

MODELOS DE TRANSIÇÃO PARA DADOS BINÁRIOS: TESTES PARA COMPARAR TRATAMENTOS

Idemauro Antonio Rodrigues de LARA¹
Clarice Garcia Borges DEMÉTRIO¹
Dalton Francisco de ANDRADE²
João Maurício Araújo MOTA³

- RESUMO: O enfoque do presente trabalho é dado aos modelos lineares generalizados de transição adequados 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. O método da máxima verossimilhança é utilizado para o ajuste dos modelos e estimação das probabilidades. Adicionalmente, são apresentados testes assintóticos para comparar tratamentos, baseados na razão de chances e na diferença das probabilidades de transição. Os métodos são ilustrados com um conjunto de dados sobre doença respiratória. Para esses dados, o processo é considerado estacionário de ordem dois e o teste proposto, sinaliza diferença estatisticamente significativa a favor do tratamento ativo.
- PALAVRAS-CHAVE: Dados longitudinais; modelo linear generalizado; processos estocásticos; probabilidades de transição; máxima verossimilhança.

1 Introdução

Dados binários são comuns em muitas áreas das ciências, nas quais, muitas vezes, há interesse em registrar a ocorrência, ou não, de um evento particular. Nos

¹Departamento de Ciências Exatas, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo – ESALQ/USP, CEP: 13418-900, Piracicaba, São Paulo, Brasil. E-mail: idemauro@esalq.usp.br / clarice@esalq.usp.br

²Departamento de Informática, Universidade Federal de Santa Catarina – UFSC, CEP: 88040-970, Florianópolis, SC, Brasil. E-mail: dandrade@inf.ufsc.br

³Departamento de Estatística, Universidade Federal do CEARÁ – UFC, CEP: 60020-181, Fortaleza, CE, Brasil. E-mail: oiCiruam@ufc.br

estudos longitudinais, com resposta dessa natureza, existe uma dicotomia conceitual e metodológica para a análise desses dados. Sabe-se que a escolha do modelo a ser utilizado depende basicamente dos objetivos do experimento por ocasião do planejamento da pesquisa. O ajuste dos modelos marginais por meio das equações de estimação generalizadas, EEGs, (LIANG; ZEGGER, 1986; ZEGGER; LIANG, 1986), permitem modelar estruturas de correlação e resultam em estimativas consistentes para os parâmetros de regressão. No entanto, esses modelos não permitem avaliar o comportamento das mudanças de respostas em cada ocasião. Portanto, para atender a esse objetivo, deve-se optar por um modelo de transição. Em particular, para dados binários, os modelos de transição se caracterizam pelas probabilidades condicionais de resposta no tempo t , dada a história passada da unidade amostral ou experimental. Devido a isso, esses modelos 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.

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. Neste trabalho, o enfoque é direcionado para a cadeia de Markov, uma vez que a variável resposta tem espaço de estados discreto, $S = \{0, 1\}$, e, o tempo, também é considerado discreto $\tau = \{0, 1, 2, \dots, T\}$. Assim, a descrição das mudanças ou transições das observações entre uma ocasião e outra é caracterizada estocasticamente por matrizes de probabilidades de transição, de dimensões 2×2 . Para estabelecer a notação, considere a propriedade de Markov para cadeias de primeira ordem:

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

a qual leva às matrizes de probabilidades,

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

em que o parâmetro (t) denota a dependência em relação ao tempo, havendo nesse caso, T matrizes de transição. A matriz \mathbf{P} é a matriz associada a um processo estacionário ou homogêneo no tempo. Registra-se que a cadeia pode ter alcance maior do que 1. Para uma cadeia de Markov de ordem 2, 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 4×2 :

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

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

Alguns trabalhos como os de Anderson e Goodman (1957), Goodman (1962), Bishop, Fienberg e Holland (1975), Agresti (1991), Fitzmaurice e Laird (1993), Lindsey (1995, 2004) descrevem que as mudanças ou transições de categorias de resposta no tempo, quando a amostra é homogênea, podem ser estudadas através de técnicas aplicadas a tabelas de contingência. Na prática, porém, conforme destaca Noleto (1991), dificilmente a amostra é considerada homogênea. Uma alternativa para esses casos é especificar um modelo para descrever a relação funcional de dependência dessa probabilidade em relação ao conjunto de covariáveis consideradas no estudo. Os modelos lineares generalizados (MLG) podem ser usados para tal propósito. Algumas referências clássicas com dados binários são Cox (1970), Korn e Whittemore (1979), Azzalini (1983), Zeger, Liang e Self (1985).

Nesse sentido, um modelo de transição envolve teoricamente métodos dos modelos lineares generalizados e técnicas de processos estocásticos, uma vez que além da estimação dos parâmetros deseja-se verificar, por exemplo, a 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 que parece relevante é poder comparar os diferentes tratamentos envolvidos a partir das matrizes de probabilidades de transição. Esses problemas, em geral, aparecem como lacunas na literatura desses modelos, justificando estudos nessa área. Nesse contexto, este trabalho tem dois objetivos: o primeiro é apresentar os fundamentos básicos dos modelos de transição para análise de dados longitudinais e o segundo é definir hipóteses para comparação de tratamentos e apresentar propostas de testes mais apropriados para esses casos.

2 Material e métodos

2.1 Material

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 (ativo, placebo) são administrados. As demais covariáveis, 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). Neste trabalho, por simplicidade, faz-se referência ao estado de saúde (respiração do paciente) como bom (1) ou ruim (0).

2.2 Métodos

2.2.1 Definição do modelo

Considere um estudo longitudinal e seja, \mathbf{y}_i o vetor de variáveis respostas do i -ésimo indivíduo, de dimensões $n_i \times 1$, isto é, $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})'$, de tal forma que numa ocasião t , a observação y_{it} esteja associada um vetor de variáveis explicativas ($p \times 1$), $\mathbf{x}_{it} = (x_{it1}, \dots, x_{itp})'$. De acordo com Diggle et al. (2002), 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:

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

em que w_i são os pesos a priori, ϕ um parâmetro de dispersão suposto conhecido, θ_{it} representa o parâmetro canônico, $b(\theta_{it})$ uma função que depende desse parâmetro e $c(y_{it}, \phi)$ uma função que depende da variável aleatória Y_{it} e do parâmetro de dispersão ϕ , seguindo as pressuposições de definição de um MLG, a menos pela inclusão do índice t , que considera as observações repetidas. Demonstra-se que:

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

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

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

$$g(\mu_{it}^C) = \eta_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \sum_{r=1}^s f_r(\mathbf{h}_{it}; \boldsymbol{\alpha}) \quad \text{e} \quad v_{it}^C = v(\mu_{it}^C)\phi, \quad (2)$$

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 (2), que a média condicional depende tanto das covariáveis como das respostas nas ocasiões anteriores. As respostas prévias ou funções delas mesmas são tratadas como variáveis explicativas adicionais. Os parâmetros de interesse, a priori, pode ser representado pelo vetor $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$, em que $\boldsymbol{\beta}$ está associado às covariáveis e $\boldsymbol{\alpha}$ às respostas prévias.

2.2.2 Estimação dos parâmetros δ

Ao se especificar um modelo de transição, não se sabe, se o processo é, ou não, estacionário. Assim, a princípio os modelos devem ser ajustados para o caso não estacionário. Para tanto, a proposta de Ware, Lipsitz e Speizer (1988), originariamente apresentada para dados categorizados ordinais, é ajustar um modelo para cada ocasião como num estudo transversal. Os parâmetros de δ são específicos para cada ocasião e são obtidos maximizando as funções de verossimilhanças em separado. Isso permite modelar as T transições possíveis, ou seja, de $t = 0$ para $t = 1$, de $t = 1$ para $t = 2$ e assim, sucessivamente, até $t = T - 1$ a T . A cadeia de Markov é caracterizada em cada modelo pela inclusão da resposta no tempo anterior como uma covariável independente em um modelo de regressão. Com variáveis ordinais pode-se usar o modelo de chances proporcionais (McCULLAGH, 1980), que se reduz ao modelo de regressão logística para o caso de variáveis binárias.

Para o caso estacionário, Diggle et al. (2002) apresentam um processo de estimação, que 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. 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 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 (1) define apenas a distribuição condicional, portanto, a verossimilhança das q primeiras observações $f(y_{i1}, y_{i2}, \dots, y_{iq})$ não é determinada diretamente, exceto para o caso da distribuição normal. Conseqüentemente, a função de verossimilhança total não pode ser especificada. Uma alternativa, é estimar β e α maximizando a função de verossimilhança condicional:

$$\prod_{i=1}^N f(y_{i(q+1)}, \dots, y_{in_i} | y_{i1}, \dots, y_{iq}) = \prod_{i=1}^N \prod_{t=q+1}^{n_i} f(y_{it} | \mathbf{h}_{it}). \quad (3)$$

Há duas possibilidades distintas para se maximizar a função (3). Numa primeira situação, se $f_r(h_{it}; \alpha, \beta) = \alpha_r f_r(h_{it})$, tal que $g(\mu_{it}^C) = \mathbf{x}'_{it} \beta + \sum_{r=1}^s f_r(h_{it})$ é uma função linear de ambos os parâmetros α e β , 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 $(\mathbf{x}_{it}, f_1(h_{it}), \dots, f_s(h_{it}))$.

De um modo geral, para uma variável aleatória seguindo o modelo (1), o algoritmo proposto por Nelder e Wedderburn (1972) para o ajuste de modelos

lineares generalizados permite obter estimativas de máxima verossimilhança (método escore de Fisher) para δ no preditor linear η por mínimos quadrados ponderados iterativamente.

Para tanto, considere \mathbf{Y}_i um vetor de respostas, com dimensões $(n_i - q)$, para $t = q + 1, \dots, n_i$ e μ_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 $\frac{\partial \mu_{i(q+k)}}{\partial \delta}$ 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 $\mathbf{Z}_i = \mathbf{X}_i^* \hat{\delta} + (\mathbf{Y}_i - \hat{\mu}_i^C)$. Então, o valor atualizado $\hat{\delta}$ pode ser obtido iterativamente pela regressão de \mathbf{Z} contra \mathbf{X}^* usando a matriz de pesos \mathbf{W} , isto é:

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

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 α e β . Um exemplo dessa situação é dado em Zeger e Qaish (1988), em que a modelagem de dados de contagem é feita a partir de um modelo de transição de Markov estacionário de primeira ordem, com $f = \alpha[\log(y_{i(t-1)} - \mathbf{x}'_{i(t-1)}\beta)]$, 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.

2.2.3 Matriz das probabilidades de transição da cadeia de Markov para dados binários

A estimação das probabilidades, $P[Y_{it} = y_{it} | \mathbf{h}_{it}, \mathbf{x}_{it}]$, 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 covariáveis). 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:

$$\eta_{it} = \text{logito} [P(Y_{it} = b | Y_{i(t-1)} = a)] = \ln \left(\frac{\pi_{ab}(t)}{1 - \pi_{ab}(t)} \right) = \mathbf{x}'_{it} \beta + \alpha y_{i(t-1)}. \quad (4)$$

A caracterização das matrizes de probabilidades de transição, 2×2 , com elementos $\pi_{ab}(t) = P(Y_{it} = b \mid Y_{i(t-1)} = a)$, em que a e b assumem valores em $S = \{0, 1\}$, pode ser feita a partir da equação definida em (4), por meio de:

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

avaliando-se η_{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:

$$\eta_{it} = \text{logito} [P(Y_{it} \mid Y_{i(t-1)})] = \ln \left(\frac{\pi_{ab}(t)}{1 - \pi_{ab}(t)} \right) = \mathbf{x}'_{it}\boldsymbol{\beta} + y_{i(t-1)}\mathbf{x}'_{it}\alpha, \quad (5)$$

sendo a equação (4) um caso particular da equação (5), que inclui todas as interações entre a resposta prévia e as covariáveis, resultando em vantagens em termos de inferência.

Os resultados apresentados partem da pressuposição de que a função de ligação adequada é a logito. Para os casos em que essa função não seja satisfatória, pode-se optar por outras funções de ligação, como probito e complemento *log-log*. 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 $f_r(h_{it}, \alpha) = \alpha_r y_{i(t-r)} \mathbf{x}_{it}$ 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:

$$\begin{aligned} \eta_{it} &= \text{logito} [P(Y_{it} = c \mid Y_{i(t-2)} = b, Y_{i(t-1)} = a)] \\ &= \mathbf{x}'_{it}\boldsymbol{\beta} + y_{i(t-1)} \mathbf{x}_{it} \boldsymbol{\alpha}_1 + y_{i(t-2)} \mathbf{x}_{it} \boldsymbol{\alpha}_2 + y_{i(t-1)} y_{i(t-2)} \mathbf{x}_{it} \boldsymbol{\alpha}_3, \end{aligned} \quad (6)$$

com $a, b, c \in S = \{0, 1\}$. Assim, considerando o preditor linear (6) 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:

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

Deve ser observado que como a matriz de probabilidades de transição depende das variáveis explicativas, $P(t)$ 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. A condição de estacionariedade do processo, para uma cadeia de alcance 1, tem como hipótese principal, $H_0 : \pi_{ab}(t) = \pi_{ab}$, ou em termos matriciais, $H_0 : P(t) = P$ para todo $t \in \tau$. Good (1955), Anderson e Goodman (1957) sugerem dois critérios para se testar essa hipótese, sendo um deles o de qui-quadrado, com estatística:

$$\xi = \sum_t \sum_a \sum_b \frac{n_a(t-1)[\hat{\pi}_{ab}(t) - \hat{\pi}_{ab}]^2}{\hat{\pi}_{ab}}, \quad (7)$$

em que $n_a(t-1) = \sum_b n_{ab}(t)$ e $n_{ab}(t)$ são as freqüências de transições observadas, $\hat{\pi}_{ab}$ e $\hat{\pi}_{ab}(t)$ são os estimadores das probabilidades de transição, sob H_0 e sob H_a , respectivamente.

Outro procedimento é usar o critério da razão de verossimilhanças. As funções de verossimilhança maximizadas, sob as hipóteses nula e alterantiva, são:

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

respectivamente. Portanto:

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

e como extensão do teorema de Neyman-Pearson, tem-se a seguinte estatística:

$$\lambda = -2 \sum_t \sum_a \sum_b n_{ab}(t) \log \left(\frac{\hat{\pi}_{ab}}{\hat{\pi}_{ab}(t)} \right), \quad (8)$$

sendo que, sob H_0 as estatísticas (7) e (8) são equivalentes e assintoticamente convergem para uma distribuição χ^2 com $2(T-1)$ graus de liberdade.

Anderson e Goodman (1957) também apresentam testes para se verificar alcance da cadeia. Assim como nos procedimentos anteriores, os testes necessitam que sejam estimadas as probabilidades de transição. Uma alternativa, para o caso em que há variáveis explicativas é construir o teste da razão de verossimilhanças, sob os logaritmos das funções de verossimilhanças dos modelos de transição. A hipótese nula de interesse é:

H_0 : A cadeia é de $(q - 1)$ -ésima ordem;

contra a alternativa de que a cadeia é de q -ésima ordem. Sejam $\ell(\hat{\beta}, \hat{\alpha}, \mathbf{y})_{(q-1)}$ e $\ell(\hat{\beta}, \hat{\alpha}, \mathbf{y})_{(q)}$ os logaritmos das funções de verossimilhança maximizadas sob as condições de ordem $(q - 1)$ e q , respectivamente. Assim, para a situação mais elementar, isto é, a menos de interações entre a resposta prévia e demais covariáveis, a estatística para o teste da razão de verossimilhanças é:

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

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. Quanto aos procedimentos computacionais, *softwares* padrões podem ser utilizados para o ajuste desses modelos, como os procedimentos *logistic* e *genmod* do SAS (SAS Institute Inc, 1999) ou os pacotes, *glm*, *gee* e *geepack* do R (R Development Core Team, 2007). Para o caso estacionário, necessita-se organizar o conjunto de dados de tal forma a incorporar as respostas prévias como covariáveis extras, ignorando as observações referentes a ordem da cadeia. Para esses casos, usa-se a macro *DROPOUT*, proposta por Molenberghs e Verbeke (2005), desenvolvida no programa SAS.

2.2.4 Testes para comparação de tratamentos

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. 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 como apresentado na Tabela 1.

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

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

Tabela 1 - 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)	$\pi_{a0}(A)$	$\pi_{a1}(A)$

No entanto, há outras formulações possíveis e equivalentes. 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 1, 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 que quando $\psi > 1$, 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 ($\pi_{a1}(A) > \pi_{a1}(P)$). Diante dessa explanação, é razoável também formular as seguintes hipóteses:

$$\text{Hipóteses}^{(2)} : \text{em termos de razão de chances} \begin{cases} H_0 : \psi = 1; \\ H_a : \psi > 1. \end{cases}$$

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

$$H_0 : \psi_{Y_{(t-1)}=0} = \psi_{Y_{(t-1)}=1}. \quad (9)$$

Considerando o modelo estocástico:

$$\eta = \ln \left(\frac{\pi_{ab}}{1 - \pi_{ab}} \right) = \beta_0 + \beta_1 \text{tratamento} + \alpha y_{(t-1)} + \gamma(y_{(t-1)}) * \text{tratamento}, \quad (10)$$

a razão de chances associada às repostas prévias podem ser obtidas em termos dos coeficientes desse modelo, de modo que o teste da hipótese (9) equivale ao teste $H_0 : \gamma = 0$ pelo modelo definido em (10). Esse teste pode ser feito pela razão de verossimilhanças ou pela diferença de *deviances*, considerando-se os modelos encaixados sem e com interação. Uma extensão para um modelo de Markov de alcance q , teria como hipótese natural para a homogeneidade de razão de chances,

$H_0 : \gamma = \mathbf{0}$, sendo γ o vetor de parâmetros de todas as interações do tratamento com as respostas prévias. 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.

Teste da razão de chances

No contexto da análise de dados categorizados, é conhecida a estatística de Mantel e Haenszel (1959), que avalia a não existência de associação entre as linhas (independentes) e as colunas em tabelas de contingência 2×2 , mantendo-se o efeito da estratificação em relação a um terceiro fator (AGRESTI, 1991 e PAULA, 2004). A idéia básica do teste, que tem como hipótese principal $H_0 : \psi = 1$, é a distribuição normal assintótica de uma distribuição condicional.

Considere a Tabela 2, 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.

Tabela 2 - Estrutura usual de uma tabela de contingência 2×2 para o teste de Mantel-Haenszel

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

Sob essa estrutura, a distribuição condicional $(y_1 | m, \psi)$ produz um modelo pertencente à família de distribuições hipergeométricas não centrais. Em particular, para $\psi = 1$, esse modelo reduz-se à distribuição hipergeométrica central:

$$f(y_1 | m, \psi = 1) = \frac{\binom{n_1}{y_1} \binom{n_2}{y_2}}{\binom{n_1 + n_2}{m}}, \quad (11)$$

com média e variância dadas por:

$$E(Y_1 | m, \psi = 1) = \frac{mn_1}{n} \quad \text{e} \quad \text{Var}(Y_1 | m, \psi = 1) = \frac{n_1 n_2 m (n - m)}{n^2 (n - 1)}.$$

Quando n_1 , n_2 , m e $n - m$ são suficientemente grandes, a distribuição (11) tende à distribuição normal e conseqüentemente:

$$Z = \frac{y_1 - \frac{mn_1}{n}}{\sqrt{\frac{n_1 n_2 m (n-m)}{n^3}}} \sim N(0, 1), \quad (12)$$

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 e Haenszel (1959), para confrontar as hipóteses⁽³⁾ $H_0 : \psi = 1$ 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 2×2 , uma para cada tipo de resposta prévia, que tem aspecto similar à Tabela 2.

Na Tabela 3, $n_{a0(j)}^*(.)$ 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 o resultado (12) do teste de Mantel-Haenszel, tem-se assintoticamente a seguinte estatística para teste:

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

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

Tabela 3 - Tabela de contingência 2×2 , 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	$n_{a0(j)}^*(P)$	$n_{a1(j)}^*(P)$	$n_{a \cdot (j)}^*(P)$
Ativo	$n_{a0(j)}^*(A)$	$n_{a1(j)}^*(A)$	$n_{a \cdot (j)}^*(A)$
Total marginal	$n_{\cdot 0(j)}^*$	$n_{\cdot 1(j)}^*$	$n_{\cdot \cdot (j)}^*$

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 3 denotam as freqüências de transições de h_{it} para os estados $S = \{0, 1\}$.

Teste da diferença das probabilidades de transição

Sem perda de generalidade, considere as hipóteses que avaliam as mudanças do estado 0 para o estado 1. 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. Como as probabilidades de transição satisfazem à equação de Chapman-Kolmogorov, elas 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:

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

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). 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, isto é,

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

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

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

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

3 Resultados e discussão

Para ilustrar as diferentes formas de se trabalhar com modelos de transição de Markov, a princípio, realiza-se uma análise considerando o efeito de tratamento. Na Tabela 4, são apresentadas as frequências de transição para os estados ruim (0) e bom (1), estratificada por tratamento.

Tabela 4 - Distribuição das transições de acordo com o tipo de tratamento no estudo epidemiológico sobre a condição respiratória

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

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 5 e 6. 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.

Ainda na Tabela 5, 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, π_{01} e π_{11} serem decrescentes no grupo com tratamento ativo pode ser decorrência da perda do efeito do tratamento. Sob a hipótese de estacionariedade, (Tabela 6), dado que o paciente está bom na ocasião $(t-1)$, a probabilidade de transição para o estado bom é cerca de 3,56 vezes maior para o grupo placebo e 2,29 vezes maior para o grupo com tratamento ativo, quando comparadas à condição prévia ruim.

Pela definição de modelos lineares generalizados de transição, os resultados apresentados nas Tabelas 5 e 6 são reproduzidos (estimados), a partir das

Tabela 5 - Número de transições e estimativas de probabilidades, de acordo com o tipo de tratamento e condição respiratória na visita anterior

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

estimativas dos parâmetros do modelo, especificado pelo preditor linear (10). Os resultados apresentados na Tabela 6 são reproduzidos a partir do modelo ajustado para a totalidade das transições.

Na Tabela 7, são apresentadas 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, enquanto que na Tabela 8, apresentam-se as estimativas desses parâmetros para o processo estacionário. Ressalta-se a significância estatística ($p < 0,01$), para as respostas no tempo $(t - 1)$.

Para estimar as probabilidades usam-se essas estimativas. Por exemplo,

Tabela 6 - Total de transições e estimativas de probabilidades, de acordo com o tipo de tratamento e condição respiratória na visita anterior

Tratamento	$(Y_{i(t-1)} = 0)$			$(Y_{i(t-1)} = 1)$		
	$Y_{it} = 0$	$Y_{it} = 1$		$Y_{it} = 0$	$Y_{it} = 1$	
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

Tabela 7 - 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	1ª transição		2ª transição	
	Estimativas	valor p	Estimativas	valor p
intercepto	-0,8938	0,0239	-1,8326	< 0,01
tratamento ativo	0,6256	0,2473	1,4759	0,0432
resposta prévia ($Y_{i(t-1)}$)	1,8923	< 0,01	2,4204	< 0,01
tratamento* $Y_{i(t-1)}$	24,7412	0,9998	-0,4215	0,6546
Parâmetros	3ª transição		4ª transição	
	Estimativas	valor p	Estimativas	valor p
intercepto	-1,0609	< 0,01	-1,9095	< 0,01
tratamento ativo	0,2724	0,6815	0,8979	0,2572
resposta prévia ($Y_{i(t-1)}$)	2,2846	< 0,01	3,3446	< 0,01
tratamento* $Y_{i(t-1)}$	0,6439	0,5152	-1,2683	0,2070

Tabela 8 - 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	Estimativa	Erro-padrão	valor p
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

sob estacionariedade, (Tabela 8), 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 que comparadas ao valor $\chi_{6; 0,01}^2 = 16,81$, mostram evidência de que o processo é estacionário. Matricialmente, tem-se

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

A partir dessas matrizes, pode-se também observar que, para os pacientes com estado de saúde ruim na visita anterior (linhas superiores das matrizes $\hat{P}_{(ativo)}$ e $\hat{P}_{(placebo)}$), o logaritmo da razão de chances, isto é, $\ln[(0,3718/0,6282)/(0,2063/0,7937)] = 0,82$, é a estimativa do coeficiente do efeito de tratamento. Para os pacientes cuja condição de saúde é boa na ocasião $(t - 1)$ (linhas inferiores das matrizes $\hat{P}_{(ativo)}$ e $\hat{P}_{(placebo)}$), tem-se que $\ln[(0,8551/0,1449)/(0,7353/0,2647)] = 0,75$ é a soma dos coeficientes referentes ao efeito de tratamento mais a interação de tratamento e resposta prévia. A razão de chances, $\exp(0,82) = 2,27$ indica que se a condição prévia é ruim, os indivíduos do grupo placebo têm 2,27 vezes maior chance de permanência nesse estado, ao passo que se a condição prévia é boa essa chance cai para 2,11. Na Tabela 9, 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 $(t - 1)$. Os efeitos de centro clínico, tratamento, e resposta prévia são significativos ($p < 0,01$). No modelo 2, adicionalmente, são considerados os efeitos das interações da resposta prévia com essas covariáveis, além do fator ocasião (visita). Nenhuma dessas interações é significativa. 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, independente da história. 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:

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

Tabela 9 - Estimativas dos parâmetros fixos e respectivos erros-padrão, para dois modelos de transição estacionários de primeira ordem, no estudo sobre a condição respiratória

Parâmetros	Modelo 1		Modelo 2	
	Estimativa	Erro-padrão	Estimativa	Erro-padrão
intercepto	-0,3248	0,4104	1,1776	0,7930
centro 1	-0,7622	0,2477	-0,9852	0,3548
sexo (feminino)	0,0506	0,3111	0,1891	0,4486
tratamento (ativo)	0,8382	0,2413	0,8040	0,3566
idade	-0,0177	0,0093	-0,0365	0,0158
visita			-0,2345	0,1509
Y_{it-1}	2,2277	0,2337	0,6575	1,0858
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

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, ou seja,

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

A diferença observada entre as probabilidades é $\pi_{01}^{(4)}(A) - \pi_{01}^{(4)}(P) = 0,2768$, 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, em se tratando de modelos de transição, é importante considerar também a possibilidade de alcance da cadeia maior do que 1.

Dando prosseguimento à análise, na Tabela 10 são apresentadas as freqüências de transições de segunda ordem, as quais permitem explorar a necessidade de se considerar a cadeia de Markov de ordem 2, bem como verificar sua estacionariedade.

A partir da Tabela 10, especificam-se as freqüências sob a hipótese de estacionariedade, conforme mostra a Tabela 11 e, portanto, obtém-se a estatística 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 10 - Frequências 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 (0,2051)	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

Tabela 11 - Total de frequências de transições de segunda ordem observadas no estudo sobre a condição respiratória

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

Sendo $\chi_{8,(0,01)}^2 = 20,09$, tem-se que mesmo para o processo de ordem 2, não se rejeitaria a hipótese de estacionariedade. Pelo critério da razão de verossimilhanças, a estatística obtida é $\lambda = 2,03$. Por outro lado, podem-se, também, aplicar 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 10, também podem ser estimadas as frequências de primeira ordem, sob H_0 , como mostra a Tabela 12.

Assim, as estimativas da Tabela 11 podem ser comparadas às frequências da Tabela 12, fixando-se $Y_{i(t-1)} = 0$ e, posteriormente $Y_{i(t-1)} = 1$. Pelo critério de qui-quadrado, tem-se:

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

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).

Tabela 12 - 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

Os resultados dessa análise exploratória preliminar indicam que é razoável considerar para esse caso um modelo estocástico estacionário de ordem 2. Assim, considerando as demais variáveis explicativas do estudo, são apresentadas na Tabela 13 as estimativas para os parâmetros do modelo de 2ª ordem.

Tabela 13 - Estimativas dos parâmetros e respectivos erros-padrão para o modelo de transição estacionários de segunda ordem, no estudo sobre a condição respiratória

Parâmetros	Estimativa	Erro-padrão	valor p
intercepto	-0,5616	0,4940	0,2557
centro 1	-0,6540	0,2967	0,0275
sexo (feminino)	0,1945	0,3709	0,5999
tratamento (ativo)	0,7339	0,2889	< 0,01
idade	-0,0252	0,0111	0,0232
$Y_{i(t-1)}$	1,9213	0,2979	< 0,01
$Y_{i(t-2)}$	0,9477	0,2974	< 0,01

Para o ajuste desse modelo, são consideradas apenas as observações referentes às visitas 3 a 5, totalizando 333 transições. Assim, comparando os modelos de 1ª e 2ª 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 $Y_{i(t-2)}$. Note que para a decisão da ordem da cadeia, incluindo as demais covariáveis, 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 0,0016. Esse teste equivale à diferença de *deviances* para os modelos encaixados de 1ª e 2ª 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. Com base nos efeitos significativos desse modelo, ($p < 0,01$), as matrizes de probabilidades de transição, para os respectivos

grupos ativo e placebo, são:

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

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. 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-Haenszel 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.

Conclusões

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 adotar 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.

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 podem não ser válidos quando não há um número razoável de freqüências de transições. 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.

Observações omisas 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 omisso como se fosse a primeira observação. Outra alternativa é aumentar a ordem da cadeia, que certamente traz penalizações pelo aumento no número de parâmetros.

LARA, I. A. R.; DEMÉTRIO, C. G. B.; ANDRADE, D. F.; MOTA, J. M. A. Transition models for binary data: tests to compare treatments. *Rev. Mat. Estat.*, São Paulo, v.25, n.4, p.77-100, 2007.

- **ABSTRACT:** *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 maximum likelihood approach is used in order to fit the models and estimate the probabilities. Furthermore, we propose asymptotic tests to compare treatments based on odds ratio and on the differences of transition probabilities. The methods are illustrated with respiratory disease data. For these data, the process is stationary of order two and the suggested test points to a significant statistical difference in favour of the active treatment.*
- **KEYWORDS:** *Longitudinal data; generalized linear model; stochastic processes; transition probabilities; maximum likelihood.*

Referências

- AGRESTI, A. *Categorical data analysis*. New York: Jonh Wiley, 1991. 558p.
- ANDERSON, T. W.; GOODMAN, L. A. Statistical Inference about Markov Chains. *Ann. Math. Stat.*. Ann Arbor, v.28, p.89-110, 1957.
- AZZALINI, A. Maximum likelihood estimation of order n for stationary stochastic processes. *Biometrika*, Cambridge, v.70, p.381-388, 1983.
- BISHOP, Y. M. M.; FIENBERG, S. E.; HOLLAND, P. W. *Discrete multivariate analysis: theory and practice*. Massachussets: The MIT Press, 1975. 557p.
- COX, D. R. *The analysis of binary data*. London: Methuen, 1970. 142p.

- DIGGLE, P. J. et al. *Analysis of longitudinal data*. New York: Oxford University Press, 2002. 379p.
- FITZMAURICE, G. M.; LAIRD, N. M. A likelihood-based method for analysing longitudinal binary responses. *Biometrika*, Cambridge, v.80, p.141-151, 1993.
- 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. *Am. J. Sociol.*, Chicago, v.68, p.57-78, 1962.
- KOCH, G. C. et al. Categorical data analysis. In: BERRY, D.A. *Statistical methodology in the pharmaceutical sciences*. New York: Marcel Dekker, 1990. chap.13, p.389-473.
- 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.
- 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. 300p.
- LINDSEY, J. K. *Statistical analysis of stochastic processes in time*. New York: Cambridge University Press, 2004. 338p.
- MANTEL, N.; HAENSZEL, B. F. Statistical aspects of the analysis of the data from retrospective studies of disease. *J. Am. Stat. Assoc.*, Boston, v.58, p.690-700, 1959.
- McCULLAGH, P. *Regression methods for ordinal data*. J. R. Stat. Soc., Ser. B Methodol., London, v.42, p.109-142, 1980.
- MOLENBERGS, G.; VERBEKE, G. *Models for discrete longitudinal data*. New York: Springer-Verlag, 2005, 683p.
- NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. *J. R. Stat. Soc.*, Ser. A Gen., London, v.135, p.370-384, 1972.
- NOLETO, A. Q. *Análise de dados longitudinais com respostas categorizadas através de processos de Markov*. 1991. 93f. Dissertação (Mestrado em Estatística)-Universidade Estadual de Campinas, Campinas, 1991.
- PAULA, G. A. *Modelos de regressão com apoio computacional*. São Paulo: IME, 2004. 245p.
- 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 ago. 2007.
- SAS INSTITUTE INC. *SAS/STAT user's guide, Version 8*. Cary NC, 1999.
- WARE, J. H.; LIPSITZ, S.; SPEIZER, F. E. Issues in the Analysis of Repeated Categorical Outcomes. *Stat. Med.*, Chichester, v.7, p.95-107, 1988.

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.; 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 aproach. *Biometrics*, Washington, v.44, p.1019-1031, 1988.

Recebido em 14.09.2007.

Aprovado após revisão em 06.03.2008.