

COMPARAÇÃO DE DUAS ABORDAGENS DOS MODELOS DE TRANSIÇÃO DE MARKOV EM EXPERIMENTOS PLANEJADOS COM DADOS BINÁRIOS CORRELACIONADOS

Maurício Santana LORDÊLO¹
Sônia Maria De Stefano PIEDADE²
Gilênio Borges FERNANDES³
Rosemeire Levolgildo FIACCONE³

- RESUMO: Os modelos de transição de Markov constituem uma ferramenta de grande importância para diversas áreas do conhecimento quando são desenvolvidos estudos com medidas repetidas. Eles caracterizam-se por modelar a variável resposta ao longo do tempo condicionada a uma ou mais respostas anteriores, conhecidas como a história do processo. Além disso, é possível a inclusão de outras covariáveis. No caso das respostas binárias, pode-se construir uma matriz com as probabilidades de transição de um estado para outro. Neste trabalho, duas abordagens diferentes de modelos de transição foram comparadas para avaliar qual estima melhor o efeito causal de tratamentos em um estudo experimental cuja variável resposta é binária e medida ao longo do tempo. Estudos de simulação foram realizados levando em consideração experimentos balanceados com três tratamentos de natureza categórica. Para avaliar as estimativas foram utilizados o erro padrão, viés e percentual de cobertura dos intervalos de confiança. Os resultados mostraram que os modelos de transição marginalizados podem ser mais indicados na situação em que um experimento é desenvolvido com um reduzido número de medidas longitudinais e de repetições por tratamentos.
- PALAVRAS-CHAVE: Dados binários; medidas repetidas; efeito causal de tratamento

¹Universidade Estadual de Feira de Santana, Área de Estatística, Departamento de Ciências Exatas, Avenida Transnordestina S/N, CEP: 44036-900, Feira de Santana, BA, Brasil. E-mail: mslordelo@uefs.br

²Universidade de São Paulo, Escola Superior de Agricultura “Luiz de Queiroz” – ESALQ, Departamento de Ciências Exatas, Programa de Pós-Graduação em Estatística e Experimentação Agronômica, CEP: 13418-900, Piracicaba, SP, Brasil. E-mail: soniamsp@usp.br

³Universidade Federal da Bahia, Instituto de Matemática, Departamento de Estatística, Av. Adhemar de Barros, CEP: 40170-110, Salvador, BA, Brasil. Email: gilenio@ufba.br; rose.fiaccone@gmail.com

1 Introdução

Nos estudos longitudinais, grande parte dos esforços para análise dos dados está relacionada com a modelagem da resposta média levando em consideração a estrutura de correlação intra-indivíduos decorrente da medição da mesma variável no mesmo indivíduo em tempos diferentes. Quando os estudos são desenvolvidos considerando que a variável resposta não tem distribuição normal, geralmente a metodologia adequada para análise é fundamentada na extensão dos modelos lineares generalizados (Nelder e Wedderburn, 1972).

Diggle et al. (2002) apresentam três extensões dos modelos lineares generalizados (MLGs) para dados longitudinais, incorporando a dependência entre as observações ao longo do tempo: os modelos marginais, os modelos de transição e os modelos de efeitos aleatórios.

Nos modelos marginais, modela-se separadamente o efeito das variáveis explicativas no valor esperado da resposta, permitindo a incorporação de uma estrutura de correlação para a dependência das observações repetidas de cada indivíduo. O ajuste é feito por meio das equações de estimação generalizadas (Liang e Zeger, 1986).

Os modelos de efeitos aleatórios ou mistos são aqueles que incorporam efeitos fixos e aleatórios. Para dados longitudinais, este tipo de modelagem permite que os coeficientes da regressão variem entre os indivíduos. Esses modelos têm dois componentes: um intra-indivíduos (descrito pelo modelo de regressão com um intercepto e inclinação populacional) e outro entre indivíduos (variação no intercepto e inclinação individual). A estimação dos parâmetros pode ser realizada usando métodos de máxima verossimilhança. Para dados não normais, como o caso dos binários, técnicas de integração numérica são necessárias. Lara (2007) apresenta uma descrição dos métodos de aproximações numéricas que em geral podem ser utilizados.

O uso dos modelos de transição ou modelos de Markov permite verificar mudanças individuais no tempo e como essas mudanças são influenciadas pelas variáveis explicativas consideradas (Lara, 2007). Além disso, quando a resposta é binária, a dinâmica do processo a ser modelado pode ser analisada estimando a matriz de transição que traduz as mudanças de estados em termos das probabilidades condicionais.

As diferentes abordagens de modelos de transição de Markov para dados binários longitudinais são: modelo de transição usual (Diggle et al., 2002), modelo de transição usual com efeito aleatório (Lara, 2007), modelo de transição marginalizado (Heagerty, 2002) e modelo de transição marginalizado com parâmetro de dependência a razão de chances, no qual também permite a incorporação de um efeito aleatório (Gonçalves e Azzalini, 2008).

Para as estratégias de modelagem de dados binários longitudinais abordadas neste trabalho, considera-se que Y_{it} representa uma variável resposta e \mathbf{x}_{it} um vetor de variáveis explanatórias com comprimento p observado no tempo t .

Assim, y_{it} representa uma observação de Y_{it} correspondente a i -ésima unidade de análise no tempo t , com $i = 1, 2, \dots, m$ e $t = 1, 2, \dots, n_i$. Além disso, $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{in_i})$ representa a matriz $n_i \times p$ de covariáveis cujas linhas são os vetores $\mathbf{x}_{it} = (x_{it1}, x_{it2}, \dots, x_{itp})'$ de dimensões $p \times 1$.

De maneira equivalente, pode-se definir para cada unidade de análise um vetor $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{in_i})$ com média $E(\mathbf{Y}_i) = \boldsymbol{\mu}_i$ e matriz de covariância $\text{Var}(\mathbf{Y}_i) = \mathbf{V}_i$. As respostas para todas as unidades de análise são denotadas por

$\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m)$ que possui dimensão N , sendo que $N = \sum_{i=1}^m n_i$.

Neste trabalho, as abordagens dos modelos de transição usual e marginalizado com parâmetro de dependência a razão de chances são apresentadas para análise de dados binários correlacionados e contextualizadas na avaliação do efeito causal de tratamentos de natureza categórica em experimentos planejados e balanceados via processo de simulação.

2 Métodos

2.1 Modelo de transição usual

2.1.1 Definição geral

Os modelos de transição são baseados na teoria dos processos estocásticos sendo que, dentre os vários tipos de processos estocásticos que podem ser considerados, a cadeia de Markov é o mais frequentemente adotado. Segundo Diggle et al. (2002), a função de densidade de probabilidade da variável resposta Y_{it} pode ser escrita na forma da família exponencial

$$f(y_{it}|\mathbf{h}_{it}) = \exp \{ [y_{it}\theta_{it} - b(\theta_{it})] / \phi + c(y_{it}, \phi) \}, \quad (1)$$

sendo que \mathbf{h}_{it} é o vetor de dimensão q das respostas prévias, ϕ um parâmetro de dispersão supostamente conhecido, θ_{it} o parâmetro canônico, $b(\theta_{it})$ e $c(y_{it}, \phi)$ funções conhecidas. A ordem de um modelo de transição pode ser definida como o número de medições anteriores que influenciam a medição atual. Considerando que a distribuição condicional de Y_{it} dado \mathbf{h}_{it} depende somente das q respostas anteriores, então q representa a ordem do modelo.

Pode ser demonstrado que a média e variância condicionais ao vetor das respostas prévias são dadas respectivamente por

$$\mu_{it}^C = E(Y_{it}|\mathbf{h}_{it}) = b'(\theta_{it}) \quad (2)$$

e

$$v_{it}^C = \text{Var}(Y_{it}|\mathbf{h}_{it}) = b''(\theta_{it})\phi. \quad (3)$$

Desta forma, o modelo linear generalizado de transição pode ser escrito com

$$g(\mu_{it}^C) = \eta_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \sum_{r=1}^s f_r(\mathbf{h}_{it}; \boldsymbol{\alpha}), \quad (4)$$

sendo que $f_r(\cdot)$ são funções apropriadas, $v_{it}^C = v(\mu_{it}^C)\phi$, $g(\cdot)$ é uma função de ligação conhecida, $g(\mu_{it}^C)$ é o preditor linear e $v(\cdot)$ a função de variância. Assume-se que o passado afeta o presente por meio da soma de s termos, cada um deles podendo depender dos q valores prévios. Além disso, os vetores de parâmetros β e α estão associados às covariáveis e as respostas prévias, respectivamente. Consequentemente, um modelo de transição expressa a média condicional como uma função das covariáveis \mathbf{x}_{it} e das respostas anteriores $Y_{i(t-1)}, Y_{i(t-2)}, \dots, Y_{i(t-q)}$.

Um modelo de transição é denominado de estacionário se a forma funcional da dependência não tem relação com o tempo real em que ocorre.

2.1.2 Estimação

Os parâmetros para este tipo de abordagem podem ser estimados usando métodos baseados em máxima verossimilhança. De acordo com Diggle et al. (2002), dada as respostas Y_{i1}, \dots, Y_{in_i} , a distribuição conjunta destas variáveis ($f(y_{i1}, \dots, y_{in_i})$) pode ser escrita na forma

$$f(y_{in_i}|y_{i(n_i-1)}, \dots, y_{i1})f(y_{i(n_i-1)}|y_{i(n_i-2)}, \dots, y_{i1}) \dots f(y_{i2}|y_{i1})f(y_{i1}) \quad (5)$$

A distribuição condicional de Y_{it} em um modelo de Markov de ordem q é escrita por

$$f(y_{it}|\mathbf{h}_{it}) = f(y_{it}|y_{i(t-1)}, y_{i(t-2)}, \dots, y_{i(t-q)}). \quad (6)$$

Dessa forma, a contribuição do i -ésimo indivíduo para a verossimilhança é expressa como

$$f(y_{i1}, \dots, y_{iq}) \prod_{t=q+1}^{n_i} f(y_{it}|f_{i(t-1)}, f_{i(t-2)}, \dots, y_{i(t-q)}). \quad (7)$$

Vale ressaltar que diferente do caso de dados gaussianos, para dados binários o primeiro termo da verossimilhança apresentada em (7) não é possível ser especificado completamente. Como alternativa, para os modelos logístico por exemplo, uma proposta é estimar β e α maximizando a verossimilhança condicional

$$\prod_{i=1}^m f(y_{i(q+1)}, \dots, y_{in_i}|y_{i1}, \dots, y_{iq}) = \prod_{i=1}^m \prod_{t=q+1}^{n_i} f(y_{it}|\mathbf{h}_{it}). \quad (8)$$

O processo de maximização desta função pode ser formulado de forma análoga ao método de mínimos quadrados ponderados iterativamente. Definindo $\delta = (\beta, \alpha)$, a derivada do logaritmo da função da verossimilhança condicional ou função escore condicional tem a forma

$$S^C(\delta) = \sum_{i=1}^m \sum_{t=q+1}^{n_i} \frac{\partial \mu_{it}^C}{\partial \delta} v_{it}^{c-1}(y_{it} - \mu_{it}^C) = 0. \quad (9)$$

Dessa forma, obtém-se o $\hat{\delta}$ por meio de um processo iterativo.

2.1.3 Modelo de primeira ordem

Um caso particular pode ser definido ao considerar um modelo de regressão logística para dados binários que compreende uma cadeia de Markov de primeira ordem (Zeger et al., 1985). Neste contexto, o preditor linear do modelo linear generalizado de transição pode ser expresso

$$g(\mu_{it}^C) = \eta_{it} = \text{logit} \{P(Y_{it} = 1 | \mathbf{h}_{it})\} = \mathbf{x}'_{it}\boldsymbol{\beta} + \boldsymbol{\alpha}y_{i(t-1)}, \quad (10)$$

sendo que $g(\mu_{it}^C) = \text{logit} \{\mu_{it}^C\} = \log \left(\frac{\mu_{it}^C}{1-\mu_{it}^C} \right)$ é a função de ligação, $v(\mu_{it}^C) = \mu_{it}^C(1 - \mu_{it}^C)$ é a função de variância e $f_r(\mathbf{h}_{it}, \boldsymbol{\alpha}) = \alpha_r y_{i(t-r)}$. Para cadeias de Markov com ordem superior a um, este resultado pode ser generalizado para a seguinte expressão

$$\text{logit} \{P(Y_{it} = 1 | \mathbf{h}_{it})\} = \mathbf{x}'_{it}\boldsymbol{\beta}_q + \sum_{r=1}^q \boldsymbol{\alpha}_r y_{i(t-r)}. \quad (11)$$

Diggle et al. (2002) propôs uma forma concisa para um modelo logístico de transição de primeira ordem:

$$\text{logit} \{P(Y_{it} = 1 | Y_{i(t-1)} = a)\} = \mathbf{x}'_{it}\boldsymbol{\beta}_0 + a\mathbf{x}'_{it}\boldsymbol{\alpha}, \quad (12)$$

sendo que $a = y_{i(t-1)}$ (que pode ser 0 ou 1) e de modo que $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0 + \boldsymbol{\alpha}$. Esta equação inclui como preditores as respostas anteriores bem como suas interações com as variáveis explanatórias. Outras funções de ligação, tais como *probit* e complemento *log-log*, podem ser utilizadas caso a função de ligação *logit* não seja a adequada.

Utilizando a definição $\pi_{ab} = P(Y_{it} = b | Y_{i(t-1)} = a)$ com a, b assumindo os valores 0 ou 1, a matriz 2×2 de probabilidades de transição pode então ser obtida da seguinte forma

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

sendo

$$\pi_{00} = \frac{1}{1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta})}, \quad \pi_{01} = 1 - \pi_{00}, \quad \pi_{11} = \frac{\exp(\mathbf{x}'_{it}\boldsymbol{\beta} + \boldsymbol{\alpha})}{1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta} + \boldsymbol{\alpha})} \quad \text{e} \quad \pi_{10} = 1 - \pi_{11}.$$

2.2 Modelo de transição marginalizado

2.2.1 Definição

Azzalini (1994) estabelece que, na análise de dados binários longitudinais, a probabilidade de sucesso em cada ocasião é formulada através de uma regressão logística, entretanto a estrutura para regular a dependência serial entre observações sucessivas pode ser feita pelo parâmetro de autocorrelação de ordem um ou pela

razão de chances. Vale ressaltar que com dados binários, o intervalo de variação do parâmetro da correlação é restrito pelas probabilidades marginais.

Baseando-se nos modelos de Azzalini (1994), Heagerty (2002) propôs um modelo alternativo que denominou de modelo de transição marginalizado. Nesta abordagem, há uma combinação de um modelo de regressão marginal utilizado para caracterizar a dependência da resposta nas covariáveis e um modelo de regressão condicional, ou modelo de transição, usado para capturar a dependência de série no processo de resposta e identificação de uma função de verossimilhança.

Gonçalves e Azzalini (2008) apresentaram outra abordagem para análise de dados binários longitudinais, também baseando-se em Azzalini (1994), e demonstraram algebricamente uma equivalência com a proposta de Heagerty (2002), podendo assim também ser denominada de modelo de transição marginalizado. De forma mais simplificada, propuseram o uso de cadeias de Markov de até segunda ordem na modelagem marginal de dados desta natureza através da razão de chances como parâmetro que estabelece a dependência.

Nesta abordagem introduz-se um modelo de regressão logística que relaciona as covariáveis com a distribuição de probabilidade da resposta

$$\text{logit}(\mu_{it}) = \mathbf{x}'_{it}\boldsymbol{\beta}, \quad (13)$$

sendo $\mu_{it} = P\{Y_{it} = 1\}$. Em uma cadeia de Markov de primeira ordem, a distribuição conjunta da resposta é especificada pela distribuição de Y_1 (pode-se aqui omitir o índice i , pois consideram-se os indivíduos independentes) e de um conjunto de probabilidades condicionais definidas

$$\pi_{1j} = P(Y_t = 1 | Y_{(t-1)} = j), \quad (t = 1, 2, \dots, n_i), \quad j = (0, 1), \quad (14)$$

sendo n_i o número de observações no tempo, $i = 1, 2, \dots, m$ e m perfis individuais. Para simplificar a notação pode-se escrever $\pi_{1j} = \pi_j$. Considerando $(Y_{(t-1)}, Y_t)$ um par de variáveis sucessivas e que a distribuição de $Y_{(t-1)}$ é conhecida, escolhe-se as π_j 's de modo que $\mu_t = E(Y_t)$ tem um valor pré-definido, normalmente especificado através de (13). Desta forma, permite-se selecionar outro componente da distribuição conjunta para regular a dependência. No caso de dados binários, a razão de chances (OR) é considerada a medida mais conveniente, expressa como

$$\psi_1 = OR(Y_t, Y_{(t-1)}) = \frac{P(Y_{(t-1)} = Y_t = 1) P(Y_{(t-1)} = Y_t = 0)}{P(Y_{(t-1)} = 0, Y_t = 1) P(Y_{(t-1)} = 1, Y_t = 0)}, \quad (15)$$

que pode ser representada em função das probabilidades de transição

$$\psi_1 = \frac{\pi_1/(1 - \pi_1)}{\pi_0/(1 - \pi_0)}. \quad (16)$$

Rescrevendo a média na forma

$$\mu_t = \mu_{(t-1)}\pi_1 + (1 - \mu_{(t-1)})\pi_0 \quad (17)$$

e resolvendo esta expressão conjuntamente com a (16) em função de π_0 e π_1 , tem-se a solução encontrada em Azzalini (1994). Rearrmando esta solução, encontra-se a seguinte forma para as probabilidades de transição

$$\pi_j = \frac{(2j-1)[1-\delta_0+(\psi_1-1)\mu_{(t-1)}] + (\psi_1-1)\mu_t}{2(\psi_1-1)[1-j+(2j-1)\mu_{(t-1)}]}, \quad \text{com } j=0,1, \quad (18)$$

sendo que $\delta_0^2 = 1 + (\psi_1 - 1)\psi_1(\mu_t - \mu_{(t-1)})^2 - (\mu_t + \mu_{(t-1)})^2 + 2(\mu_t + \mu_{(t-1)})$. Quando $\psi_1 = 1$, as variáveis são independentes e $\pi_j = \mu_t$. De forma análoga, para cadeias de Markov de segunda ordem tem-se

$$OR(Y_{(t-1)}, Y_{(t-2)}) = \psi_1 = OR(Y_{(t-1)}, Y_t) \quad (19)$$

e

$$OR(Y_{(t-2)}, Y_t | Y_{(t-1)} = 0) = \psi_2 = OR(Y_{(t-2)}, Y_t | Y_{(t-1)} = 1). \quad (20)$$

No contexto de dados binários, estas condições fornecem uma parametrização cuja interpretação é similar ao coeficiente de auto-correlação parcial de um processo Gaussiano, transferido para a escala de razão de chances.

2.2.2 Estimação

Para a estimação dos parâmetros, a função de verossimilhança baseia-se em uma amostra de m perfis individuais, assumidos serem independentes. O logaritmo da função de verossimilhança para os parâmetros (β, λ) , com $\lambda = (\lambda_1, \lambda_2) = (\log \psi_1, \log \psi_2)$, é dado por (Gonçalves et al., 2012)

$$\begin{aligned} \ell^F(\beta, \lambda) = & [y_1 \logit(\mu_1) + \log(1 - \mu_1)] + [y_2 \logit(\pi_{y_1}) + \log(1 - \pi_{y_1})] + \\ & \sum_{t=3}^{n_i} [y_t \logit(\pi_{y_{(t-2)}, y_{(t-1)}}) + \log(1 - \pi_{y_{(t-2)}, y_{(t-1)}})], \end{aligned} \quad (21)$$

sendo que $\pi_j = P(Y_t = 1 | Y_{(t-1)} = j)$, $\pi_{hj} = P(Y_t = 1 | Y_{(t-2)} = h, Y_{(t-1)} = j)$, para $h, j = 0, 1$. O logaritmo da função de verossimilhança completa é obtido com a soma das m contribuições individuais de (21), cujo processo de maximização é realizado por meio de métodos numéricos.

3 Material

As duas propostas para modelos de transição foram implementadas por meio do *software* livre R (R Core Team, 2013) e, com o propósito de facilitar a apresentação nos gráficos e tabelas, receberam as seguintes denominações:

Modelagem M1: ajuste do modelo de transição usual por meio da função *geeglm* do pacote *geepack* (Hojsgaard, 2006). Esta função ajusta equações de estimação generalizada fundamentada na teoria dos MLGs, possibilitando acrescentar a

resposta anterior como covariável e o procedimento de estimação é o método de quase-verossimilhança;

Modelagem M2: ajuste do modelo de transição marginalizado por meio da função *bild* do pacote *bild* (Gonçalves et al., 2012). Esta função ajusta um modelo de regressão logística para dados binários longitudinais, permitindo a dependência serial entre as observações de um determinado indivíduo representada através da razão de chances. Permite também a introdução de um efeito aleatório no intercepto. A maximização do logaritmo da função de verossimilhança é realizada por meio de métodos numéricos.

Foram gerados 2000 conjuntos de dados provenientes de um delineamento experimental inteiramente casualizado, balanceado, com três tratamentos, repetidos dez e vinte vezes, produzindo trinta e sessenta unidades de observação por tempo, respectivamente. Além disso, foram considerados os casos de quatro a onze medidas longitudinais em cada uma das unidades experimentais.

A primeira observação do vetor de respostas binárias repetidas foi gerada através do mecanismo probabilístico

$$P(Y_{i1} = 1) = 1 / [1 + \exp(-x_{i1}\beta)]. \quad (22)$$

Para gerar as demais observações do vetor, considerou-se a suposição de estacionariedade de primeira ordem de modo a adotar o mecanismo probabilístico

$$P(Y_{it} = 1) = 1 / \{1 + \exp [(-x_{it}\beta - \alpha y_{i(t-1)})]\}, \quad (23)$$

com $t = 4, 5, \dots, 11$, sendo que α representa o parâmetro da covariável “resposta no tempo anterior” e β representa o vetor de parâmetros para o tratamento e tempo.

Os valores dos parâmetros “resposta anterior” e “tempo” (sendo este último somente necessário para a modelagem **M2**) foram atribuídos no sentido apenas de compor a estrutura de modelagem (Tabela 1), pois o interesse concentrou-se apenas em avaliar as estimativas para o “intercepto”, “tratamento B” e “tratamento C” sem levar em consideração interações com outras covariáveis.

As comparações foram realizadas por meio do erro padrão e viés (em porcentagem) das estimativas dos parâmetros, além do percentual de cobertura dos intervalos de confiança. A média das estimativas foi calculada de modo que um por cento das observações maiores e menores foram retiradas (média “aparada”).

4 Resultados

Considerando os dois cenários, as Figuras 1 e 2 mostram um decréscimo do erro padrão à medida que aumenta o número de medidas repetidas nas duas modelagens. Este decréscimo também é observado quando se aumenta o número de repetições. Para as estimativas do efeito dos tratamentos B e C, e para um número menor de repetições no tempo, a modelagem **M2** apresentou valores menores para o erro

Tabela 1 - Valores atribuídos aos parâmetros para as modelagens **M1** e **M2** considerando dois diferentes cenários

Parâmetro	diferentes conjuntos		mesmo conjunto
	M1	M2	M1 e M2
Intercepto	-1,052	-2,243	-1,702
Tratamento B	-0,832	-0,977	0,793
Tratamento C	1,029	0,856	1,465
Resposta anterior	0,167	0,159	1,650
Tempo	-	0,381	0,133

padrão. Em contrapartida, para experimentos com um número maior de repetições no tempo, a modelagem **M1** apresenta valores menores para esta medida. No caso do intercepto, a modelagem **M1** apresentou o erro padrão menor independentemente do número de medidas repetidas e do número de repetições dos tratamentos.

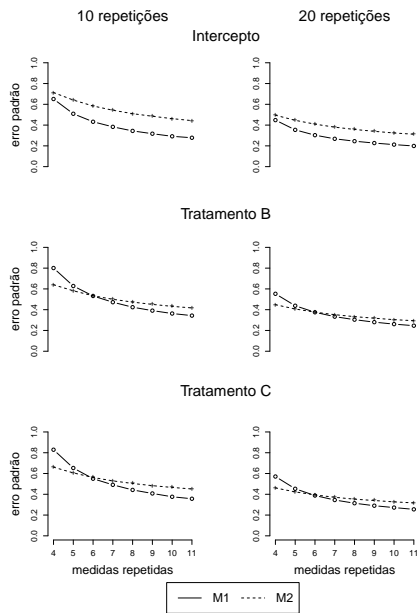


Figura 1 - Mesmo conjunto.

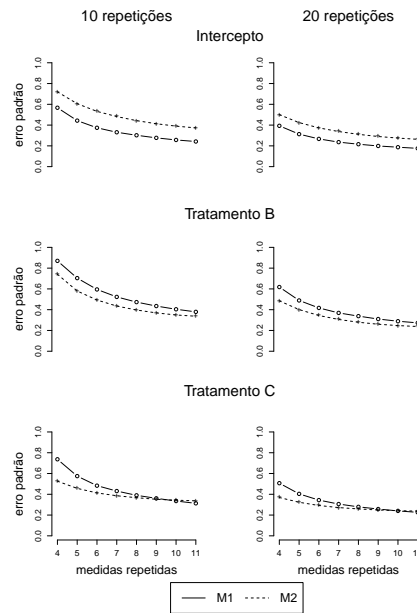


Figura 2 - Diferentes conjuntos.

Quando a comparação é feita pelo percentual de intervalos de confiança das estimativas que contém o verdadeiro valor do parâmetro (Figuras 3 e 4), para o primeiro cenário de se considerar o mesmo conjunto de parâmetros, a modelagem **M1** apresentou, de um modo geral, percentuais maiores do que a modelagem **M2**.

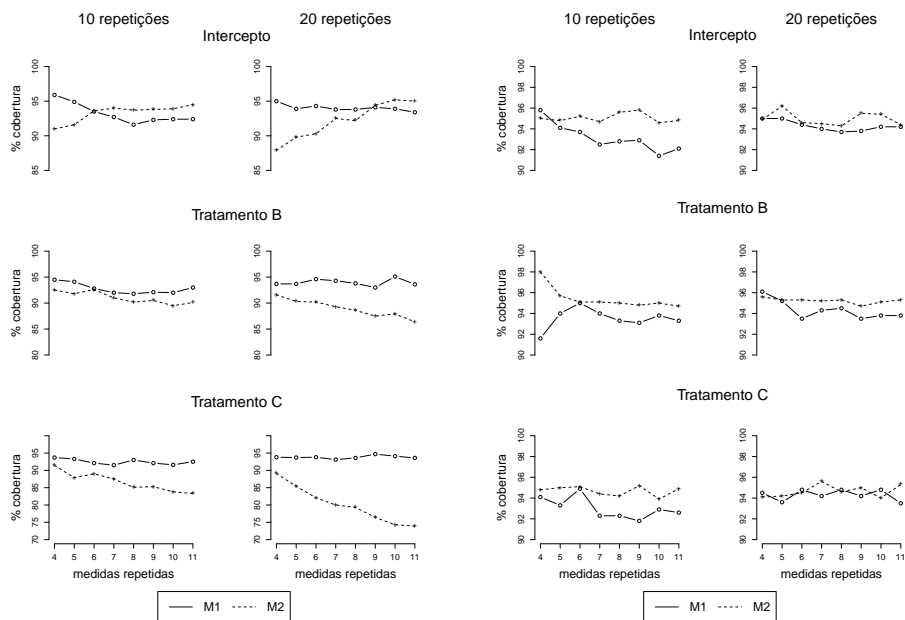


Figura 3 - Mesmo conjunto.

Figura 4 - Diferentes conjuntos.

Já para o cenário com diferentes conjuntos de parâmetros, observou-se um padrão inverso, ou seja, os percentuais de cobertura foram maiores para a modelagem **M2**.

Na comparação do viés (Tabela 2), observou-se que para o cenário com o mesmo conjunto de parâmetros, a modelagem **M1** apresentou valores menores para esta medida para os três tratamentos, independentemente do número de repetições e de medidas repetidas. Além disso, apresentou uma tendência de redução à medida que o número de medidas repetidas aumenta, fato que foi apenas constatado para a modelagem **M2** ao se analisar o intercepto.

No cenário de diferentes conjuntos de parâmetros, o decréscimo do viés ao longo do tempo é mais acentuado para a modelagem **M1** do que para a modelagem **M2**. Quando se considera um experimento com um número pequeno de repetições e também de medidas repetidas, a modelagem **M1** apresentou valores maiores para esta medida nos tratamentos B e C, com destaque para um valor muito acima dos demais valores encontrados para o viés em todos os cenários analisados.

Tabela 2 - Viés para estimativas das modelagens **M1** e **M2** considerando os dois cenários para o conjunto de parâmetros

Medidas repetidas	Número de Repetições	Parâmetros											
		Mesmo conjunto						Diferentes conjuntos					
		Intercepto		Trat B		Trat C		Intercepto		Trat B		Trat C	
		M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
4	10	7,5	29,9	14,7	30,7	13,9	30,1	5,6	6,6	248,9	13,5	14,4	9,9
	20	4,8	25,9	8,5	27,3	5,7	24,5	4,3	3,1	6,7	6,0	7,4	5,0
5	10	4,1	24,6	10,4	35,4	10,0	32,2	2,3	5,4	40,3	8,4	8,6	7,1
	20	2,1	20,5	4,0	30,7	4,7	27,3	1,6	3,1	4,8	5,7	3,0	5,1
6	10	2,7	17,7	9,5	31,6	9,1	31,0	0,8	4,4	10,1	8,1	6,5	5,2
	20	0,8	16,5	2,8	31,2	4,1	28,4	-0,4	2,7	5,0	3,6	3,7	4,2
7	10	2,8	14,1	8,4	34,3	7,8	31,2	2,1	4,5	3,8	5,3	4,8	7,7
	20	0,7	11,7	1,4	29,2	2,6	28,8	0,0	2,3	3,1	3,9	1,7	4,3
8	10	2,4	11,0	6,5	35,8	5,8	31,8	2,1	3,4	4,2	6,7	5,6	5,0
	20	1,0	8,5	2,5	31,8	2,9	28,4	0,5	1,6	2,4	4,5	2,5	3,6
9	10	2,0	9,6	5,6	35,4	4,8	32,0	1,4	2,9	4,9	6,6	3,9	4,1
	20	0,7	6,3	2,9	32,7	1,7	29,6	0,6	1,3	1,0	4,6	2,0	3,3
10	10	1,1	6,8	3,4	37,5	3,8	33,0	0,5	2,7	5,1	4,9	2,6	5,7
	20	0,3	4,0	1,0	31,8	1,8	30,1	-0,2	1,7	1,6	4,2	1,0	4,3
11	10	1,7	4,1	5,1	34,7	4,9	31,7	1,1	2,5	2,6	5,0	3,6	5,0
	20	0,5	2,4	1,5	34,5	2,2	29,4	0,1	1,1	1,9	4,3	1,7	2,8

Conclusões

De um modo geral, não ocorreram resultados seguindo um mesmo padrão para os três tratamentos de natureza categóricas incluídos no delineamento. Ficou constatado que a modelagem **M1**, caracterizada pelo ajuste do modelo de transição usual, de uma forma mais ampla, apresentou resultados melhores para o erro padrão, viés e percentual de cobertura dos intervalos de confiança mostrando ser uma boa alternativa para delineamentos com medidas repetidas no tempo quando comparado com os resultados provenientes do ajuste do modelo de transição marginalizado com parâmetro de dependência a razão de chances (modelagem **M2**).

Entretanto, vale ressaltar que a modelagem **M1** pode não ser indicado para experimentos delineados com um número reduzido de medidas longitudinais e de repetições por tratamentos. A justificativa se deve ao fato desta modelagem ter apresentado valores para o viés bastante elevados em um dos cenários avaliados, diferenciando de forma muito expressiva dos resultados da modelagem **M2**. Isto ocorreu mesmo adotando o procedimento de eliminar um por cento dos valores maiores e menores do erro padrão, viés e percentual de cobertura dos intervalos de confiança obtidos por meio dos ajustes dos dois modelos comparados.

Agradecimentos

Os autores agradecem ao Centro Nacional de Supercomputação da Universidade Federal do Rio Grande do Sul pelo suporte computacional.

LORDÉLO, M. S.; PIEDADE, S. M. S.; FERNANDES, G. B.; FIACCONE, R. L. Markov transition models: a focus on planned experiments with correlated binary data. *Rev. Bras. Biom.*, São Paulo, v.33, n.1, p.78-90, 2015.

■ **ABSTRACT:** *The transition Markov models are a very important tool for several areas of knowledge when studies are developed with repeated measures. They are characterized by modeling the response variable over time conditional to the previous response which is known as the history. In addition it is possible to include other covariates. In the case of binary responses, can be constructed a matrix of transition probabilities from one state to another. In this work, two different approach of transition models were compared in order to assess which best estimates of the causal effect of treatments in an experimental studies where the outcome is a vector of binary response measured over time. Simulation study was held taking into account a balanced experiments with three treatments of categorical nature. To assess the best estimates standard error, bias and percentage of coverage were used. The results showed that the marginal transition models may be more appropriate in situations in which an experiment is developed with a reduced number of repeated measurements.*

■ **KEYWORDS:** *Binary data; repeated measures; causal treatment effect.*

Referências

AZZALINI, A. Logistic regression for autocorrelated data with application to repeated measures. *Biometrika*, London, v. 81, n. 4, p.767-775, 1994.

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

DIGGLE, P.J.; HEAGERTY, P.J.; LIANG, K.Y.; ZEGER, S.L. *Analysis of longitudinal data*. 2.ed. New York: Oxford University Press. 2002. 379 p.

GONÇALVES, M.H.; AZZALINI, A. Using Markov chains for marginal modelling of binary longitudinal data in an exact likelihood approach. *METRON - International Journal of Statistics*, Roma, v.66, n.2, p.157-181, 2008.

GONÇALVES, M.H.; CABRAL, M.S.; AZZALINI, A. The R Package bild for the Analysis of Binary Longitudinal Data. *Journal of Statistical Software*, Los Angeles, v.46, n.9, p.1-17, 2012.

HEAGERTY, P.J. Marginalized transition models and likelihood inference for longitudinal categorical data. *Biometrics*, London, v.58, p.342-351, 2002.

HOJSGAARD, S., HALEKOH, U., YAN, J. The R Package geepack for Generalized Estimating Equations. *Journal of Statistical Software*, Los Angeles, v.15, n.2, p.1-11, 2006.

LARA, I. A. R. *Modelos de transição para dados binários*, 2007. 128 p. Tese (Doutorado em Agronomia)-Escola Superior de Agricultura 'Luiz de Queiroz', Universidade de São Paulo, Piracicaba, 2007.

LIANG, K.Y.; ZEGER, S.L. Longitudinal data analysis using generalized linear models. *Biometrika*, London, v.73, n.1, p.13-22, 1986.

R CORE TEAM. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Disponível em: <<http://www.R-project.org>>, 2013.

ZEGER, S.L.; LIANG, K.Y.; SELF, S.C. The analysis of binary longitudinal data with time-independent covariates. *Biometrika*, London, v.72, p.8-31, 1985.

Recebido em 22.09.2014.

Aprovado após revisão em 18.01.2015.