H_a : \psi > 1. \end{cases}$$

Como em uma cadeia de Markov há de se considerar os estados prévios, uma prática inicial é verificar se há homogeneidade da razão de chances em relação às respostas prévias, que definem a ordem da cadeia. Assim em uma cadeia de ordem 1, a hipótese de homogeneidade é:

$$H_0 : \psi_{Y_{(t-1)=0}} = \psi_{Y_{(t-1)=1}}. \quad (68)$$

Essa hipótese pode ser verificada por meio do teste homogeneidade de razão de chances. No entanto, uma forma mais operacional e elegante de se testar essa homogeneidade é por meio da definição do modelo de transição de Markov com ligação canônica. Assim, considerando o modelo estocástico:

$$\eta = \ln \left( \frac{\pi_{ab}}{1 - \pi_{ab}} \right) = \beta_0 + \beta_1 \text{tratamento} + \alpha y_{(t-1)} + \gamma(y_{(t-1)}) * \text{tratamento}, \quad (69)$$

a razão de chances associada às respostas prévias pode ser obtida em termos dos coeficientes desse modelo. De fato, para os estados  $Y_{(t-1)} = 0$  e  $Y_{(t-1)} = 1$  tem-se respectivamente:

$$\psi_{Y_{(t-1)=0}} = \exp(\beta_1) \quad \text{e} \quad \psi_{Y_{(t-1)=1}} = \exp(\beta_1 + \gamma),$$

de modo que o teste da hipótese (68) equivale ao teste  $H_0 : \gamma = 0$  pelo modelo definido em (69). Esse teste pode ser feito pela razão de verossimilhanças ou pela diferença de desvios, considerando os modelos encaixados sem e com interação:

$$H_0 : \gamma = 0 \begin{cases} M(1) : \eta_{(1)} = \beta_0 + \beta_1 \text{tratamento} + \alpha y_{(t-1)} \\ M(2) : \eta_{(2)} = \beta_0 + \beta_1 \text{tratamento} + \alpha y_{(t-1)} + \gamma(y_{(t-1)}) * \text{tratamento} \\ \text{Desvio } (M(1)) - \text{Desvio } (M(2)) \sim \chi^2_{(1)}. \end{cases}$$

Assim, a hipótese de homogeneidade de razão de chances é equivalente à hipótese de não interação entre tratamento e resposta prévia. Para um modelo de segunda ordem, a hipótese de homogeneidade de razão de chances em relação à história,  $\mathbf{h}_t = (y_{(t-2)}, y_{(t-1)})$  é:

$$H_0 : \psi_{\mathbf{h}_t=(0,0)} = \psi_{\mathbf{h}_t=(0,1)} = \psi_{\mathbf{h}_t=(1,0)} = \psi_{\mathbf{h}_t=(1,1)}, \quad (70)$$

sendo avaliada a partir do modelo de transição:

$$\eta = \ln \left( \frac{\pi_{abc}}{1 - \pi_{abc}} \right) = \beta_0 + \beta_1 x + \alpha_1 y_{(t-1)} + \alpha_2 y_{(t-2)} + \gamma_1 y_{(t-1)} x + \gamma_2 y_{(t-2)} x + \gamma_3 y_{(t-1)} y_{(t-2)} x,$$

em que  $x$  é a covariável associada ao tratamento. Desse modo, a hipótese (70) é equivalente à hipótese de que  $(\gamma_1, \gamma_2, \gamma_3)' = (0, 0, 0)'$ . Uma extensão para um modelo de Markov de alcance  $q$ , teria como hipótese natural para a homogeneidade de razão de chances,  $H_0 : \boldsymbol{\gamma} = \mathbf{0}$ , sendo  $\boldsymbol{\gamma}$  o vetor de parâmetros de todas as interações.

Portanto, a ausência de interação entre as respostas prévias e tratamento significa que a associação entre a resposta no tempo  $t$  e o tipo de tratamento (covariável) é independente da história da cadeia. Contudo, sendo o modelo de transição de Markov, deve haver efeito de  $\mathbf{h}_t$  sob as probabilidades de transição. Logo, essas probabilidades podem estar em patamares diferentes de acordo com o tratamento recebido e estado na ocasião  $(t - 1)$  ou em sua história,  $(\mathbf{h}_t)$ , se a cadeia é de alcance maior do que 1. Faz-se necessário, então, um teste para comparar os tratamentos.

### 2.2.2.2.1 Teste da razão de chances

A proposta de comparar os tratamentos pela razão de chances pode ser vista como uma extensão natural da estatística original de Mantel e Hanszel (1959), que avalia

a não existência de associação entre as linhas (independentes) e as colunas em tabelas de contingência  $2 \times 2$ , mantendo-se o efeito da estratificação. A idéia básica do teste, que tem como hipótese principal  $H_0 : \psi = 1$ , é a distribuição normal assintótica de uma distribuição condicional, que é descrita a seguir.

Tabela 7 - Estrutura usual de uma tabela de contingência  $2 \times 2$  para o teste de Mantel-Hanszel

Estratos	Resposta		Total marginal
	1	2	
A	$y_1$	$n_1 - y_1$	$n_1$
B	$y_2$	$n_2 - y_2$	$n_2$
Total marginal	$m$	$n - m$	$n$

Considere a Tabela 7, em que são amostrados  $n_1$  e  $n_2$  indivíduos nos estratos A e B, com  $y_1$  e  $y_2$  respostas favoráveis à primeira categoria, respectivamente. Sob essa estrutura, a distribuição condicional  $(y_1 | m, \psi)$  produz um modelo pertencente à família de distribuições hipergeométricas não centrais:

$$f(y_1 | m, \psi) = \frac{\binom{n_1}{y_1} \binom{n_2}{m - y_1} \psi^{y_1}}{\sum_s \binom{n_1}{s} \binom{n_2}{m - s} \psi^s}, \quad (71)$$

em que  $s$  varia entre o máximo  $(0, m - n_2)$  e o mínimo  $(n_1, m)$ . Em particular, para  $\psi = 1$ , o modelo (71) reduz-se à distribuição hipergeométrica central:

$$f(y_1 | m, \psi = 1) = \frac{\binom{n_1}{y_1} \binom{n_2}{y_2}}{\binom{n_1 + n_2}{m}}, \quad (72)$$

com média e variância dadas por:

$$E(Y_1 | m, \psi = 1) = \frac{mn_1}{n} \quad \text{e} \quad \text{Var}(Y_1 | m, \psi = 1) = \frac{n_1 n_2 m (n - m)}{n^3}.$$

Quando  $n_1$ ,  $n_2$ ,  $m$  e  $n - m$  são suficientemente grandes, a distribuição (72) tende à distribuição normal e conseqüentemente:

$$Z = \frac{y_1 - \frac{mn_1}{n}}{\sqrt{\frac{n_1 n_2 m (n - m)}{n^3}}} \sim N(0, 1), \quad (73)$$

é uma estatística apropriada para avaliar a hipótese de que  $\psi = 1$ .

No caso dos modelos de transição, se há homogeneidade da razão de chances em relação à história do processo, então é possível adaptar a estatística do teste de Mantel-Hanszel, para confrontar as hipóteses<sup>(3)</sup>  $H_0 : \psi = 1$  e  $H_a : \psi > 1$ . Para um processo estacionário de primeira ordem é possível descrever as freqüências de transições através de duas tabelas de contingência  $2 \times 2$ , uma para cada tipo de resposta prévia, que tem aspecto similar à Tabela 7.

Tabela 8 - Tabela de contingência  $2 \times 2$ , considerando os estratos de tratamento e as freqüências de transições para os estados 0 e 1, dado  $Y_{(t-1)} = a$ ,  $a \in \{0, 1\}$

Tratamento	$Y_t$		Total marginal
	0	1	
Placebo	$n_{a0(j)}^*(P)$	$n_{a1(j)}^*(P)$	$n_{a.(j)}^*(P)$
Ativo	$n_{a0(j)}^*(A)$	$n_{a1(j)}^*(A)$	$n_{a.(j)}^*(A)$
Total marginal	$n_{.0(j)}^*$	$n_{.1(j)}^*$	$n^{*..(j)}$

Assim, na Tabela 8,  $n_{a0(j)}^*(.)$  e  $n_{a1(j)}^*(.)$  denotam as freqüências totais de transições do estado  $a$  para os estados 0 e 1, respectivamente. O índice  $(j)$  é incluído para diferenciar as freqüências de transições em relação à história da cadeia. Usando (73) do teste de Mantel-Hanszel, tem-se assintoticamente (sob a hipótese nula) a seguinte estatística para teste:

$$Z_{MH} = \frac{\sum_{j=1}^2 n_{a0(j)}^*(P) - \sum_{j=1}^2 (E)_j}{\sqrt{\sum_{j=1}^2 (\text{Var})_j}} \sim N(0, 1), \quad (74)$$

em que  $(E)_j = \frac{n_{.0(j)}^* n_{a.(j)}^*(P)}{n^{*..(j)}}$  e  $(\text{Var})_j = \frac{n_{a.(j)}^*(P) n_{a.(j)}^*(A) n_{.0(j)}^* (n^{*..(j)} - n_{.0(j)}^*)}{(n^{*..(j)})^3}$ .

Assim para  $H_a : \psi > 1$ , o nível descritivo do teste é dado por:

$$p^* = P(Z > Z_{MH}),$$

constituindo-se num critério para discriminar entre os dois tratamentos. O teste se aplica naturalmente para cadeias de alcance maior do que 1. Nesses casos, as caselas da Tabela 8 denotam as freqüências de transições de  $h_{it}$  para os estados  $S = \{0, 1\}$ . Em caso de rejeição de  $H_0$  há evidências estatísticas de que os indivíduos do grupo ativo estão mais propensos à transição para o estado 1 do que os indivíduos do grupo placebo.

### 2.2.2.2 Teste da diferença das probabilidades de transição

Uma outra forma de comparar dois tratamentos é a partir das probabilidades de transição. Conforme descrito na seção anterior, há várias formulações de hipóteses equivalentes. Sem perda de generalidade, considere as hipóteses que avaliam as mudanças do estado 0 para o estado 1.

$$\text{Hipóteses}^{(1)} : \text{mudanças para o estado 1} \begin{cases} H_0 : \pi_{01}(A) - \pi_{01}(P) = 0; \\ H_a : \pi_{01}(A) - \pi_{01}(P) > 0. \end{cases}$$

Um teste pode ser especificado para avaliar essas hipóteses, num procedimento fundamentado na teoria assintótica com respeito à distribuição dos estimadores das probabilidades de transição, de modo análogo à idéia que se tem na comparação de duas proporções, a menos pelo fato de que, no contexto dos modelos de transição, as proporções representam probabilidades de mudanças ou transições no tempo. Como essas probabilidades satisfazem a equação de Chapman-Kolmogorov, as hipóteses<sup>(1)</sup> podem ser reescritas em termos de funções de transição em  $r = (s + v)$  passos. Essas funções parecem ser mais indicadas para avaliar o que ocorre com o indivíduo, considerando o início do estudo, as transições e, seu estado ao final do estudo. Ademais, para o caso em que as probabilidades de transição são estacionárias, a função de transição em  $r$  etapas, por Chapman-Kolmogorov é:

$$\pi_{01}^{(r)} = \sum_{k \in S} \pi_{0k}^{(s)} \pi_{k1}^{(v)}.$$

Logo, podem-se formular as hipóteses:

$$\text{Hipóteses}^{(1*)} : \text{mudanças para o estado 1 em r passos} \begin{cases} H_0 : \pi_{01}^{(r)}(A) - \pi_{01}^{(r)}(P) = 0; \\ H_a : \pi_{01}^{(r)}(A) - \pi_{01}^{(r)}(P) > 0. \end{cases}$$

Para estabelecer a estatística do teste, parte-se do resultado sobre a distribuição amostral do estimador de máxima verossimilhança das probabilidades de transição,  $\hat{\pi}_{ab}$ , descrito em Anderson e Goodman (1957) e apresentado na seção (2.1.2.4.2). Assim, quando  $n \rightarrow \infty$ , tem-se:

$$\sqrt{n\varsigma_a}(\hat{\pi}_{ab} - \pi_{ab}) \rightarrow N(0, \pi_{ab}(1 - \pi_{ab})),$$

em que  $\varsigma_a = \sum_{j=1}^k \sum_{t=1}^T \frac{n_j(0)}{n} \pi_{ja}^{[t-1]}$ . Assim, para os grupos de tratamentos, tem-se assintoticamente que:

$$\hat{\pi}_{01}^{(r)}(A) \sim N \left[ \pi_{01}^{(r)}(A), \frac{\pi_{01}^{(r)}(A)[1 - \pi_{01}^{(r)}(A)]}{n(A)\varsigma_a(A)} \right]$$

e

$$\hat{\pi}_{01}^{(r)}(P) \sim N \left[ \pi_{01}^{(r)}(P), \frac{\pi_{01}^{(r)}(P)[1 - \pi_{01}^{(r)}(P)]}{n(P)\varsigma_a(P)} \right].$$

Como há independência entre os tratamentos, a diferença entre as estatísticas  $\hat{\pi}_{01}^{(r)}(A) - \hat{\pi}_{01}^{(r)}(P)$ , quando  $n(A) \rightarrow \infty$  e  $n(P) \rightarrow \infty$ , também tem como limite uma distribuição normal:

$$\hat{\pi}_{01}^{(r)}(A) - \hat{\pi}_{01}^{(r)}(P) \sim N(\pi_D, \sigma_D^2), \quad (75)$$

em que  $\pi_D = \pi_{01}^{(r)}(A) - \pi_{01}^{(r)}(P)$  e  $\sigma_D^2 = \frac{\pi_{01}^{(r)}(A)[1 - \pi_{01}^{(r)}(A)]}{n(A)\varsigma_a(A)} + \frac{\pi_{01}^{(r)}(P)[1 - \pi_{01}^{(r)}(P)]}{n(P)\varsigma_a(P)}$ .

Desse modo, sob  $H_0$  e considerando teste unilateral à direita, o nível descritivo é dado por:

$$p^* = P \left( Z > \frac{\hat{\pi}_D}{\hat{\sigma}_D} \right). \quad (76)$$

### 2.2.2.3 Modelo de transição de Markov misto

Sob um modelo de transição de Markov misto, as variáveis aleatórias  $Y_i$  condicionadas à história do processo ( $\mathbf{h}_i$ ) e aos efeitos aleatórios ( $\mathbf{d}_i$ ) seguem um modelo linear generalizado. A parte sistemática do modelo inclui as respostas prévias como efeitos fixos, variáveis explicativas associadas exclusivamente aos efeitos aleatórios ou a efeitos fixos e

aleatórios. O intercepto do modelo também pode ser um parâmetro fixo ou aleatório, dependendo das pressuposições estabelecidas. Assim, de um modelo geral, para uma cadeia de ordem  $q$ , tem-se como preditor linear:

$$\eta_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \sum_{r=1}^q \alpha_r y_{i(t-r)} + \mathbf{z}'_{it}\mathbf{d}_i,$$

em que  $\boldsymbol{\beta}$ ,  $\boldsymbol{\alpha}$  são os parâmetros associados aos efeitos fixos e  $\mathbf{d}_i$  aos efeitos aleatórios.

Para o caso do estudo epidemiológico sobre doença respiratória, sob um modelo de Markov pode-se considerar que o efeito do tratamento tem o mesmo peso nas probabilidades de transição para o estado de doença de qualquer paciente, assim como o estado do indivíduo na ocasião  $(t - 1)$ , contribui com peso fixo para o estado do indivíduo na ocasião  $t$ . Entretanto, pode-se considerar que cada indivíduo tem uma propensão para a doença, refletindo suas predisposições genéticas e influências não mensuráveis de fatores ambientais. Assim, as probabilidades de transição são determinadas não somente por efeitos fixos, mas também por um componente aleatório (intercepto). Assumindo dependência de ordem 1, o modelo logístico de transição misto tem, então, a seguinte estrutura funcional:

$$\eta_{it} = (\beta_0 + d_{i0}) + \mathbf{x}'_{it}\boldsymbol{\beta} + \alpha y_{i(t-1)}, \quad (77)$$

em que  $\mathbf{x}'_{it}$  engloba a covariável referente ao efeito de tratamento e  $y_{i(t-1)}$  é a covariável da suposição de Markov. O termo  $d_{i0}$  representa a propensão individual para a probabilidade de doença respiratória. Da mesma forma sexo e idade também podem ser adicionados ao modelo (77) como efeitos fixos. Para simplificar a notação, considere  $\boldsymbol{\delta}$  o vetor de parâmetros de efeitos fixos, incluindo  $\boldsymbol{\beta}$ ,  $\boldsymbol{\alpha}$  e  $\mathbf{x}'_{it}$  a  $i$ -ésima linha da matriz de planejamento associada a esses efeitos:

$$\eta_{it} = \mathbf{x}'_{it}\boldsymbol{\delta} + \mathbf{z}'_{it}\mathbf{d}_i.$$

A função de verossimilhança para  $\boldsymbol{\delta}$  e  $\mathbf{G}$  no caso estacionário é proporcional a:

$$L(\boldsymbol{\delta}, \mathbf{G}; \mathbf{y}) \propto \prod_{i=1}^N \int \prod_{t=2}^{n_i} [\mu_{it}(\boldsymbol{\delta}, \mathbf{G})]^{y_{it}} [1 - \mu_{it}(\boldsymbol{\delta}, \mathbf{G})]^{1-y_{it}} f(\mathbf{d}_i; \mathbf{G}) d(\mathbf{d}_i), \quad (78)$$

em que  $\mu_{it}(\boldsymbol{\delta}, \mathbf{d}_i) = \text{E}(Y_{it} \mid \mathbf{d}_i; \boldsymbol{\delta})$ .

A forma da função de verossimilhança (78) é equivalente àquela definida para os MLGM, a menos pelo fato de que os produtórios não envolvem a primeira observação, devido



à suposição da cadeia de Markov. Focalizando a atenção sob função de ligação canônica e supondo distribuição normal para o efeito aleatório,  $\mathbf{d}_i$ , a expressão (78), reduz-se a:

$$\prod_{i=1}^N \int \exp \left[ \boldsymbol{\delta}' \sum_{t=2}^{n_i} \mathbf{x}_{it}^* y_{it} + \mathbf{d}_i' \sum_{t=2}^{n_i} \mathbf{z}_{it} y_{it} - \sum_{t=2}^{n_i} \ln \{1 + \exp(\mathbf{x}_{it}^* \boldsymbol{\delta} + \mathbf{z}_{it}' \mathbf{d}_i)\} \right] (2\pi)^{-1} |\mathbf{G}|^{-1/2} \exp \left( \frac{-\mathbf{d}_i' \mathbf{G}^{-1} \mathbf{d}_i}{2} \right) d(\mathbf{d}_i),$$

em que  $\mathbf{G}$  é matriz de covariâncias dos efeitos aleatórios,  $\mathbf{d}_i$ .

Essa conexão e analogia com os modelos lineares generalizados mistos permite que os programas desenvolvidos nos *softwares* estatísticos para o ajuste desses modelos, possam ser adaptados para o modelo de transição, a fim de se estimarem os parâmetros referentes aos efeitos fixos e aleatórios, bem como a variância desses efeitos. Como uma das opções computacionais disponíveis, para sua implementação no SAS, pode-se combinar a macro **DROPOUT** para o arranjo do conjunto de dados de acordo com a ordem de cadeia necessária e, a macro **GLIMMIX**, para o ajuste do modelo de transição misto. Nesse caso, a integral da função de verossimilhança é feita a partir da aproximação dos dados. Nos procedimentos NLMIXED e no glmML do R, a solução da integral se dá por quadratura gaussiana.

Uma vez ajustado o modelo de transição com intercepto aleatório, estimam-se as probabilidades de transição como nos modelos com efeitos fixos, ou seja, por meio de:

$$\pi_{ab}(t) = \frac{\exp(\eta_{it})}{1 + \exp(\eta_{it})},$$

Matricialmente, pode-se escrever:

$$\mathbf{P}(t) = \begin{pmatrix} \pi_{00}(t) = \frac{1}{1 + \exp(\beta_0 + \mathbf{x}'_{it} \boldsymbol{\beta})} & \pi_{01}(t) = \frac{\exp(\beta_0 + \mathbf{x}'_{it} \boldsymbol{\beta})}{1 + \exp(\beta_0 + \mathbf{x}'_{it} \boldsymbol{\beta})} \\ \pi_{10}(t) = \frac{1}{1 + \exp(\beta_0 + \mathbf{x}'_{it} \boldsymbol{\beta} + \alpha)} & \pi_{11}(t) = \frac{\exp(\beta_0 + \mathbf{x}'_{it} \boldsymbol{\beta} + \alpha)}{1 + \exp(\beta_0 + \mathbf{x}'_{it} \boldsymbol{\beta} + \alpha)} \end{pmatrix},$$

em que  $\mathbf{x}'_{it} \boldsymbol{\beta}$  representa o efeito das variáveis explicativas referentes aos efeitos fixos, exceto os termos intercepto (efeito aleatório) e covariável da suposição de Markov.

## 2.3 Resultados

### 2.3.1 Modelo de transição de primeira ordem

Para ilustrar as diferentes formas de se trabalhar com modelos de transição de Markov, a princípio, realiza-se uma análise ignorando a dependência em relação às variáveis explicativas. A Tabela 9 resume todas as frequências de transições de primeira ordem possíveis, ou seja, nesse estudo longitudinal com 555 observações (111 indivíduos e 5 ocasiões) há 444 transições. Essas transições podem ser observadas separadamente, considerando do início do estudo (*baseline*) para a visita 1 (primeira transição), da visita 1 para a visita 2 (segunda transição), da visita 2 para a 3 (terceira transição) e da visita 3 para a 4 (quarta transição), como mostra a Tabela 10. Note que dos 61 indivíduos inicialmente na condição ruim, há 53 nessa condição ao final do estudo, indicando que 8 mudam para o estado bom.

Tabela 9 - Distribuição do número de transições da condição respiratória, no exemplo do estudo epidemiológico. Probabilidades estimadas para o processo estacionário

$Y_{i(t-1)}$	$Y_{it}$		Total
	0	1	
0	149 (0,7304)	55 (0,2696)	204
1	47 (0,1958)	193 (0,8042)	240
			444

As frequências relativas da Tabela 9 constituem-se em estimativas para as probabilidades de transição,  $P(Y_{it} = y_{it} \mid Y_{i(t-1)} = y_{i(t-1)})$ . Assim, por exemplo, para um indivíduo cuja condição de respiração é ruim no tempo  $(t - 1)$ , a probabilidade de transição para o estado de saúde bom é de 0,2696, enquanto que essa probabilidade é 2,98 vezes maior se o estado prévio é bom. Essas estimativas pouco se diferem, quando observadas as 4 transições possíveis dadas pela Tabela 10. Quando as estimativas de probabilidades de transição são próximas, como nesse caso, isso pode ser um indicativo de que a cadeia de Markov é estacionária.

Tabela 10 - Número de transições e probabilidades estimadas do estudo sobre a condição respiratória, para as quatro transições

Primeira transição				Segunda transição		
$Y_{it}$				$Y_{it}$		
$Y_{i(t-1)}$	0	1	Total	0	1	Total
0	39 (0,6393)	22 (0,3607)	61	35 (0,7609)	11 (0,2391)	46
1	7 (0,1400)	43 (0,8600)	50	16 (0,2462)	49 (0,7538)	65
			111			111

Terceira transição				Quarta transição		
$Y_{it}$				$Y_{it}$		
$Y_{i(t-1)}$	0	1	Total	0	1	Total
0	37 (0,7255)	14 (0,2745)	51	38 (0,8261)	8 (0,1739)	46
1	9 (0,1500)	51 (0,8500)	60	15 (0,2308)	50 (0,7692)	65
			111			111

Para comprovar essa hipótese efetua-se o teste para estacionariedade de uma cadeia, através da estatística:

$$\xi = \sum_{t=1}^4 \sum_{a=0}^1 \sum_{b=0}^1 \frac{n_a(t-1)[\hat{\pi}_{ab}(t) - \hat{\pi}_{ab}]^2}{\hat{\pi}_{ab}},$$

e, usando as estimativas das Tabelas 9 e 10, tem-se

$$\xi = \frac{61[0,6393 - 0,7304]^2}{0,7304} + \dots + \frac{65[0,7692 - 0,8042]^2}{0,8042} = 8,27,$$

sendo o resultado não significativo (nível descritivo = 0,2185), ou seja, desconsiderando o efeito das variáveis explicativas o processo é estacionário. Nesse caso, a matriz de probabilidades de transição, resume-se a:

$$\hat{P} = \begin{pmatrix} 0,7304 & 0,2696 \\ 0,1958 & 0,8042 \end{pmatrix}.$$

No entanto, como se sabe, nesse estudo, há variáveis explicativas e uma das questões é verificar a eficiência da droga ativa e seu efeito na estimativa da probabilidade de transição. Assim, na Tabela 11 são apresentadas as freqüências de transição para os estados ruim (0) e bom (1), estratificada por tratamento. Nota-se que há indicativo de influência do tratamento sob a condição respiratória do indivíduo. A estimativa de transição para o estado bom é 1,53 vezes maior para pacientes do grupo ativo. Entretanto, há nitidamente uma forte correlação entre as observações nos tempos  $t$  e  $(t - 1)$ , conforme mostram as Tabelas 12 e 13. Analisando, separadamente, a influência do tratamento dadas as respostas prévias, percebe-se que apesar do indicativo do efeito da droga ativa, há uma maior probabilidade de transição para estado bom, se a condição prévia é boa.

Tabela 11 - Distribuição das transições de acordo com o tipo de tratamento no estudo epidemiológico sobre a condição respiratória

Tratamento	$Y_{it}$		Total
	0	1	
Placebo	127 (0,5570)	101 (0,4430)	228
Ativo	69 (0,3194)	147 (0,6806)	216
			444

Tabela 12 - Número de transições e estimativas de probabilidades, de acordo com o tipo de tratamento e condição respiratória na visita anterior

Primeira transição						
Tratamento	$Y_{it} = 0$	$Y_{it} = 1$	Total	$Y_{it} = 0$	$Y_{it} = 1$	Total
Placebo	22	9	31	7	19	26
	(0,7097)	(0,2903)		(0,2692)	(0,7308)	
Ativo	17	13	30	0	24	24
	(0,5667)	(0,4333)		(0,0000)	(1,0000)	
	$(Y_{i(t-1)} = 0)$		61	$(Y_{i(t-1)} = 1)$		50
Segunda transição						
Tratamento	$Y_{it} = 0$	$Y_{it} = 1$	Total	$Y_{it} = 0$	$Y_{it} = 1$	Total
Placebo	25	4	29	10	18	28
	(0,8621)	(0,1379)		(0,3571)	(0,6429)	
Ativo	10	7	17	6	31	37
	(0,5882)	(0,4118)		(0,1622)	(0,8378)	
	$(Y_{i(t-1)} = 0)$		46	$(Y_{i(t-1)} = 1)$		65
Terceira transição						
Tratamento	$Y_{it} = 0$	$Y_{it} = 1$	Total	$Y_{it} = 0$	$Y_{it} = 1$	Total
Placebo	26	9	35	5	17	22
	(0,7429)	(0,2571)		(0,2273)	(0,7727)	
Ativo	11	5	16	4	34	38
	(0,6875)	(0,3125)		(0,1053)	(0,8947)	
	$(Y_{i(t-1)} = 0)$		51	$(Y_{i(t-1)} = 1)$		60
Quarta transição						
Tratamento	$Y_{it} = 0$	$Y_{it} = 1$	Total	$Y_{it} = 0$	$Y_{it} = 1$	Total
Placebo	27	4	31	5	21	26
	(0,8710)	(0,1290)		(0,1923)	(0,8077)	
Ativo	11	4	15	10	29	39
	(0,7333)	(0,2667)		(0,2564)	(0,7436)	
	$(Y_{i(t-1)} = 0)$		46	$(Y_{i(t-1)} = 1)$		65

Tabela 13 - Total de transições e estimativas de probabilidades, de acordo com o tipo de tratamento e condição respiratória na visita anterior

Tratamento	$Y_{it}$		Total	$Y_{it}$		Total
	0	1		0	1	
Placebo	100 (0,7937)	26 (0,2063)	126	27 (0,2647)	75 (0,7353)	102
Ativo	49 (0,6282)	29 (0,3718)	78	20 (0,1449)	118 (0,8551)	138
	$(Y_{i(t-1)} = 0)$		204	$(Y_{i(t-1)} = 1)$		240

Sob a hipótese de estacionariedade, Tabela 13, essa probabilidade é cerca de 3,56 vezes maior para o grupo placebo e 2,29 vezes maior para o grupo ativo, quando comparadas à condição prévia ruim. Pela definição de modelos lineares generalizados de transição, os resultados apresentados nas Tabelas 12 e 13 são reproduzidos (estimados) a partir das estimativas dos parâmetros do modelo, especificado por:

$$\eta_{it} = \beta_0 + \beta_1 \text{tratamento} + \alpha y_{i(t-1)} + \gamma (y_{i(t-1)} * \text{tratamento}), \quad (79)$$

sendo, que quando ajustado marginalmente permite o estudo de todas transições no tempo.

Tabela 14 - Estimativas dos parâmetros para os modelos de transição de 1ª ordem, considerando efeitos de tratamento e interação tratamento e resposta prévia, no estudo sobre a condição respiratória

Parâmetros	Estimativas			
	1ª transição	2ª transição	3ª transição	4ª transição
intercepto	-0,8938	-1,8326	-1,0609	-1,9095
tratamento (ativo)	0,6256	1,4759	0,2724	0,8979
resposta prévia ( $Y_{i(t-1)}$ )	1,8923	2,4204	2,2846	3,3446
tratamento * $Y_{i(t-1)}$	24,7412	-0,4215	0,6439	-1,2683

A Tabela 14 descreve as estimativas dos parâmetros para os quatro modelos de transição de primeira ordem, considerando os efeitos de tratamento e interação com a resposta no tempo anterior. Ressalta-se a significância estatística pelo teste de Wald ( $p < 0,01$ ), para as respostas no tempo  $(t - 1)$ . Para estimar as probabilidades de transição, a partir desses coeficientes, usa-se a expressão (66). Por exemplo, na primeira transição, para o grupo com tratamento ativo, a probabilidade de transição para o estado bom, são dadas por  $\hat{\pi}_{01} = \frac{\exp(-0,8938+0,6256)}{1+\exp(-0,8938+0,6256)} = 0,4333$ , se o estado prévio do paciente é ruim e, por  $\hat{\pi}_{11} = \frac{\exp(-0,8938+0,6256+1,8923+24,7412)}{1+\exp(-0,8938+0,6256+1,8923+24,7412)} \simeq 1,0000$ , se o estado anterior é bom. Decorre, por diferença que  $\hat{\pi}_{00} = \frac{1}{1+\exp(-0,8938+0,6256)} = 0,5667$  e  $\hat{\pi}_{10} = \frac{1}{1+\exp(-0,8938+0,6256+1,8923+24,7412)} \simeq 0,0000$ . Para o grupo Placebo, essas probabilidades são dadas por  $\hat{\pi}_{01} = \frac{\exp(-0,8938)}{1+\exp(-0,8938)} = 0,2903$ , com probabilidade complementar  $\hat{\pi}_{00} = \frac{1}{1+\exp(-0,8938)} = 0,7097$ , e,  $\hat{\pi}_{11} = \frac{\exp(-0,8938+1,8923)}{1+\exp(-0,8938+1,8923)} = 0,7308$ , com probabilidade complementar  $\hat{\pi}_{10} = \frac{1}{1+\exp(-0,8938+1,8923)} = 0,2692$ . Do mesmo modo para as outras três transições. Simplificadamente, caracterizam-se os resultados pelas matrizes:

### Grupo tratado

$$\hat{P}_{(1)} = \begin{pmatrix} 0,5667 & 0,4333 \\ 0,0000 & 1,0000 \end{pmatrix}, \hat{P}_{(2)} = \begin{pmatrix} 0,5882 & 0,4118 \\ 0,1622 & 0,8378 \end{pmatrix}, \hat{P}_{(3)} = \begin{pmatrix} 0,6875 & 0,3125 \\ 0,1053 & 0,8947 \end{pmatrix}, \hat{P}_{(4)} = \begin{pmatrix} 0,7333 & 0,2667 \\ 0,2564 & 0,7436 \end{pmatrix}.$$

### Grupo placebo

$$\hat{P}_{(1)} = \begin{pmatrix} 0,7097 & 0,2903 \\ 0,2692 & 0,7308 \end{pmatrix}, \hat{P}_{(2)} = \begin{pmatrix} 0,8621 & 0,1379 \\ 0,3571 & 0,6429 \end{pmatrix}, \hat{P}_{(3)} = \begin{pmatrix} 0,7429 & 0,2571 \\ 0,2273 & 0,7727 \end{pmatrix}, \hat{P}_{(4)} = \begin{pmatrix} 0,8710 & 0,1290 \\ 0,1923 & 0,8077 \end{pmatrix}.$$

Essas probabilidades também podem ser estimadas usando outra função de ligação. Verifica-se, para esses dados, que o uso das funções de ligação proibito e complemento *log-log*, levam aos mesmos resultados (probabilidades estimadas). Analisando essas matrizes, percebe-se que as probabilidades de transições para o estado bom, do grupo com tratamento ativo são maiores do que as do grupo placebo, enquanto que as probabilidades de transição para o estado ruim são maiores para o grupo placebo (exceto na última transição). O fato de as probabilidades,  $\pi_{01}$  e  $\pi_{11}$  serem decrescentes no grupo ativo pode ser decorrência da

perda do efeito do tratamento. Os resultados apresentados na Tabela 13 são reproduzidos a partir do modelo com preditor linear (79), ajustado para a totalidade das transições.

Tabela 15 - Estimativas dos parâmetros para o modelo de transição estacionário de 1ª ordem, considerando efeitos de tratamento e interação tratamento e resposta prévia, no estudo sobre a condição respiratória

Parâmetros	Estimativas	Erro-padrão	nível descritivo
intercepto	-1,3471	0,2201	< 0,01
tratamento (ativo)	0,8225	0,3215	0,010
resposta prévia ( $Y_{i(t-1)}$ )	2,3687	0,3144	< 0,01
tratamento * $Y_{i(t-1)}$	-0,0692	0,4606	0,880

Para estimar as probabilidades usam-se os coeficientes da 2ª coluna da Tabela 15. Por exemplo, para os pacientes com tratamento ativo, a probabilidade de mudança do estado ruim para o estado bom é estimada por  $\hat{\pi}_{01} = \frac{\exp(-1,3471+0,8225)}{1+\exp(-1,3471+0,8225)} = 0,3718$ , enquanto que para o grupo placebo, essa probabilidade é dada por  $\hat{\pi}_{01} = \frac{\exp(-1,3471)}{1+\exp(-1,3471)} = 0,2063$ . Assim, é possível estimar todas as probabilidades de transição, que são consideradas independentes do tempo. Essa pressuposição, é ainda confirmada aplicando-se o teste da homogeneidade no tempo (ou estacionariedade) para as matrizes de probabilidades de transição em cada um dos grupos ativo e placebo, estimadas a partir dos modelos de transição marginais. As estatísticas pelo teste de qui-quadrado são  $\xi = 9,62$ , para o grupo ativo e  $\xi = 5,93$ , para o grupo placebo, sendo  $\chi_{6; 0,01}^2 = 16,81$ , os resultados são não significativos. Matricialmente, tem-se:

$$\hat{\mathbf{P}}_{(ativo)} = \begin{pmatrix} 0,6282 & 0,3718 \\ 0,1449 & 0,8551 \end{pmatrix} \quad \text{e} \quad \hat{\mathbf{P}}_{(placebo)} = \begin{pmatrix} 0,7937 & 0,2063 \\ 0,2647 & 0,7353 \end{pmatrix}.$$

A partir dessas matrizes, pode-se também observar que, para os pacientes com estado de saúde ruim na visita anterior (linhas superiores das matrizes  $\hat{\mathbf{P}}_{(ativo)}$  e  $\hat{\mathbf{P}}_{(placebo)}$ ), o logaritmo da razão de chances, isto é,  $\ln[(0,3718/0,6282)/(0,2063/0,7937)] = 0,8225$ , é a estimativa do coeficiente do efeito de tratamento. Para os pacientes cuja condição de saúde é boa em  $(t-1)$  (linhas inferiores das matrizes  $\hat{\mathbf{P}}_{(ativo)}$  e  $\hat{\mathbf{P}}_{(placebo)}$ ), tem-se que



$\ln[(0,8551/0,1449)/(0,7353/0,2647)] = 0,7533$  é a soma dos coeficientes referentes ao efeito de tratamento mais a interação de tratamento e resposta prévia.

Ainda, dos resultados apresentados na Tabela 13, têm-se as seguintes estimativas de razão de chances:

$$\psi_{Y_{(t-1)}=0} = 2,27 \quad \text{e} \quad \psi_{Y_{(t-1)}=1} = 2,12,$$

sendo consideradas homogêneas em relação à história do processo. Isso pode ser verificado pela hipótese,  $H_0 : \gamma = 0$ , da interação entre o tratamento e a resposta prévia, a qual conforme Tabela 15, não é rejeitada. Sendo assim, uma primeira idéia sobre a comparação dos grupos ativo e placebo, pode ser feita por meio do teste para razão de chances de Mantel-Hanszel, para avaliar a hipótese  $H_0 : \psi = 1$ . O valor esperado e a variância associados à  $Y_{(t-1)} = 0$ , são 92,08 e 9,48, e para  $Y_{(t-1)} = 1$ , são 19,97 e 9,23, respectivamente. A estatística do teste é dada por:

$$Z_{MH} = \frac{(100 + 27) - (92,02 + 19,97)}{\sqrt{9,48 + 9,23}} = 3,46,$$

a qual leva ao nível descritivo 0,0002, ou seja, há forte indicativo de que os indivíduos do grupo ativo estão mais propensos à transição para o estado bom.

Deve ser ressaltado que esses resultados são apenas descritivos e, na verdade, tais recursos constituem-se em técnicas exploratórias para modelos de transição. Nesse estudo, há, ainda, as covariáveis sexo, idade e centro clínico e, certamente, uma descrição sucessiva através de tabelas de contingência é impraticável. Nesses casos, é imprescindível a especificação das transições a partir do ajuste dos modelos. Na Tabela 16, são apresentadas as estimativas dos parâmetros para o modelo estacionário de primeira ordem, considerando as demais covariáveis do estudo.

O modelo 1 é ajustado considerando os efeitos do centro clínico, sexo, idade, tratamento e condição de saúde na ocasião  $(t - 1)$ . São significativos, pelo teste de Wald, os efeitos do centro clínico, tratamento, e resposta prévia. No modelo 2, adicionalmente, são considerados os efeitos das interações da resposta prévia com essas covariáveis, além do fator ocasião (visita). Nenhuma dessas interações são significativas. Quando não há interação, entre a história do processo e as covariáveis, há um indicativo de que as respostas prévias afetam a probabilidade de transição, mas os efeitos das covariáveis sob elas são os mesmos,

Tabela 16 - Estimativas dos parâmetros fixos,  $\hat{\beta}$ , e respectivos erros-padrão,  $s(\hat{\beta})$ , para três modelos de transição estacionários de primeira ordem, no estudo sobre a condição respiratória

Parâmetros	Modelo 1		Modelo 2		Modelo 3	
	$\hat{\beta}$	$s(\hat{\beta})$	$\hat{\beta}$	$s(\hat{\beta})$	$\hat{\beta}$	$s(\hat{\beta})$
intercepto	-0,3248	0,4104	1,1776	0,7930	0,0043	0,4870
centro 1	-0,7622	0,2477	-0,9852	0,3548	-0,8916	0,2952
sexo (feminino)	0,0506	0,3111	0,1891	0,4486	0,0190	0,3740
tratamento (ativo)	0,8382	0,2413	0,8040	0,3566	0,9337	0,2881
idade	-0,0177	0,0093	-0,0365	0,0158	-0,0193	0,0111
visita			-0,2345	0,1509		
$Y_{it-1}$	2,2277	0,2337	0,6575	1,0858	1,7614	0,2553
centro * $Y_{i(t-1)}$			0,4652	0,4988		
sexo * $Y_{i(t-1)}$			-0,2228	0,6350		
tratamento * $Y_{i(t-1)}$			-0,0182	0,4937		
idade * $Y_{i(t-1)}$			0,0320	0,0200		
visita * $Y_{i(t-1)}$			0,1116	0,2144		

independente da história. O modelo 3 é ajustado considerando os mesmos efeitos do modelo 1, porém supondo intercepto aleatório. Nota-se que não há grandes mudanças em relação às estimativas dos efeitos fixos, exceto no termo fixo do intercepto. A variância estimada para o efeito aleatório do intercepto é 0,5837, com erro padrão 0,4060. Um teste de razão de verossimilhanças é feito para avaliar a hipótese nula de que o modelo de transição tem intercepto aleatório contra a alternativa de que o modelo tem somente efeitos fixos. Para tanto usam-se os logaritmos das funções de verossimilhanças maximizadas sob os modelos 1 e 3. O resultado desse teste é não significativo (nível descritivo = 0,1604), orientando a escolha do modelo de efeitos fixos.

Desse modo, considerando os efeitos significativos do modelo 1, são estimadas

as probabilidades de transição, através da função:

$$\hat{\pi}_{a1} = \frac{\exp(-0,3248 - 0,7622(\text{centro}) + 0,8382(\text{tratamento}) + 2,2277Y_{i(t-1)})}{1 + \exp(-0,3248 - 0,7622(\text{centro}) + 0,8382(\text{tratamento}) + 2,2277Y_{i(t-1)})},$$

com  $\hat{\pi}_{a0} = 1 - \hat{\pi}_{a1}$ ,  $a \in S = \{0, 1\}$ . As matrizes de probabilidades de transição considerando esses efeitos, para os respectivos grupos ativo e placebo, são:

$$\hat{\mathbf{P}}_{(ativo)} = \begin{pmatrix} 0,5618 & 0,4382 \\ 0,1214 & 0,8786 \end{pmatrix} \quad \text{e} \quad \hat{\mathbf{P}}_{(placebo)} = \begin{pmatrix} 0,7478 & 0,2522 \\ 0,2422 & 0,7578 \end{pmatrix}.$$

Usando os resultados das probabilidades de transição dessas matrizes, pode-se efetuar o teste para a diferença das probabilidades considerado sob a hipótese de mudanças para o estado bom, em 4 passos. As probabilidades necessárias para aplicação do teste são obtidas pela equação de Chapman-Kolmogorov:

$$\hat{\mathbf{P}}_{(ativo)}^{(4)} = \begin{pmatrix} 0,2464 & 0,7536 \\ 0,2088 & 0,7911 \end{pmatrix} \quad \text{e} \quad \hat{\mathbf{P}}_{(placebo)}^{(4)} = \begin{pmatrix} 0,5232 & 0,4768 \\ 0,4578 & 0,5422 \end{pmatrix}.$$

A diferença observada entre as probabilidades é  $\pi_{01}^{(4)}(\text{A}) - \pi_{01}^{(4)}(\text{P}) = 0,2768$ , com erro-padrão estimado em 0,1527, levando à estatística  $Z = 1,81$ , (nível descritivo = 0,036). O resultado ratifica a evidência de diferença significativa a favor da grupo tratado.

Adicionalmente, como em toda análise estatística, após o ajuste de um modelo, faz-se necessário verificar as pressuposições estatísticas. Para os modelos de transição, técnicas usadas para diagnósticos nos MLGs podem ser adaptadas e, constituem-se em importantes recursos exploratórios. A Figura 2 mostra os gráficos semi-normal de probabilidades com envelopes simulados, ao nível de confiança de 95%, para os modelos estacionários 1 e 2, apresentados na Tabela 16. Percebe-se que os dois modelos (sem e com interação) não estão bem ajustados, uma vez que os pontos saem do envelope. Pode estar havendo algum tipo de falha sistemática (como inadequacidade da função de ligação), ou alguma discrepância isolada (pontos atípicos).

Focalizando a atenção sob Modelo 1, nota-se pelo gráfico dos resíduos, Figura 3, que não há discrepância isolada, uma vez que os resíduos do modelo se distribuem no intervalo  $(-2, +2)$ . Adicionalmente, efetua-se o teste para a adequacidade da função de ligação, através do teste da razão de verossimilhanças (diferença de desvios), considerando como variável

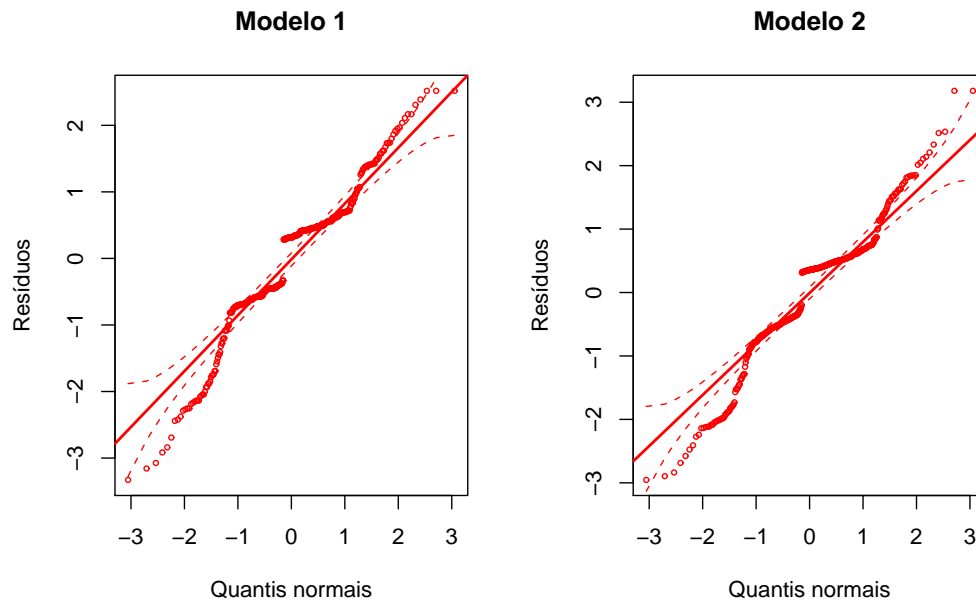


Figura 2 - Gráficos semi-normal de probabilidades com envelopes simulados, para os modelos de transição estacionários de 1<sup>a</sup> ordem, no estudo da condição respiratória

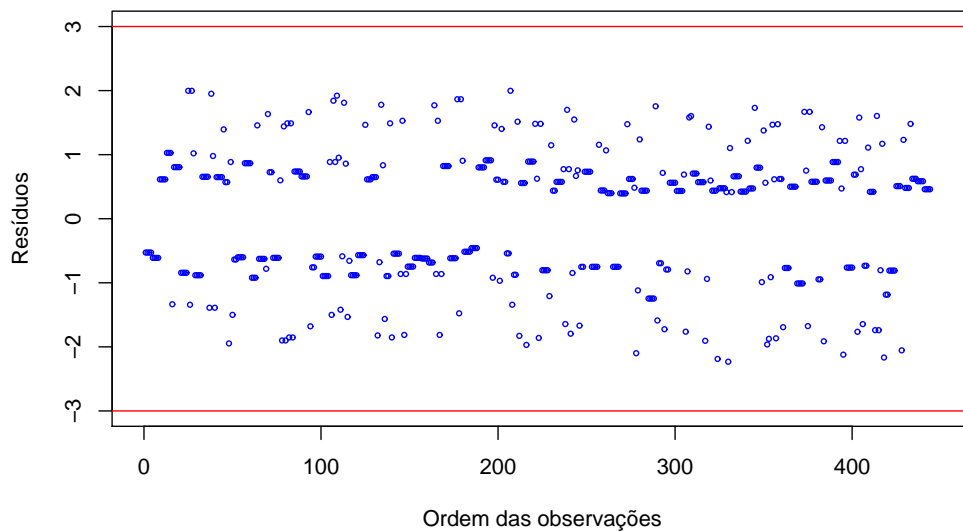


Figura 3 - Gráfico dos resíduos de Pearson versus a ordem das observações para o modelo (1) de transição estacionário de 1<sup>a</sup> ordem, no estudo da condição respiratória

adicionada  $u = \eta^2$ . O resultado é não significativo (nível descritivo = 0,4375), indicando adequidade da função logística. Assim, pode ainda existir algum outro tipo de falha. Em se tratando de modelos de transição é importante considerar também a possibilidade de alcance da cadeia maior do que 1.

### 2.3.2 Modelo de transição de segunda ordem

Um modelo de transição de 2ª ordem permite verificar se as inferências acerca dos parâmetros mudam com a dependência (ordem da cadeia). Assim, dando prosseguimento à análise, é explorada a possibilidade de se considerar um modelo de segunda ordem. Na Tabela 17 são apresentadas as freqüências de transições de segunda ordem, as quais permitem verificar a necessidade de se considerar uma cadeia de Markov de alcance 2, bem como sua estacionariedade.

Tabela 17 - Freqüências de transições de segunda ordem observadas no estudo sobre a condição respiratória

$h_t$	Primeira transição			Segunda transição			Terceira transição		
	$Y_{it} = 0$	$Y_{it} = 1$	Total	$Y_{it} = 0$	$Y_{it} = 1$	Total	$Y_{it} = 0$	$Y_{it} = 1$	Total
00	31 (0,7949)	8 (0,20510)	39	28 (0,8000)	7 (0,2000)	35	33 (0,8919)	4 (0,1081)	37
01	6 (0,2727)	16 (0,7273)	22	3 (0,2727)	8 (0,7273)	11	6 (0,4286)	8 (0,5714)	14
10	4 (0,5714)	3 (0,4286)	7	9 (0,5625)	7 (0,4375)	16	5 (0,5556)	4 (0,4444)	9
11	10 (0,2326)	33 (0,7674)	43	6 (0,1224)	43 (0,8776)	49	9 (0,1765)	42 (0,8235)	51

A partir da Tabela 17, especificam-se as freqüências sob a hipótese de estacionariedade, conforme mostra a Tabela 18. Aplicando o teste de qui-quadrado para estacionariedade:

$$\xi = \frac{39(0,7949 - 0,8288)^2}{0,8288} + \dots + \frac{51(0,1765 - 0,1748)^2}{0,1748} = 4,59,$$

Tabela 18 - Total de freqüências de transições de segunda ordem observadas no estudo sobre a condição respiratória

$h_t$	$Y_{it} = 0$	$Y_{it} = 1$	Total
00	92	19	111
	(0,8288)	(0,1712)	
01	15	32	47
	(0,3191)	(0,6809)	
10	18	14	32
	(0,5625)	(0,44375)	
11	25	118	143
	(0,1748)	(0,8252)	

sendo  $\chi_{4,(0,01)}^2 = 13,27$ , tem-se que mesmo para o processo de ordem 2, não se rejeitaria a hipótese de estacionariedade (nível descritivo = 0,3318). Pelo critério da razão de verossimilhanças, a estatística obtida é  $\lambda = 2,03$ .

Por outro lado, pode-se também aplicar os mesmos testes para se verificar a ordem da cadeia, ou seja, avaliar as hipóteses

$$H_0: \text{A cadeia é de ordem 1 } (\pi_{abc} = \pi_{bc});$$

$$H_a: \text{A cadeia é de ordem 2 } (\pi_{abc} \neq \pi_{bc}),$$

com  $a, b, c \in S = \{0, 1\}$ . Da Tabela 17, também podem ser estimadas as freqüências de primeira ordem, sob  $H_0$ , como mostra a Tabela 19.

Assim, as estimativas da Tabela 18 podem ser comparadas às freqüências da Tabela 19, fixando-se  $Y_{i(t-1)} = 0$  e, posteriormente  $Y_{i(t-1)} = 1$ . Pelo critério de qui-quadrado tem-se:

$$\xi = \frac{111(0,8288 - 0,7692)^2}{0,8288} + \dots + \frac{143(0,8252 - 0,7895)^2}{0,8252} = 14,35,$$

sendo significativo, uma vez que  $\chi_{2,(0,01)}^2 = 9,21$  (nível descritivo = 0,00076). O critério da razão de verossimilhanças leva à estatística  $\lambda = 13,17$  (nível descritivo = 0,0013).

Os resultados dessa análise exploratória preliminar indicam que é razoável considerar para esse caso um modelo estocástico estacionário de ordem 2. Assim, considerando

Tabela 19 - Total de freqüências de transições de primeira ordem observadas no estudo sobre a condição respiratória, considerando 333 observações

$h_t$	$Y_{it} = 0$	$Y_{it} = 1$	Total
$Y_{i(t-1)} = 0$	110 (0,7692)	33 (0,2308)	143
$Y_{i(t-1)} = 1$	40 (0,2105)	150 (0,7895)	190

as demais variáveis explicativas do estudo, são apresentadas na Tabela 20 as estimativas para os parâmetros do modelo de 2ª ordem.

Tabela 20 - Estimativas dos parâmetros e respectivos erros-padrão para o modelo de transição estacionário de segunda ordem, no estudo sobre a condição respiratória

Parâmetros	Estimativa	Erro-padrão
intercepto	-0,5616	0,4940
centro 1	-0,6540	0,2967
sexo (feminino)	0,1945	0,3709
tratamento (ativo)	0,7339	0,2889
idade	-0,0252	0,0111
$Y_{i(t-1)}$	1,9213	0,2979
$Y_{i(t-2)}$	0,9477	0,2974

Para o ajuste desse modelo, são consideradas apenas as observações referentes às visitas 3 a 5, totalizando 333 transições. Assim, comparando os modelos de primeira e segunda ordens, percebe-se que inclusão de  $Y_{i(t-2)}$  como covariável, acarreta numa sensível mudança nos coeficientes de regressão, sendo que  $Y_{i(t-1)}$  continua sendo a variável com maior peso para explicar as probabilidades de transição, seguida pela covariável  $Y_{i(t-2)}$ . As duas respostas prévias e o efeito de tratamento são significativos. Note que para a decisão da ordem da cadeia, incluindo as demais variáveis explicativas, pode-se usar o critério da razão de verossimilhanças, a partir dos logaritmos das funções de verossimilhanças para os modelos

de 1ª e 2ª ordens. Os valores obtidos são  $-166,79$  e  $-161,85$ , respectivamente. Logo, a estatística da razão de verossimilhanças é  $\lambda = 9,88$ , com nível descritivo igual a  $0,0016$ . Esse teste equivale à diferença de desvios para os modelos encaixados de primeira e segunda ordens, ambos ajustados com 333 observações. Logicamente, nesse estudo, é desnecessária uma extensão para ordens maiores e, os resultados, ratificam a hipótese de que um modelo baseado na cadeia de Markov de ordem 2 é satisfatório para explicar o comportamento funcional dos dados.

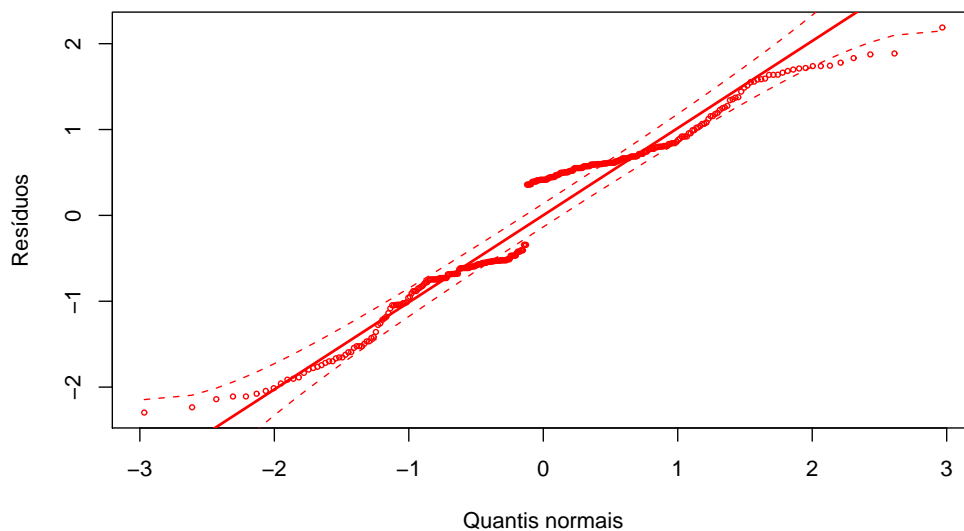


Figura 4 - Gráfico semi-normal de probabilidades com envelopes simulados para o modelo de transição estacionário de 2ª ordem, no estudo da condição respiratória

A Figura 4 apresenta o gráfico semi-normal de probabilidades com envelopes simulados, ao nível de confiança de 95%. É possível notar que há indicativo de que o modelo de segunda ordem esteja correto, embora ainda existam pontos fora do envelope. Porém, o ajuste obtido é melhor quando comparado ao modelo de primeira ordem. Considerando, então, os efeitos significativos desse modelo, as matrizes de probabilidades de transição, para



os respectivos grupos ativo e placebo, são:

$$\hat{\mathbf{P}}_{(ativo)} = \begin{pmatrix} 0,4570 & 0,5430 \\ 0,2460 & 0,7540 \\ 0,1097 & 0,8903 \\ 0,0455 & 0,9545 \end{pmatrix} \quad \text{e} \quad \hat{\mathbf{P}}_{(placebo)} = \begin{pmatrix} 0,6368 & 0,3632 \\ 0,4046 & 0,5954 \\ 0,2042 & 0,7958 \\ 0,0905 & 0,9095 \end{pmatrix}.$$

Analisando essas matrizes, nota-se que quando o histórico do paciente é ruim nas duas últimas ocasiões,  $\mathbf{h} = (0,0)$ , a probabilidade de transição (mudança) para o estado bom é estimada em 0,5430 se ele recebe a droga ativa, enquanto que para um paciente do grupo placebo essa probabilidade é 0,3632. Essas probabilidades são maiores se o indivíduo está bom em ao menos uma das ocasiões anteriores, sempre com diferença a favor dos pacientes do grupo ativo. Nota-se, também, que quando a história é  $\mathbf{h}_t = \{1,0\}$ , há probabilidades maiores de transição para o estado bom na ocasião  $t$ , em relação à  $\mathbf{h}_t = \{0,1\}$  e essa diferença, pode sinalizar que, embora, haja evidência da droga ativa e das respostas prévias sob o estado na ocasião  $t$ , essa transição para o estado bom, não ocorre de forma imediata, isto é, em um intervalo de tempo “pequeno”. Como não se tem acesso à base original de dados, com detalhes das ocasiões de observação dos dados, não se pode fazer uma análise mais detalhada com respeito ao tempo necessário para que essa transição ocorra. E, certamente, uma análise nessa linha necessitaria também de uma amostra maior de ocasiões, a fim de que seja possível aproximar o processo de tempo discreto do processo de tempo contínuo, possibilitando a estimação dessas taxas de transição.

As razões de chances estimadas para cada possibilidade de história da cadeia, mantendo-se o efeito da estratificação por tratamento são:

$$\psi_{\mathbf{h}_t=(0,0)} = 2,083; \quad \psi_{\mathbf{h}_t=(0,1)} = 2,082; \quad \psi_{\mathbf{h}_t=(1,0)} = 2,082; \quad \psi_{\mathbf{h}_t=(1,1)} = 2,087.$$

Esses valores evidenciam a homogeneidade da razão de chances em relação à história da cadeia e, conseqüentemente pode-se aplicar o teste de Mantel-Hanszel para comparação dos grupos ativo e placebo. A estatística do teste é  $Z_{MH} = 2,43$ , com nível descritivo igual a 0,0074, indicando forte evidência a favor do tratamento ativo. O aumento da ordem da cadeia, nesse caso, não muda as considerações com respeito à discriminação entre os tratamentos.

### 3 CONSIDERAÇÕES FINAIS

A análise de dados longitudinais com respostas binárias pode ser feita a partir da definição de vários modelos. A escolha do modelo, por sua vez, em geral, está relacionada aos objetivos do estudo. Os modelos de transição de Markov constituem-se em uma classe de modelos condicionais, uma vez que se tem interesse em modelar a distribuição condicional da resposta no tempo  $t$ , dada uma ou mais respostas prévias e um conjunto de covariáveis. Essa metodologia tem estreita relação com as áreas das ciências médica, veterinária e afins, uma vez que, é plausível admitir que o estado de um indivíduo numa dada ocasião possa ter influência sob seu estado num tempo futuro. Desse modo, a avaliação de drogas terapêuticas deve levar em conta a possibilidade dessa influência, que não é “mensurada”, ao se adotarem na estrutura modelos marginais. Assim, ao se empregar um modelo de transição, a hipótese nula inicial a ser explorada é a questão da dependência do tempo, bem como a ordem da cadeia. A estacionariedade do processo é sempre desejável, mas nem sempre ocorre. A análise dos resíduos também é importante para validar as inferências decorrentes de um modelo linear generalizado de transição. Quando há violação de um ou mais pressupostos, essa análise pode auxiliar na identificação de falhas ou, ainda, orientar a escolha do modelo para a dependência do tempo.

Adicionalmente, como contribuições para os pesquisadores e usuários potenciais, o presente trabalho destaca os testes para comparação de tratamentos bem como a inclusão de efeitos aleatórios nesses modelos. Assim, a suposição de que cada indivíduo tem uma propensão para o desenvolvimento ou, não, de uma determinada característica pode ser analisada, considerando na estrutura do modelo estocástico um intercepto aleatório. Os tratamentos envolvidos no estudo podem ser comparados a partir das probabilidades de transição ou por meio da razão de chances. O teste baseado nas probabilidades de transição são bem aplicados no contexto dos modelos estocásticos de primeira ordem, inclusive para processos não estacionários. Para cadeias de alcance maior do que 1, em virtude do aumento do número de parâmetros, o teste da razão de chances apresenta-se mais apropriado.

Apesar da relevância do assunto e de sua aplicabilidade, verifica-se que há, ainda, outras questões que merecem atenção e estudos nessa área. Primeiro, a metodologia apresentada neste trabalho pode ser estendida para mais de duas categorias de resposta,

usando o modelo de chances proporcionais. Segundo, os testes propostos são assintóticos e, portanto, podem não ser válidos quando o número de observações é pequeno. Fazem-se necessários estudos e simulações que possam avaliar a eficiência e o poder desses testes, para diferentes valores de  $n$ . Quanto ao teste para comparar tratamentos pelas probabilidades de transição, ele não é restrito ao caso de dois tratamentos. Para mais de dois tratamentos, podem-se efetuar comparações duas a duas, mas há necessidade de se corrigir o nível de significância. E isso também requer estudos adicionais.

Observações omissas também são comuns em estudos longitudinais. Para os modelos marginais, as equações de estimação garantem a consistência das estimativas dos parâmetros sob a condição de que essas perdas sejam completamente ao acaso, ou seja, se uma observação é perdida, ela não pode depender das respostas prévias. Assim, com modelos de transição, valores perdidos podem inviabilizar a análise, uma vez que, por definição, são modelos condicionais. Por isso, em geral, quando uma observação é perdida, o perfil individual de respostas é excluído da contribuição para a função de verossimilhança. Para o caso do modelo de Markov de primeira ordem, alguns estudos indicam como uma alternativa, tratar a observação após um valor omissivo como se fosse a primeira observação. Outra alternativa é aumentar a ordem da cadeia, que certamente traz penalizações pelo aumento no número de parâmetros. Pode-se, ainda, optar pela imputação de dados.

Nesse contexto, considera-se que, embora os objetivos propostos estejam satisfeitos, este trabalho é uma abordagem inicial para modelos de transição, que são fontes de perspectiva para pesquisas futuras.

## REFERÊNCIAS

- ANDERSON, D.A.; AITKIN, M. Variance component models with binary response: interviewer variability. **Journal of the Royal Statistical Society**, Series B, London, v. 47, p. 203 – 210, 1985.
- ANDERSON, T.W.; GOODMAN, L.A. Statistical Inference about Markov Chains. **Annals of Mathematical Statistics**. Ann Arbor, v. 28, p. 89 – 110, 1957.
- ANDRADE, D.F.; SINGER, J.M. Análise de dados longitudinais. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA (SINAPE), 7., 1986, Campinas. **Anais do SINAPE**, Campinas: Associação Brasileira de Estatística, 1986, 106 p.
- AGRESTI, A. **Categorical data analysis**. New York: John Wiley, 1990, 558 p.
- ARANDA-ORDAZ, F.J. On two families of transformations to additivity for binary response data. **Biometrika**, Cambridge, v. 68, p. 357-363, 1981.
- ATKINSON, A.C. **Plots, Transformations and Regression**. New York: Oxford University Press, 1985, 282 p.
- AZZALINI, A. Maximum likelihood estimation of order  $n$  for stationary stochastic processes. **Biometrika**, Cambridge, v. 70, p. 381 – 388, 1983.
- AZZALINI, A. Estimation and hypothesis testing for collections of autoregressive time series. **Biometrika**, Cambridge, v. 71, p. 85 – 90, 1984.
- BARTLETT, M.S. The frequency goodness of fit test for probability chains. **Proceedings of the Cambridge Philosophical Society**, Cambridge, v. 47, p. 86 – 95, 1951.
- BASU, A.K. **Introduction to Stochastic Processes**. Kolkata: Alpha Science International, 2003, 419 p.
- BHAT, U.N; MILLER, G.K. **Elements of Applied Stochastic Processes**. 3rd ed. New York: John Wiley, 2002, 488 p.
- BISHOP, Y.M.M.; FIENBERG, S.E.; HOLLAND, P.W. **Discrete multivariate analysis: theory and practice**. Massachusetts: The MIT Press, 1975, 557 p.
- BOX, G.E.P. Problems in the analysis of growth and wear curves. **Biometrics**, Washington, v. 6, p. 362 – 389, 1950.
- BRESLOW, N.E.; CLAYTON, D.G. Approximate inference in generalized linear mixed models. **Journal of the American Statistical Association**, Boston, v. 88, n. 421, p. 9 – 25, 1993.
- COLLETT, D. **Modelling Binary Data**. London: Chapman & Hall, 1991, 369 p.
- COOK, R.D.; WEISBERG, S. **Residuals and influence in regression**. London: Chapman & Hall, 1982, 230 p.

CORDEIRO, G.M. Modelos Lineares Generalizados. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA (SINAPE), 7., 1986, Campinas. **Anais do SINAPE**, Campinas: Associação Brasileira de Estatística, 1986, 286 p.

COX, D.R.; SNELL, E.J. A general definition of residuals (with discussion). **Journal of The Royal Statistical Society**, London, v. B30, p. 248 – 275, 1968.

COX, D.R. **The analysis of binary data**. London: Methuen, 1970, 142 p.

CROWDER, M.J.; HAND, D.J. **Analysis of repeated measures**. London: Chapman & Hall, 1990, 257 p.

DEMÉTRIO, C.G.B. **Modelos Lineares Generalizados em Experimentação Agrônômica**. Piracicaba: USP/ESALQ, 2002, 113 p.

DIGGLE, P.J.; HEAGERTY, P.J.; LIANG, K.Y.; ZEGER, S.L. **Analysis of longitudinal data**. New York: Oxford University Press, 2002, 379 p.

DOBSON, A. J. **An Introduction to Statistical Modelling**. 2 ed. London: Chapman & Hall, 2001, 174 p.

FEARN, T. A bayesian approach to the growth curves. **Biometrika**, Cambridge, v. 62, p. 89 – 100, 1975.

FITZMAURICE, G.M., LAIRD, N.M. A likelihood-based method for analysing longitudinal binary responses. **Biometrika**, Cambridge, v. 80, p. 141 – 151, 1993.

GEISSER, S.; GREENHOUSE, S.W. An extension of Box's results on the use of F-distribution in multivariate analysis. **Annals of Mathematical Statistics**, Ann Arbor, v. 29, p. 885 – 891, 1958.

GOOD, I.J. The likelihood ratio test for Markov chains. **Biometrika**, Cambridge, v. 42, p. 531 – 533, 1955.

GOODMAN, L.A. Statistical Methods for analysing processes of change. **American Journal of Sociology**, Chicago, v. 68, p. 57 – 78, 1962.

GRIZZLE, J.E.; ALLEN, D.M. Analysis of dose and response curves. **Biometrics**, Washington, v. 25, p. 357 – 382, 1969.

HARVILLE, T.A. Maximum likelihood approaches to variance component estimation and to related problems. **Journal of the American Statistical Association**, Boston, v. 72, p. 320 – 338, 1977.

HUYNH, H.; FELDT, L.S. Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. **Journal of the American Statistical Association**, Boston, v.65, n.332, p. 1582 – 1589, 1970.

JORGENSEN, B. Exponential dispersion models (with discussion). **Journal of the Royal Statistical Society**, London, v B.49, p. 127 – 162, 1987.

- KALBFLEISCH, J.D.; LAWLESS, J.F. The analysis of panel data under a Markov assumption. **Journal of American Statistical Association**, Boston, v. 80, p. 863 – 871, 1985.
- KARLIN, S.; TAYLOR, H.M. **A first course in Stochastic Processes**, 2nd ed. New York: Academic Press, 1975. 557 p.
- KOCH, G.C.; CARR, G.J.; AMARA, I.A.; STOKES, M.E.; URYNIAK, T.J. Categorical Data Analysis. In: BERRY, D.A. **Statistical Methodology in the Pharmaceutical Sciences**. New York: Marcel Dekker, 1990, chap. 13, p. 389 – 473.
- KOCH, G.C.; LANDIS, J.R.; FREEMAN, J.L.; FREEMAN, D.H.; LEHMAN, R.G. A general methodology for the analysis for repeated measurements of categorical data. **Biometrics**, Washington, v. 38, p. 133 – 158, 1977.
- KORN, E.L.; WHITTEMORE, A.S. Methods for analysing panel studies of acute health effects of air pollution. **Biometrics**, Washington, v. 35, p. 715 – 802, 1979.
- KOVÁCS, Z. L. **Teoria da Probabilidade e Processos Estocásticos**. São Paulo: Editora Acadêmica de São Paulo, 1996, 120 p.
- LAIRD, N.M.; WARE, J.H. Random-effects models for longitudinal data. **Biometrics**, Washington, v. 38, p. 963 – 974, 1982.
- LEE, Y.; NELDER, J.A. Hierarchical generalized linear model (with discussion). **Journal of the Royal Statistical Society**, Series B, London, v. 58, p. 619 – 678, 1996.
- LIANG, K.Y.; ZEGER, S.L. Longitudinal data analysis using generalized linear models. **Biometrika**, Cambridge, v. 73, n. 1, p. 13 – 22, 1986.
- LINDSEY, J.K. **Modelling frequency and count data**. Oxford: Oxford University Press, 1995, 300 p.
- LINDSEY, J.K. **Statistical analysis of stochastic processes in time**. New York: Cambridge University Press, 2004, 338 p.
- MANTEL, N. HANSZEL, B.F. Statistical aspects of the analysis of the data from retrospective studies of disease. **Journal of the American Statistical Association**, Boston, v. 58, p. 690 – 700, 1959.
- McCULLAGH, P. Regression Methods for Ordinal Data. **Journal of The Royal Statistical Society**, Series B, London, v. 42, p. 109 – 142, 1980.
- McCULLAGH, P. **Tensor Methods in Statistics**. London: Chapman dan Hall, 1987, 285 p.
- McCULLAGH, P.; NELDER, J.A. Quasi-likelihood functions. **Annals of Statistics**, Hayward, v. 11, p. 59 – 67, 1983.
- McCULLAGH, P.; NELDER, J.A. **Generalized linear models**. 2nd ed. London: Chapman and Hall, 1989. 511 p.

McGILCHRIST, C.A. Estimation in generalized mixed models. **Journal of the Royal Statistical Society**. Series B, London, v. 56, p. 61 – 69, 1994.

MOLENBERGHS, G. VERBEKE, G. **Models for discrete longitudinal data**. New York: Springer-Verlag, 2005, 683 p.

NELDER, J.A.; WEDDERBURN, R.W.M. Generalized linear models. **Journal of the Royal Statistical Society**, Series A, London, v. 135, p. 370 – 384, 1972.

NOLETO, A.Q. **Análise de Dados Longitudinais com Respostas Categorizadas através de Processos de Markov**. 1991, 93 p. Dissertação (Mestrado em Estatística)-Universidade Estadual de Campinas, Campinas, 1986.

OCHI, Y.; PRENTICE, R.L. Likelihood inference incorrelated probit regression. **Biometrika**, Cambridge, v. 71, p. 531 – 543, 1984.

PAULA, G.A. **Modelos de Regressão com Apoio Computacional**. São Paulo: IME, 2004, 245 p.

PAIK, M.C. Repeated Measurement Analysis for Nonnormal Data in Small Samples. **Communications in Statistics - Simulation and Computation**, Monticello, v. 17, p. 1155 – 1171, 1988.

PINHEIRO, J.C.; BATES, D.M. **Mixed-effects in S and S-PLUS**, New York: Springer-Verlag, 2000, 552 p.

POTHOFF, R.F.; ROY S.N. A generalized multivariate analysis of variance model useful especially for growth curve problems. **Biometrika**, Cambridge, v. 51, p. 313 – 326, 1964.

PREGIBON, D. Logistic regression diagnostics. **Annals of Statistics**, Hayward, v. 9, p. 705 – 724, 1981.

RAO, C.R. Some problems involving linear hypotheses in multivariate analysis. **Biometrika**, Cambridge, v. 46, p. 49 – 58, 1959.

RAO, C.R. The theory of least squares when the parameters are stochastic and its applications to the analysis of growth curves. **Biometrika**, Cambridge, v. 52, p. 447 – 458, 1965.

R Development Core Team. **R: A language and environment for statistical computing 2.5.1**. Vienna, Austria, 2007. Disponível em: <<http://www.R-project.org>>. Acesso em: 25 agos. 2007.

ROSS, S.M. **Stochastic Processes**. New York: John Wiley Sons, 1996, 510 p.

SAS Institute Inc. **SAS/STAT User's Guide, Version 8**. Cary NC, 1999.

SAS Institute Inc. **The GLIMMIX Procedure (Experimental)**. Cary NC, 2004.

SINGER, J.M.; ANDRADE, D.F. Analysis of longitudinal data. In: SEN, P.K.; RAO, C.R.(Ed.). **Handbook of statistics**. Amsterdam: Elsevier, 2000. v. 18, p. 115 – 160.



- SINGER, B.; SPILERMAN, S. The representation of social processes by Markov models. **American Journal of Sociology**, Chicago, v. 82, p. 1 – 54, 1976.
- SMITH, D.M. **Oswald oriented software the analysis of longitudinal data in S-PLUS**. S-PLUS package version 3.4, 1998. Disponível em: <<http://www.maths.lancs.ac.uk/Software/Oswald>>. Acesso em: 20 jun. 2006.
- STATACorp LP. **Stata 9 Longitudinal-Panel data reference manual**. Collage Station, Texas, 2005
- STIRATELLI, R.; LAIRD, N.; WARE, J.H. Random effects-models for serial observations with binary response. **Biometrics**, Washington, v. 40, p. 961 – 971, 1984.
- STRAM, D.O.; WEI, L.J.; WARE, J.H. Analysis of repeated ordered categorical outcomes with possibly missing observations and time-dependent covariates. **Journal of American Statistical Association**, Boston, v. 83, p. 631 – 637, 1988.
- TUMA, N.B.; HANNAN, M.T.; GROENEVELD, L.P. Dynamic analysis of event histories. **American Journal of Sociology**, Chicago, v. 84, p. 820 – 854, 1979.
- WANG, P.C. Adding a variable in generalized linear models. **Technometrics**, Baltimore, v. 27, p. 273 – 276, 1985.
- WARE, J.H. Linear models for the analysis of several measurements in longitudinal studies. **American Statistician**, Washington, v. 39, p. 95 – 101, 1985.
- WARE, J. H.; LIPSITZ, S.; SPEISER, F. E. Issues in the Analysis of Repeated Categorical Outcomes. **Statistics in Medicine**, Chichester, v. 7, p. 95 – 107, 1988.
- WEDDERBURN, R.W.M. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. **Biometrika**, Cambridge, v. 61, p. 439 – 447, 1974.
- WILLIAMS, D.A. Residuals in generalized linear models. In: PROCEEDINGS OF THE INTERNATIONAL BIOMETRICS CONFERENCE, 12., 1984, Tokyo. **Annals of the International Biometrics Conference**, Tokyo, 1984, p. 59 – 68.
- WILLIAMS, D.A. Generalized linear model diagnostic using the deviance and single case deletion. **Applied Statistics**, London, v. 36, p. 181 – 191, 1987.
- ZEGER, S.L.; LIANG, K.Y. Longitudinal data analysis for discrete and continuous outcomes. **Biometrics**, Washington, v. 42, p. 121 – 130, 1986.
- ZEGER, S.L.; LIANG, K.Y. An overview of methods for the analysis of longitudinal data. **Statistics in Medicine**, Chichester, v. 11, p. 1825 – 1839, 1992.
- ZEGER, S.L.; LIANG, K.Y.; ALBERT, P. Models for longitudinal data: a generalized estimating equation approach. **Biometrics**, Washington, v. 44, p. 1049 – 1060, 1988.
- ZEGER, S.L.; LIANG, K.Y.; SELF, S.G. The analysis of binary longitudinal data with time-independent covariates. **Biometrika**, Cambridge, v. 72, p. 31 – 38, 1985.



ZEGER, S.L.; QAQISH, B. Markov regression models for time series: a quase-likelihood approach. **Biometrics**, Washington, v. 44, p. 1019 – 1031, 1988.

## APÊNDICES

## APÊNDICE A - Conjunto de dados do experimento

centro	id	trat	sexo	idade	base	resp1	resp2	resp3	resp4
1	1	P	M	46	0	0	0	0	0
1	2	P	M	28	0	0	0	0	0
1	3	A	M	23	1	1	1	1	1
1	4	P	M	44	1	1	1	1	0
1	5	P	F	13	1	1	1	1	1
1	6	A	M	34	0	0	0	0	0
1	7	P	M	43	0	1	0	1	1
1	8	A	M	28	0	0	0	0	0
1	9	A	M	31	1	1	1	1	1
1	10	P	M	37	1	0	1	1	0
1	11	A	M	30	1	1	1	1	1
1	12	A	M	14	0	1	1	1	0
1	13	P	M	23	1	1	0	0	0
1	14	P	M	30	0	0	0	0	0
1	15	P	M	20	1	1	1	1	1
1	16	A	M	22	0	0	0	0	1
1	17	P	M	25	0	0	0	0	0
1	18	A	F	47	0	0	1	1	1
1	19	P	F	31	0	0	0	0	0
1	20	A	M	20	1	1	0	1	0
1	21	A	M	26	0	1	0	1	0
1	22	A	M	46	1	1	1	1	1
1	23	A	M	32	1	1	1	1	1
1	24	A	M	48	0	1	0	0	0
1	25	P	F	35	0	0	0	0	0
1	26	A	M	26	0	0	0	0	0
1	27	P	M	23	1	1	0	1	1
1	28	P	F	36	0	1	1	0	0

1	29	P	M	19	0	1	1	0	0
1	30	A	M	28	0	0	0	0	0
1	31	P	M	37	0	0	0	0	0
1	32	A	M	23	0	1	1	1	1
1	33	A	M	30	1	1	1	1	0
1	34	P	M	15	0	0	1	1	0
1	35	A	M	26	0	0	0	1	0
1	36	P	F	45	0	0	0	0	0
1	37	A	M	31	0	0	1	0	0
1	38	A	M	50	0	0	0	0	0
1	39	P	M	28	0	0	0	0	0
1	40	P	M	26	0	0	0	0	0
1	41	P	M	14	0	0	0	0	1
1	42	A	M	31	0	0	1	0	0
1	43	P	M	13	1	1	1	1	1
1	44	P	M	27	0	0	0	0	0
1	45	P	M	26	0	1	0	1	1
1	46	P	M	49	0	0	0	0	0
1	47	P	M	63	0	0	0	0	0
1	48	A	M	57	1	1	1	1	1
1	49	P	M	27	1	1	1	1	1
1	50	A	M	22	0	0	1	1	1
1	51	A	M	15	0	0	1	1	1
1	52	P	M	43	0	0	0	1	0
1	53	A	F	32	0	0	0	1	0
1	54	A	M	11	1	1	1	1	0
1	55	P	M	24	1	1	1	1	1
1	56	A	M	25	0	1	1	0	1
2	57	P	F	39	0	0	0	0	0
2	58	A	M	25	0	0	1	1	1

2	59	A	M	58	1	1	1	1	1
2	60	P	F	51	1	1	0	1	1
2	61	P	F	32	1	0	0	1	1
2	62	P	M	45	1	1	0	0	0
2	63	P	F	44	1	1	1	1	1
2	64	P	F	48	0	0	0	0	0
2	65	A	M	26	0	1	1	1	1
2	66	A	M	14	0	1	1	1	1
2	67	P	F	48	0	0	0	0	0
2	68	A	M	13	1	1	1	1	1
2	69	P	M	20	0	1	1	1	1
2	70	A	M	37	1	1	0	0	1
2	71	A	M	25	1	1	1	1	1
2	72	A	M	20	0	0	0	0	0
2	73	P	F	58	0	1	0	0	0
2	74	P	M	38	1	1	0	0	0
2	75	A	M	55	1	1	1	1	1
2	76	A	M	24	1	1	1	1	1
2	77	P	F	36	1	1	0	0	1
2	78	P	M	36	0	1	1	1	1
2	79	A	F	60	1	1	1	1	1
2	80	P	M	15	1	0	0	1	1
2	81	A	M	25	1	1	1	1	0
2	82	A	M	35	1	1	1	1	1
2	83	A	M	19	1	1	0	1	1
2	84	P	F	31	1	1	1	1	1
2	85	A	M	21	1	1	1	1	1
2	86	A	F	37	0	1	1	1	1
2	87	P	M	52	0	1	1	1	1
2	88	A	M	55	0	0	1	1	0

2	89	P	M	19	1	0	0	1	1
2	90	P	M	20	1	0	1	1	1
2	91	P	M	42	1	0	0	0	0
2	92	A	M	41	1	1	1	1	1
2	93	A	M	52	0	0	0	0	0
2	94	P	F	47	0	1	1	0	1
2	95	P	M	11	1	1	1	1	1
2	96	P	M	14	0	0	0	1	0
2	97	P	M	15	1	1	1	1	1
2	98	P	M	66	1	1	1	1	1
2	99	A	M	34	0	1	1	0	1
2	100	P	M	43	0	0	0	0	0
2	101	P	M	33	1	1	1	0	1
2	102	P	M	48	1	1	0	0	0
2	103	A	M	20	0	1	1	1	1
2	104	P	F	39	1	0	1	0	0
2	105	A	M	28	0	1	0	0	0
2	106	P	F	38	0	0	0	0	0
2	107	A	M	43	1	1	1	1	0
2	108	A	F	39	0	1	1	1	1
2	109	A	M	68	0	1	1	1	1
2	110	A	F	63	1	1	1	1	1
2	111	A	M	31	1	1	1	1	1

## APÊNDICE B - Programas Computacionais utilizados

### B.1 - Processo não estacionário

```

/*PROGRAMA SAS PARA AJUSTE DE MODELOS EXPLORATÓRIOS DE TRANSIÇÃO*/
/*ADPATAÇÃO DA PROPOSTA DE STRAM, WEI E WARE (1988)-MOD.MARGINAIIS */
options nodate nonumber ps=1000;
data respira;
input centro$ id$ trat$ sexo$ idade base resp1 resp2 resp3 resp4;
datalines;
1 1 P M 46 0 0 0 0 0
1 2 P M 28 0 0 0 0 0
.....
2 110 A F 63 1 1 1 1 1
2 111 A M 31 1 1 1 1 1
;
proc print;
run;

/*Análise Exploratória: frequências de transições*/
proc freq order=data;
table base*resp1; /*Primeira transição*/
table resp1*resp2; /*Segunda transição*/
table resp2*resp3; /*Terceira transição*/
table resp3*resp4; /*Quarta transição*/
/*Análise Exploratória: Estratificação por tratamento*/
table trat*base*resp1; /*Primeira transição*/
table trat*resp1*resp2; /*Segunda transição*/
table trat*resp2*resp3; /*Terceira transição*/
table trat*resp3*resp4; /*Quarta transição*/
run;

/*MODELO LINEAR GENERALIZADO PRIMEIRA TRANSIÇÃO*/
proc genmod data=respira;
class centro id sexo trat;
model resp1 = trat base base*trat
            /dist=binomial;
run;

/*MODELO LINEAR GENERALIZADO SEGUNDA TRANSIÇÃO */
proc genmod data=respira;
class centro id sexo trat;
model resp2 = trat resp1 resp1*trat
            /dist=binomial;
run;

```

```

/*MODELO LINEAR GENERALIZADO TERCEIRA TRANSIÇÃO*/
proc genmod data=respira;
class centro id sexo trat;
model resp3 = trat resp2 resp2*trat
            /dist=binomial;
run;

/*MODELO LINEAR GENERALIZADO QUARTA TRANSIÇÃO */
proc genmod data=respira;
class centro id sexo trat;
model resp4 = trat resp3 resp3*trat
            /dist=binomial;
run;

```

## B.2 - Processo estacionário

```

/*MODELOS DE TRANSIÇÃO DE MARKOV: PROCESSOS ESTACIONÁRIOS*/
options nodate nonumber ps=1000;
data respira;
input centro$ id$ trat$ sexo$ idade visit resp;
timeclass=visit;
datalines;
1 1 P M 46 1 0
1 1 P M 46 2 0
1 1 P M 46 3 0
1 1 P M 46 4 0
1 1 P M 46 5 0
.....
2 111 A M 31 1 1
2 111 A M 31 2 1
2 111 A M 31 3 1
2 111 A M 31 4 1
2 111 A M 31 5 1
;
/*Carregando a Macro DROPOUT dos Geerts*/
%macro dropout(data=,id=,time=,response=,out=);
%if %bquote(&data)= %then %let data=&syslast;
proc freq data=&data noprint;
tables &id /out=freqid;
tables &time / out=freqtime;
run;
proc iml;
reset noprint;
use freqid;
read all var {&id};

```



```

nsub = nrow(&id);
use freqtime;
read all var {&time};
ntime = nrow(&time);
time = &time;
use &data;
read all var {&id &time &response};
n = nrow(&response);
dropout = j(n,1,0);
ind = 1;
do while (ind <= nsub);
    j=1;
    if (&response[(ind-1)*ntime+j]=.)
    then print "First Measurement is Missing";
    if (&response[(ind-1)*ntime+j]^=.) then
    do;
        j = ntime;
        do until (j=2);
            if (&response[(ind-1)*ntime+j]=.) then
            do;
                dropout[(ind-1)*ntime+j]=1;
                j = j-1;
            end;
            else j = 2;
            end;
        end;
        ind = ind+1;
    end;
prev = j(n,1,1);
prev[2:n] = &response[1:n-1];
i=1;
do while (i<=n);
    if &time[i]=time[1] then prev[i]=.;
    i = i+1;
end;
create help var {&id &time &response dropout prev};
append;
quit;
data &out;
merge &data help;
run;
%mend;

```

```

/*Rotina para gerar vetor resposta prévia de ordem 1*/
%dropout(data=respira,id=id,time=visit,response=resp,out=conj1);
data conj1a;
set conj1;
prev1=prev;
drop prev;
run;
/*Rotina para gerar vetor resposta prévia de ordem 2*/
%dropout(data=conj1a,id=id,time=visit,response=prev1,out=conj2);
data conj2a;
set conj2;
prev2=prev;
drop prev;
run;
proc print;
run;
/*MODELO LINEAR GENERALIZADO CADEIA DE MARKOV DE ORDEM 1*/
proc genmod data=conj2a;
class centro id trat sexo timeclass;
model resp = trat prev1 prev1*trat / dist=binomial;
run;
/*MLG CADEIA DE MARKOV ORDEM1: COM DEMAIS EFEITOS*/
proc genmod data=conj2a;
class centro id trat sexo timeclass;
model resp = centro sexo trat idade prev1 / dist=binomial;
run;
/*MLG CADEIA DE MARKOV ORDEM1: DEMAIS EFEITOS E INTERAÇÕES*/
proc genmod data=conj2a;
class centro id trat sexo timeclass;
model resp = centro sexo trat idade visit prev1
              prev1*centro prev1*sexo prev1*trat prev1*idade prev1*visit
              / dist=binomial;
run; /*MLG CADEIA DE MARKOV ORDEM2: COM DEMAIS EFEITOS*/
proc genmod data=conj2a;
class centro id trat sexo timeclass;
model resp = centro sexo trat idade prev1 prev2 / dist=binomial;
run;

```

B.3 - Rotinas computacionais pelo *software* R: Testes de hipóteses

```
#####PROGRAMA EM R PARA MODELOS DE TRANSIÇÃO#####
##AJUSTES DE MODELOS DE TRANSIÇÃO "MARGINAIS": Processos não estacionários##
dados<-read.table("respirasep.txt", header=T)
names(dados)
attach(dados)
center<-as.factor(center)

#MODELOS EXPLORATÓRIOS 1 = APENAS COM O EFEITO DE TRATAMENTO E INTERAÇÃO#
mt1<-glm(resp1~treat+base +base*treat,family=binomial)
mt2<-glm(resp2~treat+resp1+resp1*treat, family=binomial)
mt3<-glm(resp3~treat+resp2+resp2*treat, family=binomial)
mt4<-glm(resp4~treat+resp3+resp3*treat, family=binomial)
summary(mt1)
summary(mt2)
summary(mt3)
summary(mt4)

##ESTIMATIVAS DAS PROBABILIDADES DE TRANSIÇÃO##
      #Primeira transição#
#Grupo Ativo#
p01A1<-exp(-0.8938+0.6256)/(1+exp(-0.8938+0.6256))
p00A1<-1-p01A1
p11A1<-exp(-0.8938+0.6256+1.8923+24.7412)/(1+exp(-0.8938+0.6256+1.8923+24.7412))
p10A1<-1-p11A1
a01<-matrix(rbind(p00A1,p01A1),1)
a11<-matrix(rbind(p10A1,p11A1),1)
PA1<-matrix(rbind(a01,a11),2)

#Grupo Placebo#
p01P1<-exp(-0.8938)/(1+exp(-0.8938))
p00P1<-1-p01P1
p11P1<-exp(-0.8938+1.8923)/(1+exp(-0.8938+1.8923))
p10P1<-1-p11P1
p01<-matrix(rbind(p00P1,p01P1),1)
p11<-matrix(rbind(p10P1,p11P1),1)
PP1<-matrix(rbind(p01,p11),2)

      #Segunda transição#
#Grupo Ativo#
p01A2<-exp(-1.8326+1.4759)/(1+exp(-1.8326+1.4759))
p00A2<-1-p01A2
p11A2<-exp(-1.8326+1.4759+2.4204-0.4215)/(1+exp(-1.8326+1.4759+2.4204-0.4215))
p10A2<-1-p11A2
a02<-matrix(rbind(p00A2,p01A2),1)
```

```

a12<-matrix(rbind(p10A2,p11A2),1)
PA2<-matrix(rbind(a02,a12),2)

#Grupo Placebo#
p01P2<-exp(-1.8326)/(1+exp(-1.8326))
p00P2<-1-p01P2
p11P2<-exp(-1.8326+2.4204)/(1+exp(-1.8326+2.4204))
p10P2<-1-p11P2
p02<-matrix(rbind(p00P2,p01P2),1)
p12<-matrix(rbind(p10P2,p11P2),1)
PP2<-matrix(rbind(p02,p12),2)

      #Terceira transição#
#Grupo Ativo#
p01A3<-exp(-1.0609+0.2724)/(1+exp(-1.0609+0.2724))
p00A3<-1-p01A3
p11A3<-exp(-1.0609+0.2724+2.2846+0.6439)/(1+exp(-1.0609+0.2724+2.2846+0.6439))
p10A3<-1-p11A3
a03<-matrix(rbind(p00A3,p01A3),1)
a13<-matrix(rbind(p10A3,p11A3),1)
PA3<-matrix(rbind(a03,a13),2)

#Grupo Placebo#
p01P3<-exp(-1.0609)/(1+exp(-1.0609))
p00P3<-1-p01P3 p11P3<-exp(-1.0609+2.2846)/(1+exp(-1.0609+2.2846))
p10P3<-1-p11P3
p03<-matrix(rbind(p00P3,p01P3),1)
p13<-matrix(rbind(p10P3,p11P3),1)
PP3<-matrix(rbind(p03,p13),2)

      #Quarta transição#
#Grupo Ativo#
p01A4<-exp(-1.9095+0.8979)/(1+exp(-1.9095+0.8979))
p00A4<-1-p01A4
p11A4<-exp(-1.9095+0.8979+3.3446-1.2683)/(1+exp(-1.9095+0.8979+3.3446-1.2683))
p10A4<-1-p11A4
a04<-matrix(rbind(p00A4,p01A4),1)
a14<-matrix(rbind(p10A4,p11A4),1)
PA4<-matrix(rbind(a04,a14),2)

#Grupo Placebo#
p01P4<-exp(-1.9095)/(1+exp(-1.9095))
p00P4<-1-p01P4 p11P4<-exp(-1.9095+3.3446)/(1+exp(-1.9095+3.3446))
p10P4<-1-p11P4
p04<-matrix(rbind(p00P4,p01P4),1)

```

```

p14<-matrix(rbind(p10P4,p11P4),1)
PP4<-matrix(rbind(p04,p14),2)

##MODELOS EXPLORATÓRIOS 2 = CONSIDERANDO DEMAIS EFEITOS##
mt11<-glm(resp1~center+sex+treat+idade+base, family=binomial)
mt21<-glm(resp2~center+sex+treat+idade+resp1, family=binomial)
mt31<-glm(resp3~center+sex+treat+idade+resp2, family=binomial)
mt41<-glm(resp4~center+sex+treat+idade+resp3, family=binomial)
summary(mt11)
summary(mt21)
summary(mt31)
summary(mt41)
detach(dados)
rm(list=ls(all=TRUE))

##MODELOS ESTACIONÁRIOS##
dados1<-read.table("dadosdroup1.txt",header=T)
names (dados1)
attach(dados1)
centro<-as.factor(centro)
library(geepack)

#MODELO AUTO REGRESSIVO ESTACIONÁRIO COM FATOR TRAT E INTERAÇÃO#
mod0<-geeglm(resp~trat+prev1+prev1*trat,id=id, family=binomial)
summary(mod0)

#ESTIMATIVAS DAS PROBABILIDADES DE TRANSIÇÃO#
#Grupo Ativo#
p01A<-exp(-1.3471+0.8225)/(1+exp(-1.3471+0.8225))
p00A<-1-p01A
p11A<-exp(-1.3471+0.8225+2.3687-0.0692)/(1+exp(-1.3471+0.8225+2.3687-0.0692))
p10A<-1-p11A
a0<-matrix(rbind(p00A,p01A),1)
a1<-matrix(rbind(p10A,p11A),1)
PA<-matrix(rbind(a0,a1),2)

#Grupo Placebo#
p01P<-exp(-1.3471)/(1+exp(-1.3471))
p00P<-1-p01P
p11P<-exp(-1.3471+2.3687)/(1+exp(-1.3471+2.3687))
p10P<-1-p11P
p0<-matrix(rbind(p00P,p01P),1)
p1<-matrix(rbind(p10P,p11P),1)
PP<-matrix(rbind(p0,p1),2)

```

```

##TESTE PARA ESTACIONARIEDADE 1-ANÁLISE EXPLORATÓRIA##
nab<-c(39,22,7,43,35,11,16,49,37,14,9,51,38,8,15,50)
na.<-c(61,61,50,50,46,46,65,65,51,51,60,60,46,46,65,65)
pabt<-c(0.6393,0.3607,0.1400,0.8600,0.7609,0.2391,0.2462,0.7538,
        0.7255,0.2745,0.1500,0.8500,0.8261,0.1739,0.2308,0.7692)
pab<-rep(c(0.7304,0.2696,0.1958,0.8042),4)
Q2<-(na.*(pabt-pab)^2)/(pab)
X2<-sum(Q2)
1-pchisq(X2,6) #NÍVEL DESCRITIVO

##TESTE DE HOMOGENEIDADE OU ESTACIONARIEDADE POR TRAT##
#Grupo ativo#
nabA<-c(17,13,0,19,10,7,6,31,11,5,4,34,11,4,10,29)
na.A<-c(30,30,24,24,17,17,37,37,16,16,38,38,15,15,39,39)
pA<-rep(c(p00A,p01A,p10A,p11A),4)
Q2A<-((nabA-na.A*pA)^2)/(na.A*pA)
X2A<-sum(Q2A)
1-pchisq(X2A,6) #NÍVEL DESCRITIVO

#Grupo placebo#
nabP<-c(22,9,7,19,25,4,10,18,26,9,5,17,27,4,5,21)
na.P<-c(31,31,26,26,29,29,28,28,35,35,22,22,31,31,26,26)
pP<-rep(c(p00P,p01P,p10P,p11P),4)
Q2P<-((nabP-na.P*pP)^2)/(na.P*pP)
X2P<-sum(Q2P)
1-pchisq(X2P,6) #NÍVEL DESCRITIVO

##AJUSTE COM OS DEMAIS EFEITOS##
mod1<-geeglm(resp~centro+sexo+trat+idade+prev1,id=id,family=binomial)
summary(mod1)

##COM INTERAÇÕES##
mod2<-geeglm(resp~centro+sexo+trat+idade+visit+prev1+prev1*centro+prev1*sexo
             +prev1*trat+prev1*idade+prev1*visit,id=id, family=binomial)
summary(mod2)

##MODELO COM INTERCEPTO ALEATÓRIO##
library(glmmML)
#via Laplace
efal1<-glmmML(resp~centro+sexo+trat+idade+prev1,cluster=id,family=binomial)
#via quadratura gaussiana
efal2<-glmmML(resp~centro+sexo+trat+idade+prev1,cluster=id,family=binomial,
              method="ghq",n.points=2)

efal1
efal2

```

```
#Teste da razão de verossimilhanças: Ho: Modelo é de efeitos fixos
x<-451.97-450
1-pchisq(x,1)
```

```
##COMPARAÇÃO DOS TRATAMENTOS (ESTACIONÁRIOS):TRAT E INTERAÇÃO##
```

```
#Função de transição em 4 passos: Chapman-Kolmogorov#
```

```
P4A<-PA%*%PA%*%PA%*%PA
```

```
P4P<-PP%*%PP%*%PP%*%PP
```

```
#Estimativa do erro-padrão#
```

```
kA<-0.5556*P4A[1,1]+0.4444*P4A[2,1]
```

```
nA<-54
```

```
kP<-0.5439*P4P[1,1]+0.4561*P4P[1,1]
```

```
nP<-57
```

```
nAc<-nA*kA
```

```
nPc<-nP*kP
```

```
pid<-P4A[1,2]-P4P[1,2]
```

```
s2<-(P4A[1,2]*(1-P4A[1,2])/nAc)+(P4P[1,2]*(1-P4P[1,2])/nPc)
```

```
s<-sqrt(s2)
```

```
##nível descritivo##
```

```
zo<-pid/s
```

```
1-pnorm(zo)
```

```
## COMPARAÇÃO DE TRATAMENTOS: COM OS DEMAIS EFEITOS##
```

```
#Grupo Ativo#
```

```
p01A<-exp(-0.3248-0.7622+0.8382)/(1+exp(-0.3248-0.7622+0.8382))
```

```
p00A<-1-p01A
```

```
p11A<-exp(-0.3248-0.7622+0.8382+2.2277)/(1+exp(-0.3248-0.7622+0.8382+2.2277))
```

```
p10A<-1-p11A
```

```
a0<-matrix(rbind(p00A,p01A),1)
```

```
a1<-matrix(rbind(p10A,p11A),1)
```

```
PA<-matrix(rbind(a0,a1),2)
```

```
#Transição em 4 passos: Equação de Chapman-Kolmogorov#
```

```
P4A<-PA%*%PA%*%PA%*%PA
```

```
#Grupo Placebo#
```

```
p01P<-exp(-0.3248-0.7622)/(1+exp(-0.3248-0.7622))
```

```
p00P<-1-p01P
```

```
p11P<-exp(-0.3248-0.7622+2.2277)/(1+exp(-0.3248-0.7622+2.2277))
```

```
p10P<-1-p11P
```

```
p0<-matrix(rbind(p00P,p01P),1)
```

```
p1<-matrix(rbind(p10P,p11P),1)
```

```
PP<-matrix(rbind(p0,p1),2)
```

```
#Transição em 4 passos: Equação de Chapman-Kolmogorov#
```

```
P4P<-PP%*%PP%*%PP%*%PP
```

```

#DIFERENÇA DE PROBABILIDADES:MUDANÇAS PARA O ESTADO BOM #
pid<-P4A[1,2]-P4P[1,2]
#Estimativa do erro padrão#
kA<-0.5556*P4A[1,1]+0.4444*P4A[2,1]
nA<-54
kP<-0.5439*P4P[1,1]+0.4561*P4P[1,1]
nP<-57
nAc<-nA*kA nPc<-nP*kP
s2<-(P4A[1,2]*(1-P4A[1,2])/nAc)+(P4P[1,2]*(1-P4P[1,2])/nPc)
s<-sqrt(s2)
##nível descritivo##
zo<-pid/s
1-pnorm(zo)

##ENVELOPES SIMULADOS
library(car)
res1<-residuals(mod1,type="pearson")
res2<-residuals(mod2,type="pearson")
par(mfrow=c(1,2))
qq.plot(res1,xlab="quantis normais",ylab="resíduos", main="Modelo 1")
qq.plot(res2,xlab="quantis normais",ylab="resíduos", main="Modelo 2")

#TESTE DA FUNÇÃO DE LIGAÇÃO - VARIÁVEL ADICIONADA
eta1<-predict(mod1) u<-(eta1)^2
mod4<-geeglm(resp~centro+sexo+trat+idade+prev1+u,id=id,
             family=binomial)
anova(mod1,mod4, test="Chisq") #nível descritivo para Ho:gama=0(coef. de u)

##GRÁFICO DOS RESÍDUOS VERSUS ORDEM DAS OBSERVAÇÕES
ord<-1:444
plot(ord,res1,xlab="ordem das observações",ylab="resíduos", col="blue")
abline(h=-3,col="red")
abline(h=3,col="red")
detach(dados1)
rm(list=ls(all=TRUE))

##TESTE PARA ORDEM DE CADEIA:RAZÃO DE VEROSSIMILHANÇAS##
#a menos dos efeitos das covariáveis
nabc<-c(92,19,18,14,15,32,25,118)
pabc<-c(0.8288,0.1712,0.5625,0.4375,0.3191,0.6809,0.1748,0.8252)
pbc<-c(0.7692,0.2308,0.7692,0.2308,0.2105,0.7895,0.2105,0.7895)
logp1<-log(pabc)
logp2<-log(pbc)
RV<-2*sum(nabc*(logp1-logp2))
1-pchisq(RV,2) ##nível descritivo do teste

```



```

#considerando os efeitos das covariáveis:via MLG
dados2<-read.table("dadosdroup2.txt",header=T)
names(dados2)
attach(dados2)
centro<-as.factor(centro)
cad1<-glm(resp~centro+sexo+trat+idade+prev1,family=binomial)
cad2<-glm(resp~centro+sexo+trat+idade+prev1+prev2,family=binomial)
summary(cad1)
summary(cad2)
anova(cad1,cad2,test="Chisq")

##ENVELOPE SIMULADO
res3<-residuals(cad2, type="pearson")
qq.plot(res3,xlab="quantis normais",ylab="resíduos")

##CADEIA DE ORDEM 2: Estimativas de probabilidades##
#Grupo ativo
p001A<-exp(-0.5616+0.7339)/(1+exp(-0.5616+0.7339))
p000A<-1-p001A
p011A<-exp(-0.5616+0.7339+0.9477)/(1+exp(-0.5616+0.7339+0.9477))
p010A<-1-p011A
p101A<-exp(-0.5616+0.7339+1.9213)/(1+exp(-0.5616+0.7339+1.9213))
p100A<-1-p101A
p111A<-exp(-0.5616+0.7339+1.9213+0.9477)/(1+exp(-0.5616+0.7339+1.9213+0.9477))
p110A<-1-p111A
c1a<-c(p000A,p010A,p100A,p110A)
c2a<-c(p001A,p011A,p101A,p111A)
PA<-matrix(cbind(c1a,c2a),4)
#Grupo placebo
p001P<-exp(-0.5616)/(1+exp(-0.5616))
p000P<-1-p001P
p011P<-exp(-0.5616+0.9477)/(1+exp(-0.5616+0.9477))
p010P<-1-p011P
p101P<-exp(-0.5616+1.9213)/(1+exp(-0.5616+1.9213))
p100P<-1-p101P
p111P<-exp(-0.5616+1.9213+0.9477)/(1+exp(-0.5616+1.9213+0.9477))
p110P<-1-p111P c1p<-c(p000P,p010P,p100P,p110P)
c2p<-c(p001P,p011P,p101P,p111P)
PP<-matrix(cbind(c1p,c2p),4)

```