

UMA ABORDAGEM BAYESIANA NA ANÁLISE DE DADOS LONGITUDINAIS OBTIDOS POR AMOSTRAGEM COM REPOSIÇÃO

*Davi Casale Aragon**

*Adriana de Fátima Lourençon***

*Eliza Omai****

*Jorge Alberto Achcar*****

*Edson Zangiacomi Martinez******

Introdução

O presente estudo discute a aplicabilidade de um modelo de efeitos mistos proposto para a análise de um conjunto de dados médicos longitudinais com uma estrutura não usual. Tais dados foram coletados de 16 indivíduos, divididos em 4 grupos, submetidos a transplante de medula óssea no Hospital das Clínicas da Faculdade de Medicina da USP de Ribeirão Preto. Foram coletadas amostras de sangue dos pacientes, no instante em que foi detectado o episódio de febre e 2, 4, 6, 12 e 24 horas após esse evento. A partir dos dados clínicos e do resultado da hemocultura colhida no momento do pico febril, as amostras foram classificadas em quatro grupos: (1) HN (pacientes que não receberam substâncias pirogênicas, ou seja, causadoras de febre, e apresentaram hemocultura negativa); (2) GP (pacientes que não receberam substâncias pirogênicas e apresentaram hemocultura positiva para bactérias Gram-

* CEMEQ/FAEPA/FMRP/USP, Mestrando em Medicina Social pela Faculdade de Medicina da USP de Ribeirão Preto (FMRP-USP), dcaragon@terra.com.br.

** CEMEQ/FAEPA/FMRP/USP, Mestrando em Medicina Social pela Faculdade de Medicina da USP de Ribeirão Preto (FMRP-USP), adlourencon@yahoo.com.br.

*** CEMEQ/FAEPA/FMRP/USP, Mestre em Estatística pela Universidade Federal de São Carlos (UFSCAR), Estatística do Hospital da UNICAMP, elizaomai@hotmail.com.

**** Departamento de Medicina Social, Faculdade de Medicina de Ribeirão Preto-USP, PhD em Estatística pela Universidade de Wisconsin, EUA, professor titular do ICMC-USP, estatístico da faculdade de medicina da USP, Ribeirão Preto, achcar@fmrp.usp.br.

***** Departamento de Medicina Social, Faculdade de Medicina de Ribeirão Preto-USP, Doutor em medicina pela UNICAMP, Professor Doutor da Faculdade de Medicina da USP, Ribeirão Preto, edson@fmrp.usp.br.

positivas); (3) GN (pacientes que não receberam substâncias pirogênicas e apresentaram hemocultura positiva para bactérias Gram-negativas); (4) A (pacientes que tiveram febre devido à administração de substâncias pirogênicas, como a Anfotericina). (5) Controle (grupo controle). Em cada momento de coleta de sangue foram dosadas algumas quimiocinas (MIG, RANTES, MIP-1 alfa, MCP-1, IP-10 e Eotaxin - 2), que são substâncias liberadas pelo corpo para combater a febre. É importante lembrar que as respostas do grupo controle foram admitidas, pelo pesquisador, constantes, para cada tempo e quimiocina. Em cada momento de coleta de sangue foram dosadas algumas quimiocinas, que são substâncias liberadas pelo corpo para combater a febre.

A amostragem com reposição foi caracterizada pelo fato de um mesmo paciente, na coleta de sangue, em dois episódios diferentes de febre, poder ser classificado, num primeiro instante, no grupo GP e em outro pico febril, no GN, por exemplo. Assim, houve casos em que o indivíduo teve tantos eventos de febre que pôde ser classificado nos quatro grupos. Essa situação pode ser visualizada de maneira mais clara na Tabela 1. Esta estrutura não usual de dados motivou a aplicabilidade de um modelo de efeitos mistos para a análise dos dados. O objetivo do estudo é avaliar o padrão da cinética das quimiocinas, durante os episódios de febre, verificando possíveis diferenças entre os tempos e os agentes causadores.

Tabela 1 – Classificação dos 16 indivíduos nos grupos por números de episódios de febre.

Indivíduo	Grupos	Indivíduo	Grupos	Indivíduo	Grupos	Indivíduo	Grupos
I	2	V	1,4	IX	4	XIII	2,3
II	1,1,3	VI	1,1,2,3,4,4,4	X	4	XIV	1,1,1
III	1,1,3	VII	3	XI	1,2	XV	3,3
IV	1,1,2,4	VIII	1,3,4	XII	2	XVI	1

Metodologia

Informações sobre o uso de métodos Bayesianos na pesquisa médica podem ser obtidos em Freedman, 1996 ou Spiegelhalter *et al.*, 1999. A análise bayesiana, para um determinado modelo estatístico, consiste na obtenção de medidas resumos ou densidades a posteriori para os parâmetros do modelo de interesse. Estas medidas ou densidades são obtidas combinando informações a priori sobre os parâmetros de interesse (densidade a priori) e informações contidas na amostra (função de verossimilhança).

Neste caso, para cada paciente, têm-se 6 respostas (MIG, Rantes, MIP – 1 alfa, MCP – 1, IP – 10 e Eotaxin – 2), medidas em 6 tempos (T0, T2, T4, T6,

T12 e T24), para cada um dos 4 grupos de agentes causadores de febre (HN, GP, GN e A). Cada um dos 16 indivíduos foi considerado uma ou mais vezes no estudo, resultando numa correlação entre algumas medidas.

O modelo proposto foi:

$$y_{ijk} = \alpha + \omega_i + \tau_j + \beta_t + \lambda_k + (\beta\tau)_{ij} + (\tau\lambda)_{jk} + (\beta\lambda)_{ik} + \varepsilon_{ijk},$$

onde, y_{ijk} é a variável resposta para i -ésimo indivíduo ($i = 1, \dots, 16$), da dosagem da j -ésima quimiciona ($j = 1, \dots, 6$), no t -ésimo instante ($t = 1, \dots, 6$), na k -ésima causa de febre ($k = 1, \dots, 4$); α é o intercepto; ω_i é o efeito aleatório para cada indivíduo i ($i = 1, \dots, 16$); τ_j é o efeito da j -ésima quimiciona ($j = 1, \dots, 6$); β_t é o efeito do t -ésimo tempo ($t = 1, \dots, 6$); λ_k é o efeito da k -ésima causa de febre ($k = 1, \dots, 4$); $(\beta\tau)_{ij}$ é o efeito de interação entre tempo e quimiciona; $(\tau\lambda)_{jk}$ é o efeito de interação entre quimiciona e causa da febre; $(\beta\lambda)_{ik}$ é o efeito de interação entre tempo e causa da febre; ε_{ijk} é o erro aleatório, com distribuição $N(0; \sigma^2)$. A função de verossimilhança para θ , onde $\theta = (\alpha, \tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \beta\tau_{11}, \dots, \beta\tau_{66}, \tau\lambda_{11}, \dots, \tau\lambda_{64})$, é dada por:

$$L(\theta) \propto (\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{16} \sum_{j=1}^6 \sum_{k=1}^4 \varepsilon_{ijk}^2\right\},$$

onde $N = 1332$ observações, e

$$\varepsilon_{ijk}^2 = y_{ijk} - \alpha + \omega_i + \tau_j + \beta_t + \lambda_k + (\beta\tau)_{ij} + (\tau\lambda)_{jk} + (\beta\lambda)_{ik}, \quad i = 1, \dots, 16; j, t = 1, \dots, 6; k = 1, \dots, 4.$$

A partir de uma análise preliminar detalhada dos dados, foram escolhidas distribuições a priori não informativas para os parâmetros do modelo, que são: $\alpha \sim N(\mu_\alpha, \sigma_\alpha^2)$, $\omega \sim N(\mu_\omega, \sigma_\omega^2)$, $\tau_j \sim N(\mu_{\tau_j}, \sigma_{\tau_j}^2)$, $\beta_t \sim N(\mu_{\beta_t}, \sigma_{\beta_t}^2)$, $\lambda_k \sim N(\mu_{\lambda_k}, \sigma_{\lambda_k}^2)$, $\beta\tau_{ij} \sim N(\mu_{\beta\tau_{ij}}, \sigma_{\beta\tau_{ij}}^2)$, $\tau\lambda_{jk} \sim N(\mu_{\tau\lambda_{jk}}, \sigma_{\tau\lambda_{jk}}^2)$, $\beta\lambda_{ik} \sim N(\mu_{\beta\lambda_{ik}}, \sigma_{\beta\lambda_{ik}}^2)$, $\sigma_\omega^2 \sim IG(l, m)$, $\sigma^2 \sim IG(n, o)$, onde $\mu_\alpha, \sigma_\alpha^2, \mu_\omega, \sigma_\omega^2, \mu_{\tau_j}, \sigma_{\tau_j}^2, \mu_{\beta_t}, \sigma_{\beta_t}^2, \mu_{\lambda_k}, \sigma_{\lambda_k}^2, \mu_{\beta\tau_{ij}}, \sigma_{\beta\tau_{ij}}^2, \mu_{\tau\lambda_{jk}}, \sigma_{\tau\lambda_{jk}}^2, \mu_{\beta\lambda_{ik}}, \sigma_{\beta\lambda_{ik}}^2, l, m, n, o$ são constantes conhecidas, $N(\mu, \sigma^2)$ denota uma distribuição normal com média μ e variância σ^2 e $IG(l, m)$ denota uma distribuição gama inversa com média $l/(l-1)$ e variância $m^2/[(l-1)(l-2)]$. Assumiu-se independência, a priori, entre os parâmetros.

Para o ajuste do modelo Bayesiano utilizou-se o software Winbugs versão 1.3, (Spiegelhalter et al., 1995). Foram geradas, pelo método de simulação intensiva, 170000 amostras com saltos de 30. As primeiras 20000 foram descartadas ("burn-in") para que as inferências estivessem livres de possíveis efeitos dos valores iniciais. Além disso, foram consideradas as iterações 30ª, 60ª, 90ª, ..., resultando em 5000 amostras para cada um dos parâmetros. Foi utilizado o algoritmo Gibbs Sampling para a obtenção das amostras simuladas das distribuições a posteriori para θ . A convergência deste foi verificada a partir de gráficos temporais das amostras geradas.

Resultados e Discussão

A Figura 1, a seguir, ilustra o comportamento das quimiocinas, ao longo do tempo, para cada agente causador de febre. As estimativas dos parâmetros do modelo, exceto interações, são dadas na Tabela 2.

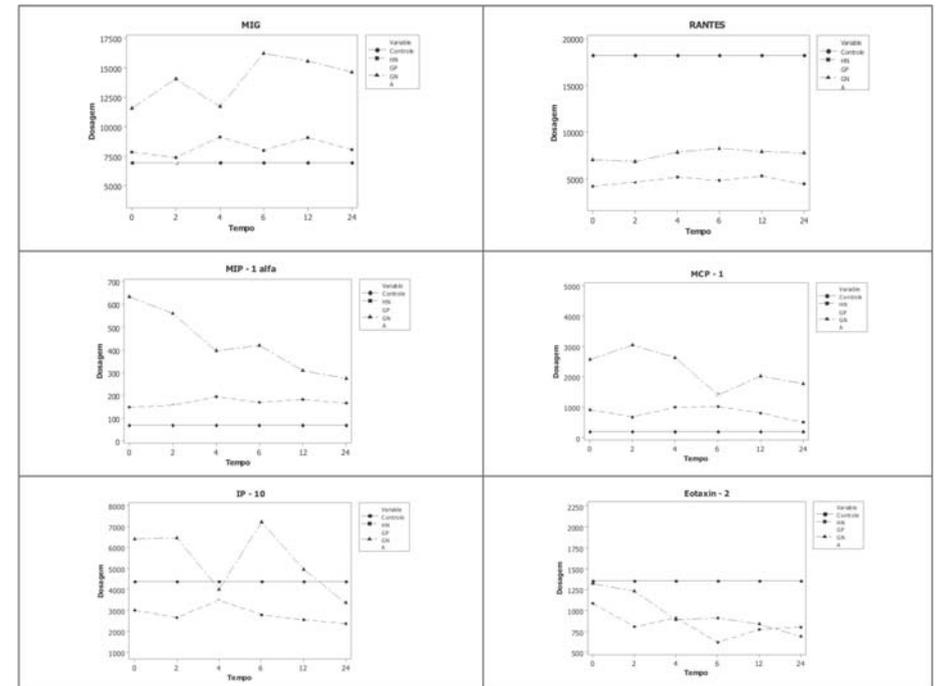


Figura 1. Comportamento das quimiocinas, no tempo, por agente causador de febre.

As curvas foram comparadas com o grupo controle através de intervalos de credibilidade para a diferença entre $\mu_{ijk} - \mu_0$, onde μ_0 é a média do controle, em cada resposta j , no tempo t na causa k e $\mu_{ijk} = E[y_{ijk}]$. Assim, alguns dos resultados observados foram:

- Na quimiciona MIG, foram detectadas possíveis diferenças entre os causadores GN e A, quando comparados com o controle, nos tempos 0, 2, 4 e 6;
- Na quimiciona RANTES, foram detectadas possíveis diferenças entre todas as quatro causas, quando comparadas com o controle, em todos os tempos;

Tabela 2. Resultados obtidos para as estimativas dos parâmetros do modelo.

Parâmetros	Estimativa	Desvio Padrão	IC (95%)	
			LI	LS
α	2,977	2,550	-2,014	7,924
τ_1	5,667	2,898	0,210	11,640
τ_2	2,133	2,867	-3,200	7,885
τ_3	-2,663	2,887	-8,316	3,002
τ_4	-1,262	2,852	-6,843	4,303
τ_5	0,125	2,816	-5,409	5,567
τ_6	-1,986	2,870	-7,490	3,701
β_1	0,091	2,852	-5,496	5,727
β_2	-0,131	2,880	-5,566	5,470
β_3	0,004	2,865	-5,672	5,432
β_4	0,059	2,914	-5,678	5,590
β_5	0,052	2,826	-5,531	5,592
β_6	-0,117	2,896	-5,867	5,483
λ_1	-0,221	2,850	-5,830	5,402
λ_2	0,065	2,870	-5,623	5,680
λ_3	1,245	2,806	-4,154	6,736
λ_4	-0,863	2,817	-6,300	4,554

- Já na quimiocina MIP 1- alfa, não foi detectada nenhuma diferença entre as causas e o grupo controle, bem como na Eotaxin – 2, em todos os momentos;

- Numa comparação entre as causas, na quimiocina MIG, só não houve diferença entre os causadores GP e GN, em todos os tempos. Já na MCP1 foram observadas diferenças entre HN e GP, em todos os instantes de coleta. Na IP – 10, foram detectadas diferenças entre GN e A em todos os tempos.

Conclusões

A abordagem Bayesiana, neste tipo de análise, mostrou-se uma poderosa ferramenta na obtenção de resultados, sendo capaz de captar algumas informações que seriam inviáveis caso fosse utilizada uma alternativa clássica.

Referências:

FREEDMAN, L. Bayesian Statistical methods. **British Medical Journal**, v. 313, p. 569-570, 1996.

GELFAND, A. E. Gibbs Sampling. **Journal of the American Statistical Association**, v. 95, n° 452, p. 1300 – 1304, 2000.

SPIEGELHALTER, D.J., MYLES, J. P., TONES, D.R., ABRAMS, K.R. An introduction to Bayesian methods in health technology assessment. **British Medical Journal**, v. 319, p.508-512, 1999.

SPIEGELHALTER, D.J., THOMAS, A., BEST, N.G., GILKS, W.R. **WinBugs version 1.3:** Bayesian inference using Gibbs sampling. MRC Biostatistics Unit, Cambridge, 1995.

Resumo:

O objetivo deste estudo é mostrar a aplicabilidade de um modelo Bayesiano, proposto para a análise de um conjunto de dados médicos longitudinais com uma estrutura não usual. Um único indivíduo pode vir a participar da pesquisa mais de uma vez e, a cada participação, dados clínicos e laboratoriais podem classificá-lo em um dentre quatro grupos de interesse. Tal experimento foi assim delineado por utilizar uma população restrita (pacientes submetidos a transplante de medula óssea). Observa-se que o modelo proposto mostrou-se uma importante ferramenta para a análise destes dados.

Palavras-chave:

Modelos de Efeitos Aleatórios, Dados Longitudinais, Amostragem com Reposição, Análise Bayesiana.