

# Assimetria e curtose em modelos não-lineares generalizados: estudos de simulação envolvendo os preditores para o teste de *Salmonella/microsossoma*

Davi Butturi-Gomes<sup>1 2</sup>

Silvio Sandoval Zocchi<sup>1</sup>

## 1 Introdução

O teste de *Salmonella/microsossoma*, conhecido também por teste de Ames (TA), surgiu no contexto da avaliação de risco, desempenhando um importante papel na identificação de substâncias capazes de induzir mutações e danos no DNA. O TA conquistou uma posição de destaque na comunidade científica e nas organizações governamentais, sendo utilizado para diversos fins, entre os quais podem ser citados a liberação de medicamentos e de defensivos agrícolas. Estes ensaios consistem, simplificada, na exposição de linhagens específicas da bactéria *Salmonella typhimurium* a diferentes doses de um produto ou de uma amostra ambiental. Após um período de incubação pré-determinado, o número de colônias revertentes em cada placa é contado, constituindo a variável resposta do experimento.

Diversos modelos estatísticos foram propostos para a análise dos dados provenientes do TA e todos são casos particulares de um modelo não-linear generalizado (MNLG). Os elementos de um MNLG, segundo [12], são: (i) um componente aleatório, cuja distribuição deve pertencer à família exponencial; (ii) um preditor tipicamente não-linear; e (iii) uma função de ligação, conectando as realizações do componente aleatório ao preditor. Esta definição permite modelos bastante flexíveis, que podem ser entendidos como uma combinação das teorias do modelo não-linear (MNL) e do modelo linear generalizado (MLG).

Considerando particularmente os dados do TA, no qual a variável resposta é constituída por contagens, as distribuições propostas para o componente aleatório são geralmente Poisson e, na presença de superdispersão, Binomial Negativa, Poisson generalizada ou Quase-Poisson. Com relação aos preditores, há uma variedade de propostas, todas seguindo as mesmas suposições probabilísticas como recomendado por [4], que incluem uma variável *offset* para representar o número inicial de bactérias em cada placa, uma função  $F$  para o acréscimo de colônias revertentes devido à mutagenicidade e um função de sobrevivência  $\bar{G}$  para o decréscimo dessas colônias devido à toxicidade. É possível observar na Tabela 1 todos os modelos propostos, de acordo com a autoria do preditor, a verossimilhança escolhida por seus autores e, por último, a função de variância para contemplar ou não a presença de superdispersão.

<sup>1</sup>LCE - ESALQ/USP. e-mail: [davibg@usp.br](mailto:davibg@usp.br)

<sup>2</sup>Agradecimento à CAPES pelo apoio financeiro.

Tabela 1: Modelos não-lineares generalizados propostos para a análise dos dados do teste de *Salmonella*/microssoma. O preditor de cada modelo é obtido pelo produto Offset  $\cdot F \cdot \bar{G}$ . A coluna  $\hat{L}$  representa a verossimilhança proposta por cada autor (BN: binomial negativa; Quase: quase-Poisson) e a coluna  $V(\mu)$  contém as funções de variância consideradas, em que  $\mu$  é a esperança da variável aleatória com a distribuição considerada,  $x$  é a dose de uma substância e  $m$  é o número inicial de bactérias na placa.

Modelo	Offset	$F$	$\bar{G}$	$\hat{L}$	$V(\mu)$
Margolin [6]	$m$	$1 - e^{-(\beta_0 + \beta_1 x)}$	$e^{-\gamma x}$	BN	$\mu(1 + \mu\kappa^{-1})$
Myers [8]	1	$\beta_0 + \beta_1 x$	$e^{-\gamma x}$	Quase	$a\mu^b$
Krewski [5]	1	$\beta_0 + \beta_1 x$	$e^{-\gamma x^\alpha}$	Quase	$a\mu^b$
Breslow [2]	1	$e^{\beta_0}(x + \delta)^{\beta_1}$	$e^{-\gamma x}$	Quase	$\mu(1 + c\mu)$
Bernstein [1]	1	$\beta_0 + \beta_1 x$	1	Quase	$a\mu^b$
Stead [10]	1	$\beta_0 + \beta_1 x^{\beta_2}$	$e^{-\gamma x}$	Poisson	$\mu$

O nível de não-linearidade de modelos pode afetar as distribuições amostrais dos estimadores dos parâmetros, de forma que testes de hipóteses baseados na normalidade assintótica podem não ser confiáveis [9]. Os modelos propostos para o TA possuem um ou dois parâmetros de interesse, aqueles diretamente ligados ao potencial mutagênico de uma substância, para os quais se deseja testar hipóteses. Uma vez que as distribuições amostrais dos estimadores dos parâmetros desses modelos em particular ainda não são completamente conhecidas, o objetivo desse trabalho foi avaliar suas respectivas assimetria e curtose.

## 2 Material e métodos

Para se avaliar a assimetria e a curtose das distribuições amostrais dos estimadores dos parâmetros dos modelos apresentados na Tabela 1, foram realizadas simulações de dados. Foram consideradas as distribuições Poisson e Binomial Negativa para o componente aleatório de cada modelo e, em todos os casos, a função de ligação logarítmica. Para representar as doses, foram utilizados 6 níveis de uma variável explanatória quantitativa com três repetições cada, no esquema proposto em [7], o protocolo internacional para o TA. Os parâmetros utilizados para tais simulações estão disponíveis na Tabela 2. A escolha desses valores específicos para os parâmetros se deu devido aos dados geralmente observados na prática, particularmente para a linhagem TA100 de *S. typhimurium* [7].

O objetivo em realizar simulações a partir da distribuição Binomial Negativa consiste em contemplar a possível presença de superdispersão: à medida que um  $\kappa$  fixo aumenta, melhor é aproximação da Binomial Negativa para a Poisson. Neste trabalho, foram considerados 3 valores para  $\kappa$ : 100, 50 e 25, portanto, as simulações em que  $\kappa = 25$  são as em que a superdispersão é mais evidente. Note que não há interesse em se estimar  $\kappa$ , apenas não impor a restrição de que, na distribuição Poisson, o parâmetro de dispersão seja 1. Para tanto, os modelos da Tabela 1 foram ajustados aos dados simulados, por máxima verossimilhança e máxima quase-verossimilhança respectivamente, para os componentes aleatórios, Poisson e quase-Poisson,

Tabela 2: Valores dos parâmetros utilizados para os estudo de simulação para cada modelo.

Modelo	Número de parâmetros	$\theta$			
		$\beta_0$	$\beta_1$	$\beta_2$	$\gamma$
Bernestein	2	4,7	0,4	-	-
Breslow	3	4,7	2	-	0,5
Krewski	3	4,7	1,2	-	0,03
Margolin	3	0,295	0,3	-	0,25
Myers	3	4,7	2	-	0,2
Stead	4	4,7	2	1,5	0,3

utilizando o algoritmo disponível em [11].

Assim, cada combinação (Preditor  $\times$  Componente aleatório  $\times \kappa$ ) foi repetida 1000 vezes ( $N = 1000$ ). Seguindo as recomendações de [9], foram reportados os coeficientes de assimetria e curtose, respectivamente:  $g_1 = \frac{m_3}{\sqrt{m_2^3}}$  e  $g_2 = \frac{m_4}{\sqrt{m_2^2}} - 3$ , em que  $m_t = \frac{\sum_i^N (\hat{\theta}_i - \tilde{\theta})^t}{N}$  é o  $t$ -ésimo momento amostral;  $\theta$  é o parâmetro utilizado para realizar as simulações (Tabela 2);  $\hat{\theta}_i$  é a estimativa desse parâmetro na  $i$ -ésima simulação;  $\tilde{\theta}$  é a média dessas estimativas em todas as  $N$  simulações. Tendo essas medidas em mãos, foram realizados testes de hipóteses para verificar a normalidade assintótica dos estimadores, considerando que assintoticamente  $g_1 \sim N(0, 6/N)$  e que  $g_2 \sim N(0, 24/N)$ , para  $N$  suficientemente grande.

### 3 Resultados e discussões

Os resultados obtidos nas simulações de cada cenário para os estimadores dos parâmetros diretamente associados ao potencial mutagênico de substâncias estão apresentados na Tabela 3.

Considerando o modelo de Bernestein, um MLG, não ocorreu afastamento da normalidade assintótica para nenhum dos estimadores de máxima ou de quase-máxima verossimilhança. As propriedades assintóticas dos estimadores em MLGs, entretanto, já foram amplamente estudadas [3] e, neste trabalho, servem como ponto de partida para a comparação com os demais modelos em estudo.

Para os modelos de Breslow e de Krewski, apenas no cenário de maior superdispersão ( $\kappa = 25$ ) houve um caso em que o afastamento da simetria normal foi significativo, com valor de  $p < 0,05$ . Este caso se refere ao estimador de quase-máxima verossimilhança do parâmetro de intercepto do modelo ( $\beta_0$ ), para o qual geralmente não há interesse em se realizar teste de hipótese, uma vez que os interceptos nos modelos para o TA são interpretados como o número de revertentes espontâneos, que já bastante conhecidos na literatura da área.

Os modelos de Margolin e de Myers apresentaram casos de assimetria significativa (valor de  $p < 0,05$ ) na distribuição amostral dos estimadores de quase-máxima verossimilhança do intercepto, respectivamente, nos cenários de menor superdispersão ( $\kappa = 100$ ) e superdispersão intermediária ( $\kappa = 50$ ). Além disso, a distribuição do estimador de  $\beta_1$  do modelo de Margolin

no cenário de maior superdispersão apresentou assimetria significativa (valor de  $p < 0,05$ ), o que pode causar falha do teste de hipótese.

Com relação ao modelo de Stead, o único com 4 parâmetros, sendo dois de interesse ( $\beta_1$  e  $\beta_2$ ), foram observados diversos casos de afastamento da simetria normal. Além daqueles destacados na Tabela 3, também não se apresentaram simétricas as distribuições amostrais dos estimadores de quase-máxima verossimilhança do parâmetro  $\gamma$ , em todos os cenários de superdispersão.

Por fim, as distribuições amostrais de todos os estimadores, em quaisquer modelo e cenário estudados, apresentaram curtose normal e, pelo menos para os estimadores de  $\beta_1$  as curvas tendem a ser simétricas. Estas boas propriedades permitem afirmar que os testes de hipóteses para os parâmetros de interesse são confiáveis, embora alguma atenção é necessária na construção de intervalos de confiança conjuntos, uma vez que há um padrão assimétrico na distribuição amostral dos interceptos dos modelos.

Tabela 3: Coeficientes de assimetria ( $g_1$ ) e curtose ( $g_2$ ) obtidos para as distribuições amostrais dos estimadores do parâmetros diretamente associados ao potencial mutagênico. O parâmetro  $\kappa$  é inversamente proporcional à superdispersão, de forma que, para  $\kappa \rightarrow \infty$ , a verossimilhança foi Poisson e, para os demais valores, quase-Poisson.

Modelo	$\theta$	$\kappa$							
		$\infty$		100		50		25	
		$g_1$	$g_2$	$g_1$	$g_2$	$g_1$	$g_2$	$g_1$	$g_2$
Bernstein	$\beta_1$	0,0306	-0,0226	0,0420	-0,2372	-0,1073	-0,1585	0,0430	-0,1498
Breslow	$\beta_1$	-0,0408	0,1589	0,0346	-0,0154	0,0909	-0,2347	0,0809	0,2318
Krewski	$\beta_1$	0,0756	-0,1959	-0,0010	-0,1420	-0,0422	-0,0935	0,0020	0,1310
Margolin	$\beta_1$	-0,0012	-0,0442	0,0369	-0,1148	-0,0071	0,2000	0,1611*	-0,1162
Myers	$\beta_1$	0,0830	0,0789	0,0923	0,1158	0,0712	0,0276	-0,0080	0,0506
Stead	$\beta_1$	-0,0203	-0,2733	-0,1433	0,1259	-0,2963**	-0,1967	-0,3290**	-0,0856
	$\beta_2$	-0,1415	0,0291	-0,1410	0,1288	-0,2497**	0,1008	-0,1336	0,1177

\*valor de  $p < 0,05$ ; \*\*valor de  $p < 0,01$

## 4 Conclusões

Ao menos quando a verossimilhança e o preditor estão corretamente especificados, os estimadores do parâmetro de interesse ( $\beta_1$ ) dos modelos propostos para o TA com até três parâmetros apresentam curvas simétricas e mesocúrticas, de forma que os testes de Wald ou similares podem ser empregados. Já para o modelo de Stead, que possui 4 parâmetros, caso seja observada superdispersão, recomenda-se cautela no uso desses testes, uma vez que as distribuições amostrais dos estimadores podem não ser simétricas.

É necessário que, para complementar o presente estudo, sejam ainda realizados os testes de hipóteses para verificar se os estimadores são viesados ou se possuem excesso de variância, considerando, é claro, suas distribuições amostrais. Além disso, pode ser contemplada uma

gama maior de valores para os parâmetros das simulações, desde que sejam respeitados os processos biológicos.

## 5 Referências

- [1] BERNSTEIN, L.; KALDOR, J.; MACCANN, J.; PIKE, M. C. An empirical approach to the statistical analysis of mutagenesis data from the *Salmonella* test. **Mutation Research**, v. 97, p. 267–281, 1982.
- [2] BRESLOW, N. E. Extra-poisson variation in log-linear models. **Journal of the Royal Statistical Society – Series C (Applied Statistics)**, v. 33, n. 1, p. 38–44, 1984.
- [3] CORDEIRO, H. H.; CORDEIRO, G. M. Skewness for parameters in generalized linear models. **Communications in Statistics – Theory and Methods**, v. 30, n. 7, p. 1317–1334, 2001.
- [4] ECKARDT, F.; HAYNES, R. H. Kinetics of mutation induction by ultraviolet in excision-deficient yeast. **Genetics**, v. 85, p. 225–247, 1977.
- [5] KREWSKI, D.; LEROUX, B. G.; BLEUR, S. R.; BROEKHAVEN, L. H. Modelling the Ames *Salmonella*/microsome assay. **Biometrics**, v. 49, p. 499–510, 1993.
- [6] MARGOLIN, B. H.; KAPLAN, N.; ZEIGER, E. Statistical analysis of the Ames *Salmonella*/microsome test. **Proceedings of the National Academy of Sciences of the United States of America**, v. 78, n. 6, p. 3779–3783, 1981.
- [7] MORTELMANS, K.; ZEIGER, E. The Ames *Salmonella*/microsome mutagenicity assay. **Mutation Research**, v. 455, p. 29–60, 2000.
- [8] MYERS, L. E.; SAXTON, N. H.; SOUTHERLAND, L. I.; WOLFF, T. J. Regression analysis of Ames test data. **Environ. Mol. Mutagen.**, v. 3, p. 575–586, 1981.
- [9] RATKOWSKY, D. A. **Nonlinear regression modelling – a unified practical approach**. Marcel Dekker Inc., 1983.
- [10] STEAD, A. G.; HASSELBLAD, V.; CREASON, J. P.; CLAXTON, L. Modelling the Ames test. **Mutation Research**, v. 85, p. 13–27, 1981.
- [11] TURNER, H. L.; FIRTH, D. gnm, a package for generalized nonlinear models. **Rnews**, v. 7, n. 2, p. 8–12, 2007.
- [12] TURNER, H. L.; FIRTH, D. Tools for computing generalized nonlinear models in R. **Statistical Computing and Graphics**, v. 18, n. 1, p. 11–16, 2007.