ESTATÍSTICA GRADIENTE: PROPRIEDADES E APLICAÇÕES

Michel Helcias MONTORIL[1]
Estéfano Alves de SOUZA[1]

▪ RESUMO: Neste trabalho exibimos de forma simples e intuitiva, com ilustrações gráficas, a formulação da estatística gradiente, com base no raciocínio de Buse (1982), também comentando em resumo a construção das estatísticas da razão de verossimilhanças de Wald e escore. No decorrer do texto mencionamos algumas propriedades da estatística gradiente, ilustrando-as com exemplos simples e comparando, sempre que possível, às outras três estatísticas.

▪ PALAVRAS-CHAVE: Estatística gradiente; estatística da razão de verossimilhanças; estatística de Wald; estatística escore; distribuição Qui-quadrado.

1 Introdução

Seja $(\mathcal{X}, \mathcal{F})$ um espaço mensurável tal que $\mathcal{X}$ corresponde ao espaço amostral de uma certo experimento $\boldsymbol{X}$ e $\mathcal{F}$ a uma σ-álgebra de subconjuntos de $\mathcal{X}$. Associando a esse espaço mensurável uma família de medidas de probabilidades $\mathcal{P}$, temos o modelo estatístico $(\mathcal{X}, \mathcal{F}, \mathcal{P})$. É muito comum supor que a família de medidas $\mathcal{P}$ seja indexada por uma quantidade θ, que pode ser escalar ou um vetor, e na prática é desconhecida (e de interesse). Desse modo, pode-se escrever $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$, em que Θ representa o espaço paramétrico, i.e., o conjunto de todos os possíveis valores de θ. Além disso, faz-se necessário supor correspondência biunívoca entre θ e $\mathcal{P}$, a fim de garantir a identificabilidade do modelo. Muitas vezes há o interesse em inferir acerca da distribuição de um conjunto de dados observados, o que é feito com respeito a θ, que define, neste caso, a distribuição P_θ dos dados. Comumente, tal interesse se dá em avaliar o quão plausível é a hipótese de que o parâmetro da distribuição geradora dos dados pertença a um subconjunto do espaço paramétrico.

Suponha que estejamos interessados em testar a hipótese $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, para algum $\boldsymbol{\theta}_0$ pertencente ao espaço paramétrico. Na literatura, há diversas formas para se fazer isso,

[1]Universidade de São Paulo – USP, Instituto de Matemática e Estatística – IME, Departamento de Estatística, CEP: 05508-090, São Paulo, SP, Brasil. E-mail: *michelcias@gmail.com* / *estefano@ime.usp.br*

dentre as quais destacamos três testes que já são bastante conhecidos e divulgados, sendo eles: o teste da razão de verossimilhanças, proposto originalmente por Neyman e Pearson (1928), e discutido com mais detalhes por Wilks (1938); o teste de Wald, proposto por Wald (1943); e o teste escore, proposto por Rao (1948). Essas estatísticas podem ser escritas na forma

$$\begin{aligned} \mathrm{LR} &= 2(l(\hat{\boldsymbol{\theta}}) - l(\tilde{\boldsymbol{\theta}})), \\ \mathrm{S} &= \left[\mathbf{U}(\tilde{\boldsymbol{\theta}})\right]^{\top} \left[\mathcal{I}(\tilde{\boldsymbol{\theta}})\right]^{-1} \mathbf{U}(\tilde{\boldsymbol{\theta}}), \\ \mathrm{W} &= (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})^{\top} \mathcal{I}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}), \end{aligned}$$

em que $l(\boldsymbol{\theta})$ corresponde à função de log-verossimilhança dos dados avaliada em $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$ é o estimador de máxima verossimilhança (EMV) de $\boldsymbol{\theta} \in \Theta$ e $\mathrm{U}(\boldsymbol{\theta}) = \partial l(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ corresponde à função escore. Ainda, sob condições adequadas (veja, por exemplo, Lehmann e Casella, 2003), temos que

$$\mathcal{I}(\boldsymbol{\theta}) = \mathrm{E}\left[-\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\top}} l(\boldsymbol{\theta})\right],$$

que corresponde à matriz de informação de Fisher.

Em problemas complexos, obter (ou inverter) a matriz de informação geralmente é uma tarefa complicada. Situações em que os dados apresentam censura costumam dificultar o cálculo da matriz de informação de Fisher, fazendo com que seja mais viável o uso da matriz de informação observada para as estatísticas de Wald e escore. Ainda, de acordo com Terrell (2002), apesar de parecer relativamente simples, lidar com a log-verossimilhança em algumas situações pode vir a ser mais complicado do que trabalhar com a função escore, como no caso dos modelos lineares generalizados de postos (*generalized linear models for ranks*). Em decorrência desses problemas, daremos ênfase neste trabalho à estatística gradiente (F), proposta originalmente por Terrell (2002). Segundo Rao (2005), seria interessante um estudo mais detalhado a respeito de tal estatística, a qual pode ser escrita na forma

$$\mathrm{F} = \left[\mathbf{U}(\tilde{\boldsymbol{\theta}})\right]^{\top} (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}).$$

Todas estas estatísticas (F, LR, S, W) são assintoticamente equivalentes, ou seja, todas elas possuem a mesma distribuição assintótica, sendo esta uma χ^2_q, em que o número de graus de liberdade da distribuição qui-quadrado, q, é igual ao número de restrições impostas sob a hipótese nula, H_0. Mais detalhes a respeito podem ser obtidos em Sen e Singer (1994) e Terrell (2002).

Recentemente, alguns estudos têm sido dedicados a estatística gradiente, especialmente no que diz respeito ao poder local sob alternativas de Pitman, que são sequências de hipóteses alternativas que convergem para a hipótese nula com taxa $n^{-1/2}$, em que n corresponde ao tamanho amostral. Junto às estatísticas da razão de verossimilhanças, de Wald e escore, o poder local da estatística gradiente foi estudado por Lemonte e Ferrari

(2012a), onde foi verificado que nenhum dos quatro testes é uniformemente mais poderoso. Comparações dessas quatro estatísticas são feitas em modelos paramétricos específicos, como é o caso de modelos não lineares da família exponencial (Lemonte, 2011), modelos de dispersão (Lemonte e Ferrari, 2012b), modelos lineares generalizados com dispersão variável (Lemonte, 2012), entre outros. Mais recentemente, Lemonte (2013) propôs uma estatística gradiente robusta para má especificação de modelos, além de derivar sua distribuição assintótica sob a hipótese nula, e Vargas et. al. (2013) obtiveram a expansão assintótica da estatística gradiente sob a hipótese nula, para testar hipóteses compostas na presença de parâmetros de perturbação.

Essa variedade de artigos que vêm sendo desenvolvidos ultimamente nos motiva na produção deste trabalho. De cunho fundamentalmente didático, temos como principal objetivo familiarizar o leitor à estatística gradiente, de modo a deixá-lo ciente da existência dessa outra interessante possibilidade de teste, bem como incentivar seu uso.

No decorrer deste trabalho, destacamos algumas características da estatística gradiente ilustrando com exemplos simples e, sempre que possível, comparando com as outras três estatísticas supracitadas. Na Seção 2, seguindo o raciocínio de Buse (1982), exibimos uma forma simples e intuitiva da construção das estatísticas da razão de verossimilhanças, de Wald e escore, para então justificar a formulação da estatística gradiente. Na Seção 3, mencionamos algumas das propriedades relacionadas a estatística gradiente, sempre que conveniente, comparando às demais. Estudos de simulação são apresentados na Seção 4. Na Seção 5, aplicamos os testes da razão de verossimilhanças, de Wald, escore e gradiente a um conjunto de dados reais, utilizando um modelo de regressão logística. Na Seção 5, encerramos o trabalho com alguns comentários.

2 Comparações gráficas para hipóteses simples

Com base no trabalho de Buse (1982), exibimos brevemente uma interpretação simples e intuitiva, via métodos gráficos, para as estatísticas da razão de verossimilhanças, de Wald e escore. Em seguida fazemos o mesmo com a estatística gradiente.

Suponha que o parâmetro de interesse seja um escalar e queiramos testar $H_0 : \theta = \theta_0$ contra a hipótese alternativa $H_a : \theta \neq \theta_0$. Analisando inicialmente a estatística LR, que neste caso é escrita como $\mathrm{LR} = 2(l(\hat{\theta}) - l(\theta_0))$, podemos observar na Figura 1 que a distância entre as log-verossimilhanças, $(1/2)\mathrm{LR}$, depende tanto da distância $\hat{\theta} - \theta_0$ quanto da curvatura $J(\hat{\theta})$ da função, a qual pode ser escrita da forma

$$J(\theta) = -\frac{d^2}{d\theta^2} l(\theta).$$

Em outras palavras, para uma curvatura fixada, quanto maior a distância entre $\hat{\theta}$ e θ_0, maior será o valor de LR. Ainda, para uma distância entre $\hat{\theta}$ e θ_0 fixada, quanto maior o valor da curvatura $J(\hat{\theta})$, maior o valor da estatística LR.

Com base na estatística LR, poder-se-ia pensar em avaliar evidências contra a hipótese nula apenas através de $\hat{\theta} - \theta_0$. Contudo, como o sinal da diferença não é de interesse, uma alternativa razoável seria fazer uso da distância quadrática $(\hat{\theta} - \theta_0)^2$, pois valores elevados desta distância poderiam servir como indícios de que H_0 não seria

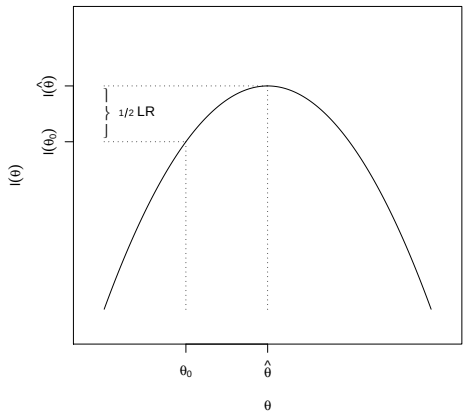

Figura 1 - Razão de verossimilhanças.

verdadeira, visto que, assim como observado na estatística da razão de verossimilhanças, à medida que θ_0 se distancia de $\hat{\theta}$, a distância entre as log-verossimilhanças $l(\theta_0)$ e $l(\hat{\theta})$ também cresce, aumentando o valor de LR. Todavia, como pode ser observado na Figura 2, dois conjuntos de dados, com um deles menos favorável à hipótese nula, podem fornecer a mesma distância quadrática, daí a necessidade da ponderação pela curvatura avaliada em $\hat{\theta}$, pois os dados que fornecem uma curvatura maior em $\hat{\theta}$, apresentam uma log-verossimilhança menor, quando avaliada em θ_0, o que implica numa maior distância entre $l(\hat{\theta})$ e $l(\theta_0)$, justificando a forma da estatística de Wald, que corresponde à distância quadrática, $(\hat{\theta} - \theta_0)^2$, ponderada pela curvatura da função de log-verossimilhança em $\hat{\theta}$, ou seja, $\mathrm{W} = (\hat{\theta} - \theta_0)^2 J(\hat{\theta})$. No entanto, o mais usual é que a distância quadrática seja ponderada pela curvatura média – informação de Fisher (ou informação esperada) – avaliada em $\hat{\theta}$, que aqui é denotada por $\mathcal{I}(\hat{\theta})$. Note que

$$\mathcal{I}(\theta) \;=\; \mathrm{E}\left(J(\theta)\right),$$

e pelo fato de $J(\theta)$ ser um estimador consistente de $\mathcal{I}(\theta)$, as duas versões da estatística de Wald são assintoticamente equivalentes (Sen e Singer, 1994, p. 236).

Para introduzirmos a estatística escore, note que, se θ_0 estiver próximo de $\hat{\theta}$, a função escore avaliada em θ_0, $\mathrm{U}(\theta_0)$, estará aproxima de zero, visto que $\mathrm{U}(\hat{\theta}) = 0$. Em outras palavras, o coeficiente angular da reta tangente à função de log-verossimilhança no ponto θ_0 deve estar próximo de zero. Em decorrência disso, pode-se utilizar a função escore para medir a distância entre $\hat{\theta}$ e θ_0. Como o sinal da função escore não é de interesse, uma alternativa para avaliar tal distância poderia ser $[\mathrm{U}(\theta_0)]^2$. Contudo, como pode ser observado na Figura 3, seria possível que dois conjuntos de dados apresentassem o mesmo valor para $\mathrm{U}(\theta_0)$, embora a distância entre $\hat{\theta}$ e θ_0 fosse maior no caso em que a log-verossimilhança apresentasse menor curvatura em θ_0. Além disso, quanto maior essa

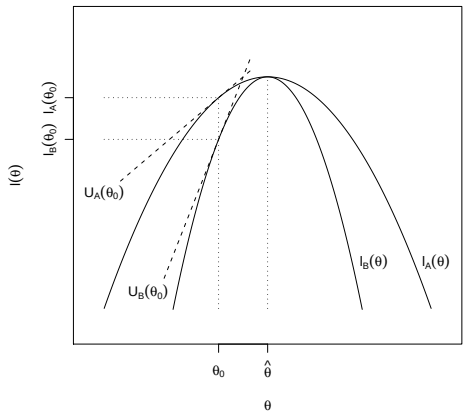

Figura 2 - Wald e gradiente.

distância, maior a distância entre $l(\theta_0)$ e $l(\hat{\theta})$. Uma forma de contornar essa situação seria poderando $[\mathrm{U}(\theta_0)]^2$ pelo inverso da curvatura da função de log-verossimilhança em θ_0, que resultaria na estatística escore, $\mathrm{S} = \dfrac{[U(\theta_0)]^2}{J(\theta_0)}$. Assim como na estatística de Wald, existem duas versões para a estatística escore, sendo uma delas ponderada pelo inverso da informação observada, e a outra, ponderada pelo inverso da informação de Fisher, ambas avaliadas em θ_0, e ambas assintoticamente equivalentes.

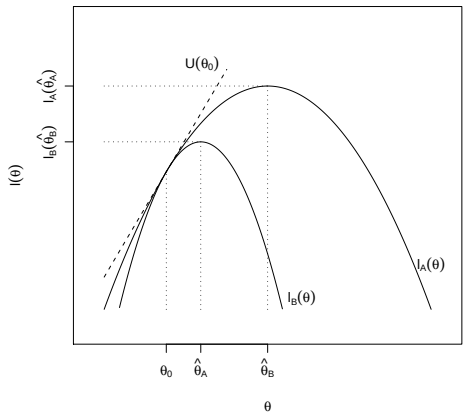

Figura 3 - Escore e gradiente.

No caso em que foi discutida a formulação da estatística de Wald, poder-se-ia inicialmente levar em consideração apenas a diferença $\hat{\theta} - \theta_0$, ao invés da distância quadrática. Como pode ser observado na Figura 2, uma forma de contornar o problema do sinal dessa diferença, e ao mesmo tempo levar em conta a curvatura da log-verossimilhança em $\hat{\theta}$, seria poderar $(\hat{\theta} - \theta_0)$ por $\mathrm{U}(\theta_0)$. Note que o sinal da diferença é o mesmo sinal da função escore avaliada em θ_0. Também, quanto maior a curvatura $J(\hat{\theta})$, maior $\mathrm{U}(\theta_0)$ em valor absoluto. Por outro lado, no caso da formulação da estatística escore, se o interesse fosse levar em consideração apenas $\mathrm{U}(\theta_0)$, ao invés de seu quadrado, pode ser observado na Figura 3 que uma alternativa seria a ponderação pela diferença $\hat{\theta} - \theta_0$, a qual levaria em conta a curvatura da função de log-verossimilhança em θ_0, assim como a estatística escore, e ainda reverteria o problema do sinal de $\mathrm{U}(\theta_0)$. Os dois casos levam à estatística gradiente $\mathrm{F} = U(\theta_0)(\hat{\theta} - \theta_0)$.

3 Propriedades da estatística gradiente

A partir de agora mencionamos algumas das propriedades da estatística gradiente. Ao invés de demonstrações, utilizamos exemplos simples, que auxiliarão na compreensão do que queremos transmitir.

Não entraremos em detalhes quanto às suposições necessárias com relação à estatística gradiente neste trabalho. Todavia, gostaríamos de ressaltar que em todos os exemplos utilizados para ilustrar as propriedades mencionadas aqui satisfazem tais suposições. Mais detalhes sobre essas suposições e resultados podem ser encontrados em Terrell (2002).

3.1 Distribuição exata e assintótica

Assim como as estatísticas da razão de verossimilhanças, de Wald e escore, a estatística gradiente tem distribuição exata qui-quadrado quando os dados são oriundos de distribuição normal.

Exemplo 1: (Buse, 1982) Considere $X_1, X_2, \ldots, X_n$ uma amostra aleatória oriunda de uma distribuição normal de média μ (pertencente aos reais), e variância 1. Neste caso, a log-verossimilhança corresponde a

$$l(\mu) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(X_i - \mu)^2,$$

uma função quadrática em μ, da qual obtemos a função escore

$$\mathrm{U}(\mu) = \frac{d}{d\mu}l(\mu) = \sum_{i=1}^{n}(X_i - \mu),$$

de onde obtemos que o EMV de μ, $\hat{\mu}$, corresponde a $\bar{X} = \sum_{i=1}^{n} X_i/n$. Logo, temos que a estatística gradiente será

$$\mathrm{F} = \mathrm{U}(\mu)(\hat{\mu} - \mu) = n(\bar{X} - \mu)^2 = \left(\frac{\bar{X} - \mu}{1/\sqrt{n}}\right)^2 = Z^2,$$

em que Z tem distribuição normal padrão. Assim, F possui distribuição χ^2_1. Além disso, como pode ser verificado em Buse (1982), neste caso, as estatísticas coincidem, isto é, $\mathrm{F} = \mathrm{LR} = \mathrm{W} = \mathrm{S}$.

É importante lembrar que, sob suposições adequadas, a estatística gradiente converge em distribuição para uma qui-quadrado, quando o conjunto de dados em questão não for oriundo de distribuição normal (Terrell, 2002). Um exemplo interessante que pode ilustrar isso segue logo abaixo.

Exemplo 2:(Terrell, 2002) Considere um experimento de contagem com p categorias independentes, em que o número de contagens na i-ésima categoria, X_i, tem distribuição de Poisson com média θ_i, $i = 1, 2, \ldots, p$. Assim, a log-verossmilhança dos dados, $l(\boldsymbol{\theta})$, a menos de uma constante aditiva que não depende de $\boldsymbol{\theta}$, corresponde a

$$\sum_{i=1}^{p} X_i \log \theta_i - \sum_{i=1}^{p} \theta_i,$$

da qual obtemos a função escore

$$U(\boldsymbol{\theta}) = \left(\frac{X_1}{\theta_1} - 1, \ldots, \frac{X_p}{\theta_p} - 1\right)^\top.$$

O EMV de θ_i é $\hat{\theta}_i = X_i$, $i = 1, \ldots, p$. Com isso, a estatística gradiente pode ser escrita na forma

$$\mathrm{F} = [U(\theta)]^\top(\hat{\theta} - \theta) = \sum_{i=1}^{p} \frac{(X_i - \theta_i)^2}{\theta_i},$$

a qual corresponde à conhecida estatística χ^2 de Pearson, que converge em distribuição para χ^2_p.

3.2 Não invariância da estatística gradiente

Muitas vezes, uma transformação dos parâmetros pode simplificar a estimação sem afetar o valor resultante da estatística. Contudo, assim como a estatística de Wald, a estatística gradiente não é invariante para transformações não lineares dos parâmetros, ao contrário das estatísticas da razão de verossimilhanças e escore.

Utilizamos um exemplo simples, baseado na distribuição exponencial, para ilustrar a não invariância da estatística gradiente.

Exemplo 3: Considere $X_1, X_2, \ldots, X_n$, uma amostra aleatória com distribuição exponencial de média θ. Suponha inicialmente que estejamos interessados em testar $H_0 : \theta = \theta_0$. A função de log-verossimilhança dos dados corresponde a

$$l(\theta) = -n\log\theta - \frac{n}{\theta}\bar{X},$$

da qual obtemos a função escore

$$\mathrm{U}(\theta) = \frac{n}{\theta^2}(\bar{X} - \theta),$$

que garante $\bar{X}$ como EMV de θ. Assim, a estatística gradiente para testar se $\theta = \theta_0$ será

$$\mathrm{F}_{(1)} = n(\bar{X} - \theta_0)^2.$$

De um modo geral, poderíamos ter interesse em testar $H_0 : \theta^k = \theta_0^k$, para algum k diferente de zero. Com base na função de log-verossimilhança, podemos obter a função escore para θ^k, que corresponde a

$$\mathrm{U}(\theta^k) = \frac{n}{k\theta^{k+1}}(\bar{X} - \theta).$$

Pelo princípio da invariância, o EMV de θ^k corresponde a $\bar{X}^k$ e, portanto, seque que a estatística gradiente nesse caso será

$$\mathrm{F}_{(k)} = \frac{n}{k\theta_0^{k+1}}(\bar{X} - \theta_0)(\bar{X}^k - \theta_0^k).$$

Desse modo, é possível observar que $\mathrm{F}_{(1)} = \mathrm{F}_{(k)}$ apenas quando $k = 1$. No geral, para valores $k \neq k^{'}$, não é difícil verificar que $\mathrm{F}_{(k)} \neq \mathrm{F}_{(k^{'})}$, o que mostra a não invariância da estatística gradiente.

A título de informação, para testar $\theta^k = \theta_0^k$, a estatística de Wald resultante será

$$n\Big(\frac{\bar{X}^k - \theta_0^k}{k\bar{X}^k}\Big)^2,$$

e não é difícil verificar que a mesma também é não invariante.

Aqui foi utilizado o EMV para o cálculo da estatística gradiente, mas Terrell (2002) mostra que F terá distribuição assintótica qui-quadrado para qualquer estimador consistente que for utilizado.

3.3 Não viciosidade

Da Seção 3.1, sabemos que nem sempre a estatística gradiente terá distribuição qui-quadrado, embora sua distribuição assintótica sempre o será (sob condições adequadas). Isto quer dizer que nem sempre a média da estatística F será a média de sua distribuição assintótica. Contudo, considere $\bar{\theta}$ como sendo o estimador (consistente) utilizado para a obtenção da estatística gradiente. Se tal estimador for não viciado, então $\mathrm{E}(\mathrm{F}) = q$, em que o número q de restrições sob a hipótese nula corresponde aos graus de liberdade da distribuição assintótica χ^2. De acordo com Terrell (2002), isso sugere que é possível obter uma melhor aproximação da estatística gradiente pela qui-quadrado, usando estimadores

de θ com menor viés em valor absoluto. Isso implicaria que, utilizando o fato de que a estatística não é invariante sob transformações do parâmetro, seria possível escolher uma reparametrização em que o EMV fosse não viciado (o uso dos EMV no cálculo será discutido na Seção 3.4).

Ao verificar a demonstração no artigo supracitado, o leitor pode vir a ter dúvidas com relação ao valor esperado da estatística gradiente, se o mesmo corresponde ao número de restrições sob a hipótese nula ou ao tamanho do vetor paramétrico. Como já mencionamos, tal valor corresponde ao número de restrições ao espaço paramétrico sob H_0. Isto pode ser verificado no exemplo que segue.

Exemplo 4: (Buse, 1982) Considere o modelo de regressão linear múltipla

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \\ \boldsymbol{u} &\sim N(\mathbf{0}; \boldsymbol{\Sigma}), \end{aligned}$$

em que $\boldsymbol{y}$ é um vetor aleatório $n \times 1$, $\boldsymbol{X}$ é uma matriz de posto completo $n \times k$ $(k < n)$, $\boldsymbol{\beta}$ um vetor de parâmetros $k \times 1$ e $\boldsymbol{\Sigma}$ uma matriz de covariâncias conhecida de ordem $n \times n$ positiva definida. Suponha que as hipóteses nula e alternativa são, respectivamente,

$$\begin{aligned} H_0 &: \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}, \\ H_a &: \boldsymbol{R}\boldsymbol{\beta} \neq \boldsymbol{r}, \end{aligned}$$

em que $\boldsymbol{R}$ é uma matriz $s \times k$ $(s < k)$ e $\boldsymbol{r}$ é um vetor $s \times 1$, representando as restrições lineares para os parâmetros.

Como a matriz de covariâncias é conhecida, a função de log-verossimilhança, a menos de uma constante aditiva que não depende de $\boldsymbol{\beta}$, é igual a

$$-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}),$$

de onde obtemos a função escore

$$\mathrm{U}(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta}) = \boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{y} - \boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{\beta}. \tag{1}$$

Os estimadores de máxima verossimilhança geral e sob H_0, respectivamente, são

$$\hat{\boldsymbol{\beta}} = \boldsymbol{C}^{-1}\boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{y} \tag{2}$$

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - \boldsymbol{C}^{-1}\boldsymbol{R}^\top (\boldsymbol{R}\boldsymbol{C}^{-1}\boldsymbol{R}^\top)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}), \tag{3}$$

em que $\boldsymbol{C} = (\boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}$.

Desse modo, por (1), (2) e (3), temos que a estatística gradiente será

$$\begin{aligned} \mathrm{F} &= \left[\mathrm{U}(\tilde{\boldsymbol{\beta}})\right]^\top (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) \\ &= (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top (\boldsymbol{R}\boldsymbol{C}^{-1}\boldsymbol{R}^\top)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}). \end{aligned} \tag{4}$$

Todos os resultados utilizados referentes ao processo de estimação neste exemplo podem ser encontrados em Buse (1982), onde também pode ser verificado que as estatísticas coincidem, isto é, $\mathrm{F} = \mathrm{LR} = \mathrm{W} = \mathrm{S}$. Consequentemente, se H_0 for verdadeira, as estatísticas possuem distribuição assintótica χ^2_s, em que s corresponde ao número de restrições da hipótese nula.

Além disso, a média e a variância de $\hat{\boldsymbol{\beta}}$ correspondem, respectivamente, a

$$\begin{aligned} \mathrm{E}(\hat{\boldsymbol{\beta}}) &= \mathrm{E}\left[\boldsymbol{C}^{-1}\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{y}\right] = \boldsymbol{C}^{-1}\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{\beta} \\ &= \boldsymbol{C}^{-1}\boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{\beta} \end{aligned} \tag{5}$$

e

$$\begin{aligned} \mathrm{Var}(\hat{\boldsymbol{\beta}}) &= \boldsymbol{C}^{-1}\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\mathrm{Var}(\boldsymbol{y})\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{C}^{-1} \\ &= \boldsymbol{C}^{-1}\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{C}^{-1} = \boldsymbol{C}^{-1}\boldsymbol{C}\boldsymbol{C}^{-1} \\ &= \boldsymbol{C}^{-1}. \end{aligned} \tag{6}$$

Por (5), podemos observar que $\hat{\boldsymbol{\beta}}$ é um estimador não viesado de $\boldsymbol{\beta}$. Além disso, com base em (4) e (6), sob H_0, temos que

$$\begin{aligned} \mathrm{E}(\mathrm{F}) &= \mathrm{E}\left[(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top(\boldsymbol{R}\boldsymbol{C}^{-1}\boldsymbol{R}^\top)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})\right] \\ &= \mathrm{E}\left[\mathrm{tr}\left\{(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top(\boldsymbol{R}\boldsymbol{C}^{-1}\boldsymbol{R}^\top)^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})\right\}\right] \\ &= \mathrm{E}\left[\mathrm{tr}\left\{(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top(\boldsymbol{R}\boldsymbol{C}^{-1}\boldsymbol{R}^\top)^{-1}\right\}\right] \\ &= \mathrm{tr}\left[\mathrm{E}\left\{(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top\right\}(\boldsymbol{R}\boldsymbol{C}^{-1}\boldsymbol{R}^\top)^{-1}\right] \\ &= \mathrm{tr}\left[\mathrm{Var}(\boldsymbol{R}\hat{\boldsymbol{\beta}})(\boldsymbol{R}\boldsymbol{C}^{-1}\boldsymbol{R}^\top)^{-1}\right] = \mathrm{tr}\left[\boldsymbol{R}\boldsymbol{C}^{-1}\boldsymbol{R}^\top(\boldsymbol{R}\boldsymbol{C}^{-1}\boldsymbol{R}^\top)^{-1}\right] \\ &= \mathrm{tr}\left[\boldsymbol{I}_{\boldsymbol{s}}\right] = s. \end{aligned}$$

3.4 EMV e a estatística gradiente

Terrell (2002) mostra que, para $\hat{\boldsymbol{\theta}}$ estimador de máxima verossimilhança, a estatística gradiente é sempre não negativa. Todavia, nenhuma garantia com relação ao sinal da estatística é dada para o caso em que seja empregado um outro estimador, apesar de no mesmo trabalho ser demonstrado que, para qualquer estimador consistente de $\boldsymbol{\theta}$, a estatística gradiente converge em distribuição para uma Qui-quadrado.

Considere o caso em que θ é um escalar. Seja $l(\theta),\ \theta \in \Theta$, a função de log-verossimilhança dos dados. Na Figura 4 é possível observar que, se o interesse for testar a hipótese $H_0 : \theta = \theta_0$, o estimador $\mathring{\theta}$ fará com que F seja negativo, visto que, neste caso, $\mathrm{U}(\theta_0) < 0$ e $\mathring{\theta} - \theta_0 > 0$.

Ainda, é interessante observar que, nesse caso, $l(\mathring{\theta}) < l(\theta_0)$. Isso indica que o mesmo problema pode acontecer com a estatística LR, caso seja usado um EMV corrigido pelo

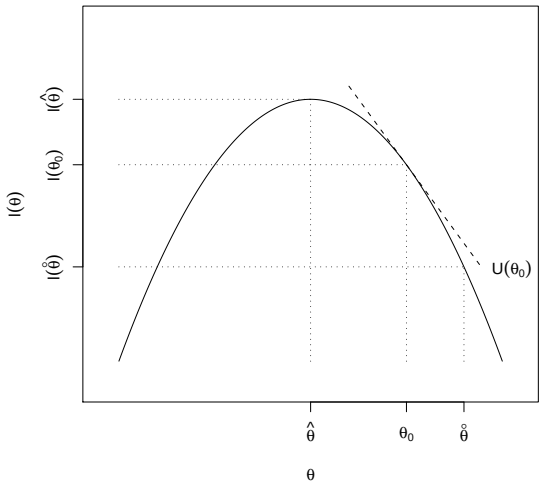

Figura 4 - Gradiente negativo

viés. Por outro lado, podemos observar na Figura 5 que a estatística gradiente será positiva sempre que a distância entre $\hat{\theta}$ e $\bar{\theta}$ for menor do que a distância entre $\hat{\theta}$ e θ_0. O mesmo pode ser observado se levarmos em consideração a log-verossimilhança, satisfazendo $l(\theta_0) < l(\bar{\theta}) \leq l(\hat{\theta})$.

Assim como a estatística gradiente, a estatística escore também pode resultar em valores negativos. Morgan et. al. (2007) mostram isso através de uma aplicação simples.

4 Simulações

Nesta seção, realizamos dois estudos de simulação, a fim de ilustrar através de estudos numéricos o desempenho da estatística gradiente, além de compará-lo com as outras estatísticas.

Exemplo 5 (dados com censura): No primeiro estudo, consideramos a distribuição Weibull, cuja função densidade de probabilidade corresponde a

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\exp\left\{-\left(\frac{x}{\beta}\right)^{\alpha}\right\}, x > 0,$$

em que $\alpha > 0$ e $\beta > 0$. Neste estudo, simulamos 10000 amostras de tamanhos $n = 20, 40, 60, 80$ e 100, com parâmetros $\alpha = 2$ e $\beta = 1$. Aqui testamos a hipótese H_0 : $\alpha = 2$ e $\beta = 1$, para cada uma das réplicas. O objetivo é comparar as taxas de erro tipo I (empíricas) das quatro estatísticas para dados com censura, utilizando um nível de

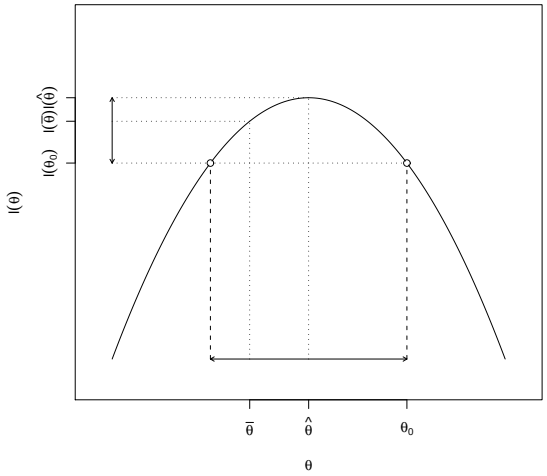

Figura 5 - Gradiente positivo

significância de 0,05. Consideramos aqui amostras com censura do tipo II à direita, com a porcentagem de censura (pc) de 10%, 20%, 30%, 40% e 50% do tamanho amostral. Os resultados encontram-se na Tabela 1, em que podemos observar as melhores taxas de erro tipo I sendo obtidas pelas estatísticas da razão de verossimilhanças e gradiente. Note que, para porcentagens de censura menores, a estatística gradiente apresenta os melhores resultados. Por outro lado, para porcentagens de censura maiores, a estatística da razão de verossimilhanças apresenta as taxas mais próximas do verdadeiro nível de significância.

Embora não tenhamos apresentado aqui, também realizamos estudos de simulação considerando outros valores de α (no caso, 1/2, 1, 3 e 5) e mantendo $\beta = 1$. Em geral, observamos que as estatísticas da razão de verossimilhanças e gradiente apresentaram as melhores taxas de erro tipo I.

Exemplo 6 (MLG): Além do estudo de simulação com dados apresentando censura, também simulamos um MLG, utilizando distribuição gama na variável resposta e função de ligação logarítmica. Em tal estudo, utilizamos o preditor linear $\eta = 2 + 3x$, em que a covariável x foi gerada de uma distribuição exponencial de média 0,5. Os tamanhos amostrais utilizados foram $n = 20, 40, 60, 80$ e 100. Foram geradas 5000 réplicas da variável resposta para cada tamanho amostral e para cada parâmetro de dispersão utilizado, $\phi = 3/2, 3$ e 5, o qual foi considerado desconhecido no processo de estimação. Na Tabela 2, encontram-se as taxas de erro tipo I (empíricas) do teste de hipótese relacionado ao coeficiente angular, em que avaliamos $H_0 : \beta = 3$ contra $H_a : \beta \neq 0$. Podemos observar, no geral, que as melhores taxas obtidas foram das estatísticas escore e gradiente. No caso, o diferencial da estatística gradiente é que ela é mais simples de ser calculada do que a estatística escore, o que a torna mais vantajosa.

Tabela 1 - Taxas de erro tipo I (empíricas) para testar $H_0 : \alpha = 2$ e $\beta = 1$, ao nível de significância de 0,05

pc	Estatística	$n = 20$	$n = 40$	$n = 60$	$n = 80$	$n = 100$
10 %	LR	0,0468	0,0471	0,0484	0,0531	0,0513
	S	0,0929	0,0755	0,0657	0,0629	0,0595
	F	0,0513	0,0507	0,0486	0,0544	0,0509
	W	0,0557	0,0535	0,0522	0,0550	0,0523
20 %	LR	0,0470	0,0478	0,0470	0,0538	0,0519
	S	0,1021	0,0825	0,0687	0,0683	0,0646
	F	0,0531	0,0537	0,0487	0,0544	0,0528
	W	0,0579	0,0538	0,0503	0,0539	0,0536
30 %	LR	0,0479	0,0476	0,0485	0,0524	0,0513
	S	0,1099	0,0908	0,0757	0,0695	0,0673
	F	0,0570	0,0516	0,0506	0,0514	0,0524
	W	0,0648	0,0569	0,0541	0,0555	0,0520
40 %	LR	0,0501	0,0480	0,0485	0,0500	0,0532
	S	0,1322	0,0985	0,0882	0,0777	0,0750
	F	0,0624	0,0551	0,0531	0,0518	0,0548
	W	0,0734	0,0592	0,0556	0,0549	0,0550
50 %	LR	0,0480	0,0474	0,0485	0,0516	0,0507
	S	0,1536	0,1142	0,0991	0,0895	0,0821
	F	0,0661	0,0575	0,0545	0,0544	0,0540
	W	0,0826	0,0685	0,0617	0,0610	0,0586

Tabela 2 - Taxas de erro tipo I (empíricas) para o teste $H_0 : \beta = 3$, ao nível de significância de $0,05$

Dispersão	Estatística	n=20	n=40	n=60	n=80	n=100
$\phi = 1,5$	LR	0,0638	0,0526	0,0510	0,0544	0,0562
	S	0,0446	0,0432	0,0474	0,0508	0,0486
	F	0,0494	0,0442	0,0498	0,0508	0,0526
	W	0,0910	0,0736	0,0604	0,0652	0,0596
$\phi = 3$	LR	0,0718	0,0582	0,0578	0,0574	0,0530
	S	0,0532	0,0498	0,0536	0,0554	0,0496
	F	0,0570	0,0514	0,0536	0,0544	0,0504
	W	0,0992	0,0654	0,0666	0,0634	0,0566
$\phi = 5$	LR	0,0676	0,0614	0,0546	0,0504	0,0506
	S	0,0528	0,0538	0,0496	0,0464	0,0484
	F	0,0544	0,0550	0,0504	0,0468	0,0480
	W	0,0822	0,0730	0,0580	0,0534	0,0556

Neste exemplo, além das taxas de erro tipo I, também avaliamos o poder dos quatro testes. Fixamos o tamanho amostral $n = 100$ e o nível de significância de 0,05.

Consideramos os casos $\phi = 3$ e $\phi = 5$ para o parâmetro de dispersão. Foram geradas 5000 réplicas sob a hipótese alternativa do coeficiente angular, em que $H_a : \beta = \beta_a$, para diferentes valores de β_a entre 2,25 e 3,75. Para cada réplica, testamos a hipótese nula $H_0 : \beta = 3$, e estimamos o poder do teste com base na taxa de rejeição empírica das réplicas geradas. O resultado pode ser observado na Figura 6. Note que, nos dois gráficos, nenhum dos testes apresentou-se uniformemente mais poderoso, concordando, assim, com o que já havia sido observado por Lemonte e Ferrari (2012a).

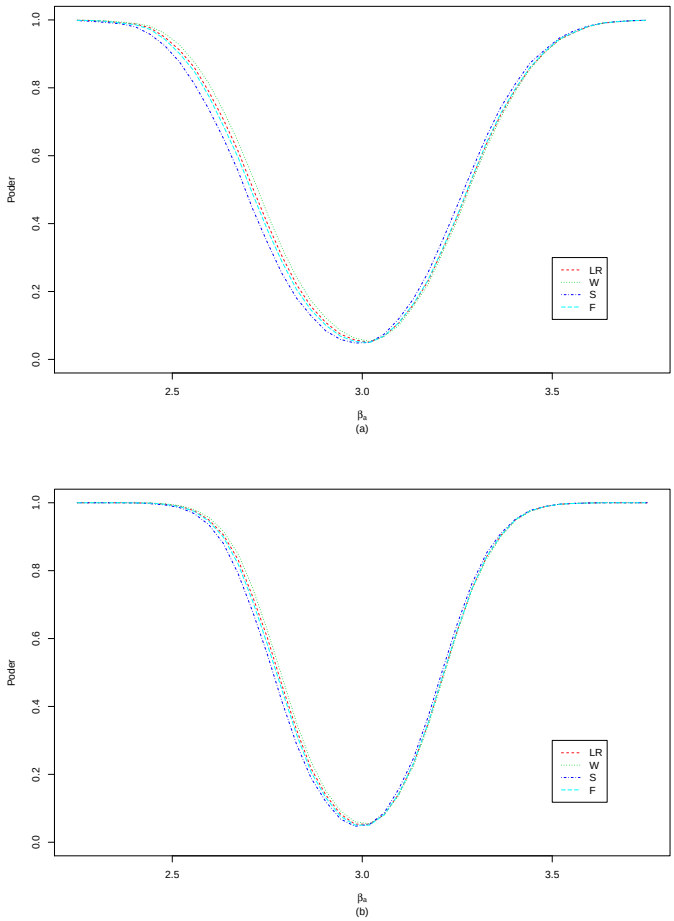

Figura 6 - Poder do teste ao nível de significância de 0,05: (a) com parâmetro de dispersão $\phi = 3$; (b) com parâmetro de dispersão $\phi = 5$.

De um modo geral, os dois exemplos indicam nas simulações que, para dados com censura, as estatísticas da razão de verossimilhanças e gradiente tendem a apresentar as melhores taxas de erro tipo I, enquanto que, para um MLG com resposta gama, as melhores

taxas são apresentadas pelas estatísticas escore e gradiente. Além disso, observamos no caso do MLG que nenhum dos quatro testes é uniformemente mais poderoso. Os resultados dos estudos indicam, também, que a estatística gradiente pode ser considera uma forte concorrente, frente às outras três estatísticas utilizadas.

5 Aplicação a dados reais

Nesta seção, ilustramos por meio de um conjunto de dados reais o uso da estatística gradiente, comparando-a com as estatísticas da razão de verossimilhanças, de Wald e escore. Os dados são referentes a um estudo realizado no Setor de Anatomia e Patologia do Hospital Heliópolis, em São Paulo, entre 1970 e 1982 (Paula e Tuder, 1986). O objetivo do estudo está relacionado ao processo infeccioso pulmonar, que pode ser considerado maligno ou benigno, em uma amostra de 175 pacientes. No caso, o interesse está em explicar a probabilidade de processo infeccioso maligno, de acordo com a idade (em anos), o sexo (masculino ou feminino), a intensidade da célula histiócitos–linfócitos (ausente, discreta, moderada ou intensa) e a intensidade da célula fibrose–frouxa (ausente, discreta, moderada ou intensa).

Visto que nosso objetivo é meramente ilustrativo, consideraremos neste trabalho apenas a covariável idade, a fim de avaliar se a mesma influencia na probabilidade de processo infeccioso maligno, que será denotado por p. Uma análise detalhada desse conjunto de dados é feita em Paula (2013, Cap. 1 e 3). Utilizamos um modelo de regressão logística e, nesse caso, temos que a probabilidade de processo infeccioso maligno para o i-ésimo indivíduo corresponde a

$$p_i = p_i(\alpha, \beta) = \frac{\exp(\alpha + \beta\ \text{Idade}_i)}{1 + \exp(\alpha + \beta\ \text{Idade}_i)}.$$

O interesse aqui é testar se $\beta = 0$.

Assim, denotemos por Y_i a variável indicadora que é igual a um, se o tumor apresentado pelo i-ésimo indivíduo for maligno, ou zero, caso contrário. Logo, temos que

$$Y_i \sim \text{Bernoulli}(p_i).$$

Assumindo independência no processo infeccioso entre os indivíduos, temos que Y_i, $i = 1, \ldots, 175$, será uma amostra de variáveis aleatórias independentes. Nesse caso, denotando $\boldsymbol{x}_i = (1,\ \text{Idade}_i)^\top$ e $\boldsymbol{\beta} = (\alpha, \beta)^\top$, a função de log-verossimilhança será

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{175} Y_i \boldsymbol{x}_i^\top \boldsymbol{\beta} - \sum_{i=1}^{175} \log(1 + \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})).$$

A função escore é obtida a partir da log-verossimilhança apresentada, resultando em

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{175} (Y_i - p_i)\boldsymbol{x}_i.$$

Assim, a estatística gradiente fica escrita da forma

$$\mathrm{F} = \sum_{i=1}^{175} (Y_i - p_i(\boldsymbol{\beta}_0)) \boldsymbol{x}_i^\top (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0),$$

em que $p_i(\boldsymbol{\beta}_0) = p_i(\alpha_0, \beta_0)$, e $\hat{\boldsymbol{\beta}}$ e $\boldsymbol{\beta}_0$ correspondem aos estimadores de máxima verossimilhança geral e sob H_0, respectivamente, para o vetor $\boldsymbol{\beta}$.

No caso do conjunto de dados apresentado, para testar se $\beta = 0$, obtivemos $\hat{\boldsymbol{\beta}} = (-3,752, 0,068)^\top$ e $\boldsymbol{\beta}_0 = (-0,382; 0)^\top$. Com as estimativas obtidas, calculamos o valor da estatística gradiente, $\mathrm{F} = 50,987$. A título de informação, as estatísticas da razão de verossimilhanças, de Wald e escore correspondem a $\mathrm{LR} = 45,419$, $\mathrm{W} = 40,892$ e $\mathrm{S} = 32,711$, respectivamente. Todas as estatísticas têm distribuição assintótica χ^2_1 e são bem significativas, indicando que a idade influencia na probabilidade de processo infeccioso maligno.

Conclusões e comentários

Neste trabalho, exploramos o uso da estatística gradiente, motivando sua formulação e destacando algumas de suas propriedades com o auxílio de exemplos simples. Verificamos em algumas situações que o valor da estatística gradiente é o mesmo das estatísticas da razão de verossimilhanças, de Wald e escore.

Destacamos ainda que, embora seja relativamente simples de se calcular, quando comparadas às estatísticas de Wald e escore, que precisam da matriz de informação, a estatística gradiente possui algumas características, tais como a possibilidade de vir a ser negativa (o que nunca acontece quando utilizamos estimadores de máxima verossimilhança no cálculo da mesma) e sua não invariância sob reparametrizações, embora este último possa ser usado como vantagem, e que estas mesmas características também podem ser verificadas na estatística de Wald (não invariância) e nas estatísticas da razão de verossimilhanças e escore (valores negativos).

Estudos de simulação apresentaram bons indicativos de que a estatística gradiente pode ser tratada como um forte competidor frente às outras três estatísticas mencionadas, devido a seus interessantes resultados numéricos. Em um estudo de simulação para dados com censura, a estatística gradiente apresentou as melhores taxas de erro tipo I, junto à estatística da razão de verossimilhanças. Em outro estudo, utilizando MLG, a estatística gradiente também se destacou, dessa vez junto com a estatística escore, apresentando as melhores taxas de erro tipo I. Ainda no estudo de simulação do MLG, observamos que nenhum dos quatro testes é uniformemente mais poderoso, realçando ainda mais a competitividade da estatística gradiente junto às outras três. Entendemos, assim, que o potencial da estatística gradiente fica evidenciado, haja visto seu bom desempenho nos dois estudos.

Utilizando dados reais, ilustramos o uso das quatro estatísticas por meio de um modelo de regressão logística, para testar se a variável Idade influencia na probabilidade do processo infeccioso maligno de pacientes tratados no Setor de Anatomia e Patologia

do Hospital Heliópolis, em São Paulo. Todas as quatro estatísticas foram fortemente significativas.

Agradecimentos

O primeiro autor agradece ao CNPq e à Fapesp, pelo apoio financeiro. O segundo agradece ao CNPq. Os autores gostariam de agradecer aos pareceristas pelas valiosas sugestões, as quais contribuíram substancialmente para a melhora deste artigo.

MONTORIL, M. H.; SOUZA, E.A. Gradient statistic: properties and applications. *Rev. Bras. Biom.,* São Paulo, v.31, n.1, p.43-60, 2013.

- *ABSTRACT: In this work we show in a simple and intuitive way, with graphical illustrations, the gradient statistic, based on Buse (1982), and we also comment about the formulation of the likelihood ratio, Wald and score statistics. Some properties of the gradient statistic are presented, making use of simple examples and comparisons with the other three statistics mentioned.*

- *KEYWORDS: Gradient statistic; likelihood ratio statistic; Wald statistic; score statistic; chi-squared distribution.*

Referências

BUSE, A. The likelihood ratio, Wald, and Lagrange multiplier tests: an expository note. *Am. Stat.*, Washington, v. 36, n. 3, p. 153-157, 1982.

LEHMANN, E. L.; CASELLA, G. *Theory of point Estimation*. New York: Springer, 2003. 589p.

LEMONTE, A. Local power of some tests in exponential family nonlinear models. *J. Stat. Plann. Infer.*, Amsterdam, v. 141, n. 5, p.1981-1989, 2011.

LEMONTE, A. J. Nonnull asymptotic distributions of the LR, Wald, score and gradient statistics in generalized linear models with dispersion covariates. *Statistics*, Abingdon, v.47, n.6, p.1-17, 2012.

LEMONTE, A. J. On the gradient statistic under model misspecification. *Stat. Probab. Lett.*, Amsterdam, v. 83, n. 1, p.390-398, 2013.

LEMONTE, A.; FERRARI, S. The local power of the gradient test. *Ann. Inst. Stat. Math.*, Heidelberg, v. 64, n. 2, p.373-381, 2012a.

LEMONTE, A. J.; FERRARI, S. L. P. Local power and size properties of the LR, wald, score and gradient tests in dispersion models. *Stat. Methodol.*, Amsterdam, v. 9, n. 5, p.537-554, 2012b.

LEMONTE, A. J.; FERRARI, S. L. P. A note on the local power of the LR, wald, score and gradient tests. *Electron. J. Stat.*, Beachwood, v. 6, p.421-434, 2012c.

MORGAN, B.; PALMER, K.; RIDOUT, M. Negative score test statistic. *Am. Stat.*, Washington, v. 61, p.285-288, 2007.

NEYMAN, J.; PEARSON, E. S. On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, Oxford, v. 20a, n. 1/2, p.175-240, 1928.

PAULA, G. A.; TUDER, R. M. Utilização da regressão logística para aperfeiçoar o diagnóstico de processo infeccioso pulmonar. *Ciên. Cult.*, São Paulo, v.40, p.1046-1050, 1986.

PAULA, G. A. *Modelos de regressão*: com apoio computacional. São Paulo: IME/USP, [2013]. Disponível em: <http://www.ime.usp.br/ giapaula/texto2013.pdf>. Acesso em: 30 jun. 2013.

RAO, C. R. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Math. Proc. Camb. Philos. Soc.*, Cambridge, v. 44, p.50-57, 1948.

RAO, C. R. Score test: Historical review and recent developments. In: BALAKRISHNAN, N.; NAGARAJA, H.N.; KANNAN, N. (Ed.). *Advances in ranking and selection, multiple comparisons, and reliability*. Boston: Birkhäuser, 2005. p.3-20. (Statistics for Industry and Technology).

SEN, P. K.; SINGER, J. M. *Large sample methods in statistics*: an introduction with applications. New York: Chapman & Hall/CRC, 1994. 382p.

TERRELL, G. R. The gradient statistic. *Comput. Sci. Stat.*, Kingston, v. 34, p.206-215, 2002.

VARGAS, T. M., FERRARI, S. L. P.; LEMONTE, A. J. Gradient statistic: higher-order asymptotics and bartlett-type correction. *Electron. J. Stat.*, Beachwood, v. 7, p.43-61, 2013.

WALD, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Am. Math. Soc.*, v. 54, n. 3, p.426-482, 1943.

WILKS, S. S. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Stat.*, v. 9, n. 1, p.60-62, 1938.

Recebido em 26.04.2013.

Aprovado após revisão em 11.09.2013.