**Correções de Bartlett
e Tipo-Bartlett em
Modelos Lineares Generalizados**

Audrey Helen Mariz de Aquino Cygneiros

Dissertação apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo
para
obtenção do grau
de
Mestre em Estatística

Área de Concentração: Estatística

Orientadora: **Prof^ª Dr^ª Silvia Lopes de Paula Ferrari**

São Paulo - Junho - 1997

**Correção de Bartlett
e Tipo-Bartlett em
Modelos Lineares Generalizados**

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por **Audrey Helen Mariz de Aquino Cysneiros** e aprovada pela Comissão Julgadora.

São Paulo, 23 de junho de 1997

Banca Examinadora:

- Prof^ª Dr^ª Silvia Lopes de Paula Ferrari (Orientadora) IME/USP
- Prof^ª Dr^ª Mônica Carneiro Sandoval IME/USP
- Prof. Dr. Enrico Colosimo UFMG

*Aos meus pais, avô e tia,
com gratidão.*

*Ao meu marido Francisco,
com carinho.*

*Ao meu filho Rafael,
com muito amor.*

Agradecimentos

- À professora **Silvia Ferrari**, pela confiança e excelente orientação dedicada na elaboração desta dissertação.
- Ao meu marido **Francisco**, pelo grande apoio a mim concedido e, em especial, ao meu filho **Rafael**, pela compreensão e carinho por ele oferecido.
- Aos meus pais **Lauricy** e **Washington**, que me forneceram princípios básicos e fundamentais para minha formação moral e dedicação integral ao meu objetivo.
- A minhas irmãs **Mary** e **Walma**, tia **Cleide**, vovô **Hermano**, sogra **Gilvanete** e sogro **Gilberto** (in memoriam), pelo carinho e confiança em mim depositados.
- Aos meus amigos, que me apoiaram e ajudaram permitindo que este passo da minha vida fosse dado.
- Aos colegas do Depto. de Matemática e Estatística da Universidade Federal da Paraíba (campus II), pelo apoio dado durante este curso e, em especial, aos amigos e professores da Área de Estatística.
- A **Alba Teixeira**, **Cláudia Lima**, **Érika Fukunaga**, **Francisco Souza**, **Gil de Luna**, **Josenita de Luna**, **Loretta Gasco**, **Manoel Sena**, **Miguel Opazo**, **Rosana Marques** e **Thelma Sáfiadi**, que sempre me incentivaram nesta caminhada.
- Ao Laboratório de Computação Científica Avançada (**LCCA**) e, em especial, a **Rosemary Nowak**, **Rosa Perez** e **Francisco Ribacionka**.
- À **CAPES** pelo apoio financeiro.

Resumo

Esta dissertação tem dois objetivos. O primeiro é reunir resultados importantes sobre correções de Bartlett e tipo-Bartlett em modelos lineares generalizados para testes da razão de verossimilhança e escore e apresentar algumas aplicações a dados reais. O segundo é apresentar resultados de simulação para averiguar o efeito das correções em modelos lineares generalizados contínuos e discretos.

Abstract

This dissertassion has two purposes. The first purpose is to describe important results on Bartlett and Bartlett-type corrections to the likelihood ratio and score tests in generalized linear models and to present some applications to real data. The second purpose is to present simulation results in order to evaluate the effect of the corrections in continuous and discrete generalized linear models.

Conteúdo

Lista de Figuras	vii
Lista de Tabelas	ix
Introdução	1
1 Modelo linear generalizado	3
1.1 Introdução	3
1.2 Definição	3
1.3 Estimação dos parâmetros de regressão e de dispersão	8
1.4 Testes da razão de verossimilhança e escore em MLGs	11
2 Correções de Bartlett e tipo-Bartlett	14
2.1 Introdução	14
2.2 Correção de Bartlett para a estatística da razão de verossimilhança	16
2.3 Correção tipo-Bartlett para a estatística escore	22
2.4 Alguns casos especiais	27
2.5 Modelo de regressão linear simples	29
2.6 Modelo de classificação a um fator	33
2.7 Modelo de regressão com dois parâmetros	39
3 Simulações em modelos contínuos	43
4 Simulações em modelos discretos	62
Conclusões	75
Referências	77

Lista de Figuras

3.1	Tamanhos dos testes - $n=30$ - modelo normal - nível nominal de 1%	48
3.2	Tamanhos dos testes - $n=30$ - modelo normal - nível nominal de 5%	49
3.3	Tamanhos dos testes - $n=30$ - modelo normal - nível nominal de 10%	50
3.4	Tamanhos dos testes - $n=30$ - modelo gama com índice $\phi=2,0$ - nível nominal de 1%	56
3.5	Tamanhos dos testes - $n=30$ - modelo gama com índice $\phi=2,0$ - nível nominal de 5%	57
3.6	Tamanhos dos testes - $n=30$ - modelo gama com índice $\phi=2,0$ - nível nominal de 10%	58
3.7	Tamanhos dos testes - $n=30$ - modelo gama com índice $\phi=1,0$ - nível nominal de 1%	59
3.8	Tamanhos dos testes - $n=30$ - modelo gama com índice $\phi=1,0$ - nível nominal de 5%	60
3.9	Tamanhos dos testes - $n=30$ - modelo gama com índice $\phi=1,0$ - nível nominal de 10%	61
4.1	Tamanhos dos testes - $n=20$ - modelo de Poisson - nível nominal de 1%	69
4.2	Tamanhos dos testes - $n=20$ - modelo de Poisson - nível nominal de 5%	70
4.3	Tamanhos dos testes - $n=20$ - modelo de Poisson - nível nominal de 10%	71

LISTA DE FIGURAS

viii

4.4	Tamanhos dos testes - $n=30$ - modelo binomial com $m=5$ - nível nominal de 1%	72
4.5	Tamanhos dos testes - $n=30$ - modelo binomial com $m=5$ - nível nominal de 5%	73
4.6	Tamanhos dos testes - $n=30$ - modelo binomial com $m=5$ - nível nominal de 10%	74

Lista de Tabelas

1.1	Características de algumas distribuições da família exponencial	5
2.1	Modelos lineares - funções w , g , j , h e b	28
2.2	Modelos canônicos - funções w , f , j , h e b	29
2.3	Tempo de sobrevivência em semanas e $\log_{10}$ do número de glóbulos brancos para 17 pacientes com leucemia	32
2.4	Quantidades $A_{1,\beta}$, $A_{2,\beta}$, $A_{3,\beta}$ e $\epsilon_{p,\beta}$ para um modelo de classificação a um fator com 2 populações	36
2.5	Tempo até a perda de velocidade de 5 tipos de motores	37
2.6	Informações sobre taxa, licença, renda e estrada obtidas de 40 estados norte-americanos	41
3.1	Tamanhos dos testes - $p=10$ - modelo normal	46
3.2	Tamanhos dos testes - $n=30$ - modelo normal	47
3.3	Tamanhos dos testes - $p=10$ - modelo gama com índice $\phi=2,0$	52
3.4	Tamanhos dos testes - $n=30$ - modelo gama com índice $\phi=2,0$	53
3.5	Tamanhos dos testes - $p=10$ - modelo gama com índice $\phi=1,0$	54
3.6	Tamanhos dos testes - $n=30$ - modelo gama com índice $\phi=1,0$	55
4.1	Tamanhos dos testes - $p=10$ - modelo de Poisson	65
4.2	Tamanhos dos testes - $n=20$ - modelo de Poisson	66
4.3	Tamanhos dos testes - $p=10$ - modelo binomial com $m=5$	67
4.4	Tamanhos dos testes - $n=30$ - modelo binomial com $m=5$	68

Introdução

Nelder e Wedderburn (1972) definiram os modelos lineares generalizados (MLGs) que são uma extensão dos modelos normais lineares. A idéia básica consiste em abrir o leque de opções para a distribuição da variável resposta, permitindo que a mesma pertença à família exponencial de distribuições, bem como dar maior flexibilidade para a relação funcional entre a média da variável resposta e o preditor linear η . Nelder e Wedderburn também propuseram um processo iterativo para a estimação dos parâmetros e introduziram o conceito de desvio ('deviance') que tem sido largamente utilizado na avaliação da qualidade do ajuste dos MLGs. Inúmeros artigos relacionados com os MLGs foram publicados desde 1972. Procuramos, no Capítulo 1, revisar os principais resultados teóricos relacionados com os MLGs. Em particular, mostramos como ficam os testes da razão de verossimilhança e escore nesta classe de modelos.

As correções de Bartlett e tipo-Bartlett são utilizadas para melhorar a qualidade das aproximações das estatísticas da razão de verossimilhança e escore pela distribuição qui-quadrado. Para melhorar a aproximação por qui-quadrado para a distribuição da estatística da razão de verossimilhança, pode-se multiplicar a estatística por um fator de correção, conhecido como correção de Bartlett, que não depende do valor da estatística. Desta forma, é obtida uma estatística modificada tendo distribuição qui-quadrado até ordem n^{-1} sob a hipótese nula, onde n é o tamanho da amostra. Entretanto, existe uma correção do tipo-Bartlett para a estatística escore que é dada por um polinômio na própria estatística. Assim, é obtida uma estatística escore modificada tendo distribuição qui-quadrado até ordem n^{-1} sob a hipótese nula. No Capítulo 2, revisamos os principais artigos que tratam da obtenção das correções nos MLGs. As fórmulas das correções são dadas em notação

matricial e podem ser implementadas em sistemas de computação algébrica para se obterem expressões em forma fechada quando aplicadas a modelos especiais. Ilustramos o uso das fórmulas através de exemplos numéricos.

Nos Capítulos 3 e 4, apresentamos resultados de simulação⁽¹⁾ para avaliar o efeito das correções para testes em MLGs contínuos e discretos e analisamos a influência do número de parâmetros de perturbação no desempenho dos testes.

⁽¹⁾ Todas as simulações apresentadas nesta dissertação foram desenvolvidas a partir de programas construídos na linguagem *FORTRAN* utilizando as sub-rotinas do IMSL e processados no computador CRAY do Laboratório de Computação Científica Avançada - LCCA - da Universidade de São Paulo.

CAPÍTULO 1

Modelo linear generalizado

1.1 Introdução

Os Modelos Lineares Generalizados (MLGs), definidos por Nelder e Wedderburn (1972), podem ser interpretados como uma generalização do modelo tradicional de regressão linear. Estes modelos são definidos por um conjunto de variáveis aleatórias independentes tendo distribuição da família exponencial, mais ampla que a normal, e uma estrutura linear nos parâmetros relacionada à média da variável resposta através de uma função de ligação. A ligação entre a média e o preditor linear não é necessariamente a identidade, podendo assumir qualquer forma monótona não-linear. O processo iterativo para a estimação dos parâmetros do preditor linear pode ser visto como um método de mínimos quadrados ponderados. Dois testes de hipóteses estatísticas em MLGs são apresentados: o da razão de verossimilhança e o escore. Estudos abrangentes sobre MLGs são encontrados nos livros de McCullagh e Nelder (1989), Cordeiro (1986) e Dobson (1990).

1.2 Definição

Consideremos n variáveis aleatórias independentes $y_1, \dots, y_n$, cada uma com função densidade (ou de probabilidade) na família exponencial da forma

$$f(y; \theta_i, \phi) = \exp\{\phi[y\theta_i - b(\theta_i) + c(y)] + a(y, \phi)\}, \quad i = 1, \dots, n \quad (1.1)$$

onde $a(\cdot, \cdot)$, $b(\cdot)$ e $c(\cdot)$ são funções conhecidas. Assumimos, por enquanto, que ϕ ($\phi > 0$) é conhecido e chamamos ϕ^{-1} de parâmetro de dispersão. Para o modelo em (1.1), valem as seguintes relações:

$$E(y_i) = \mu_i = b'(\theta_i), \quad \text{Var}(y_i) = \phi^{-1}V_i,$$

onde $V = d\mu/d\theta$ é chamada função de variância e $\theta = \int V^{-1}d\mu = q(\mu)$ sendo $q(\mu)$ uma função biunívoca. A função de variância caracteriza a distribuição e desempenha um importante papel na família exponencial. Ou seja, dada a função de variância, tem-se uma classe de distribuições correspondente, e vice-versa.

Os MLGs são definidos por (1.1) e pela componente sistemática

$$g(\mu_i) = \eta_i, \quad (1.2)$$

onde $\eta = X\beta$ é o preditor linear sendo $\beta = (\beta_1, \dots, \beta_p)'$ ($p < n$) um vetor de parâmetros desconhecidos a serem estimados, $X = (x_1, \dots, x_p)$ a matriz modelo de dimensão $n \times p$ conhecida de posto completo, onde $x_j = (x_{1j}, \dots, x_{nj})'$ e $g(\cdot)$ é uma função monótona e diferenciável, denominada função de ligação, relacionando a média μ_i de uma observação com o preditor linear η_i do modelo.

As distribuições normal, Poisson, binomial, gama e normal inversa são casos especiais de (1.1). Características destas distribuições são apresentadas na Tabela 1.1 .

Consideremos um MLG definido por (1.1) e (1.2). O logaritmo da função de verossimilhança como função de β pode ser expresso na forma

$$L(\beta; y) = \sum_{i=1}^n \phi[y_i\theta_i - b(\theta_i) + c(y_i)] + a(y_i, \phi).$$

Um caso importante dos MLGs ocorre quando o parâmetro natural θ da família exponencial e o preditor linear coincidem, isto é, $\theta_i = \eta_i = \sum_{j=1}^p x_{ij}\beta_j$ para $i = 1, \dots, n$. Então, $L(\beta; y)$ fica dada por

$$L(\beta; y) = \sum_{i=1}^n \phi[y_i \sum_{j=1}^p x_{ij}\beta_j - b(\sum_{j=1}^p x_{ij}\beta_j)] + \phi \sum_{i=1}^n c(y_i) + \sum_{i=1}^n a(y_i, \phi)$$

que podemos reexpressar na forma

$$L(\beta; y) = \sum_{j=1}^p S_j\beta_j - \phi \sum_{i=1}^n b(\sum_{j=1}^p x_{ij}\beta_j) + \phi \sum_{i=1}^n c(y_i) + \sum_{i=1}^n a(y_i, \phi),$$

onde $S_j = \phi \sum_{i=1}^n y_i x_{ij}$.

Tabela 1.1 *Características de algumas distribuições da família exponencial*

	Normal	Poisson	Binomial	Gama	Normal Inversa
Notação	$N(\mu, \phi^{-1})$	$P(\mu)$	$B(n, \mu)$	$G(\mu, \phi)$	$N^-(\mu, \phi)$
Suporte de y	$(-\infty, \infty)$	$0(1)\infty$	$\frac{0(1)n}{n}$	$(0, \infty)$	$(0, \infty)$
$\phi c(y) + a(y, \phi)$	$-\frac{1}{2}(\phi y^2 + \log \frac{2\pi}{\phi})$	$-\log(y)!$	$\binom{\log n}{ny}$	$(\phi - 1) \log(y) + \phi \log \phi - \log \Gamma(\phi)$	$\frac{1}{2}(\log \frac{\phi}{2\pi y^3} - \frac{\phi}{y})$
$b(\theta)$	$\theta^2/2$	e^θ	$\log(1 + e^\theta)$	$-\log(-\theta)$	$-(-2\theta)^{1/2}$
$\mu = E(y)$	θ	e^θ	$e^\theta/(1 + e^\theta)$	$-1/\theta$	$(-2\theta)^{-1/2}$
$V(\mu)$	1	μ	$\mu(1 - \mu)$	μ^2	μ^3

Nota: No modelo binomial, a variável aleatória corresponde à proporção de sucessos em n ensaios de Bernoulli e $\phi = n$.

A parametrização do modelo gama é tal que a sua variância é dada por μ^2/ϕ .

Fonte : McCullagh e Nelder (1989; Tabela 2.1, pg. 30).

Assim, a estatística $S = (S_1, \dots, S_p)$ é suficiente minimal (Dudewicz e Mishra, 1988, Cap. 8) para o vetor $\beta = (\beta_1, \dots, \beta_p)'$. As ligações que fornecem estatísticas suficientes são denominadas canônicas. Uma das vantagens de usar ligações canônicas é que elas garantem a concavidade de $L(\beta; y)$, isto é, garantem a unicidade da estimativa de máxima verossimilhança de β , quando essa existe e, conseqüentemente, muitos resultados assintóticos são obtidos mais facilmente.

As ligações canônicas para os modelos normal, Poisson, binomial, gama e normal inverso são dadas por

$$\eta = \mu, \quad \eta = \log \mu, \quad \eta = \log \left[\frac{\mu}{1 - \mu} \right], \quad \eta = \mu^{-1} \text{ e } \eta = \mu^{-2},$$

respectivamente. Algumas ligações usuais são:

Potência: $\eta = \mu^\lambda$, onde λ é um número real. Casos importantes da ligação potência são identidade, recíproca e raiz quadrada, correspondentes a $\lambda = 1, -1$ e $1/2$, respectivamente;

Probit: $\eta = \Phi^{-1}(\mu)$ sendo $\Phi(\cdot)$ a função distribuição normal padrão;

Logística: $\eta = \log[\mu/(1 - \mu)]$;

Complemento log-log: $\eta = \log[-\log(1 - \mu)]$;

Logaritmo: $\eta = \log \mu$.

Os MLGs podem ser ajustados pelo GLIM (Payne, 1986), sistema computacional largamente difundido na análise de dados via os MLGs, como também pelo S-Plus (Chambers e Hastie, 1992). Mais detalhes sobre como ajustar MLGs no GLIM são encontrados em Aitkin et al (1990).

Os MLGs são ajustados no S-Plus através da função *glm*, onde devemos especificar *formula* e *family* que são, respectivamente, a definição do modelo e a distribuição assumida pela variável resposta com a função de ligação a ser usada. Se a função de ligação usada for diferente do ‘default’, basta especificar a função de ligação desejada através do comando *link*.

A qualidade do ajuste de um MLG é avaliada através da função desvio (‘de-

viance') dada por $S_p = D^*(y; \hat{\mu}) = \phi D(y; \hat{\mu})$, onde

$$D(y; \hat{\mu}) = 2 \sum_{i=1}^n [y_i(\tilde{\theta}_i - \hat{\theta}_i) + (b(\tilde{\theta}_i) - b(\hat{\theta}_i))],$$

denotando por $\hat{\theta}_i = \theta_i(\hat{\mu}_i)$ e $\tilde{\theta}_i = \theta_i(\tilde{\mu}_i)$, respectivamente, as estimativas de máxima verossimilhança de θ_i para os modelos com p parâmetros ($p < n$) e saturado ($p = n$). Um valor pequeno para a função desvio indica que, para um menor número de parâmetros, obtém-se um ajuste tão bom quanto o ajuste com o modelo saturado. Em geral, $D(y; \hat{\mu})$ não segue assintoticamente uma distribuição χ_{n-p}^2 , embora seja usual comparar os valores observados da função desvio com os quantis desta distribuição. Geralmente, para os casos em que $D^*(y; \hat{\mu})$ depende do parâmetro de dispersão ϕ^{-1} , o seguinte resultado (Jørgensen, 1987) pode ser utilizado:

$$D^*(y; \hat{\mu}) \sim \chi_{n-p}^2, \text{ quando } \phi \rightarrow \infty.$$

Ou seja, quando a dispersão é pequena, é razoável comparar os valores observados de $D^*(y; \hat{\mu})$ com os percentis da distribuição χ_{n-p}^2 .

Apresentaremos a seguir a função desvio dos casos especiais citados anteriormente:

Normal :

$$D(y; \hat{\mu}) = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2.$$

Binomial :

$$D(y; \hat{\mu}) = 2 \sum_{i=1}^n \left[y_i \log \left(\frac{y_i}{n_i \hat{\mu}_i} \right) + (n_i - y_i) \log \left(\frac{1 - \frac{y_i}{n_i}}{1 - \hat{\mu}_i} \right) \right].$$

Todavia, quando $y_i = 0$ ou $y_i = n_i$, o i -ésimo termo de $D(y; \hat{\mu})$ se iguala a $-2n_i \log(1 - \hat{\mu}_i)$ ou $-2n_i \log \hat{\mu}_i$, respectivamente.

Gama :

$$D(y; \hat{\mu}) = 2 \sum_{i=1}^n \left[-\log \left(\frac{y_i}{\hat{\mu}_i} \right) + \frac{(y_i - \hat{\mu}_i)}{\hat{\mu}_i} \right].$$

Poisson :

$$D(y; \hat{\mu}) = 2 \sum_{i=1}^n \left[y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right].$$

Normal Inversa :

$$D(y; \hat{\mu}) = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{y_i \hat{\mu}_i^2}.$$

A análise de desvio (*ANODEV*) é uma generalização da análise de variância para os MLGs. Considere a partição do vetor de parâmetros β tal que $\beta = (\beta'_1, \beta'_2)'$, onde β_1 é um vetor q dimensional, enquanto que β_2 tem dimensão $p - q$, e considere a hipótese nula $H_0 : \beta_1 = \beta_1^{(0)}$ a ser testada contra a alternativa $H_1 : \beta_1 \neq \beta_1^{(0)}$. As funções desvio correspondentes aos modelos sob H_0 e H_1 são dadas por $D(y; \mu^{(0)})$ e $D(y; \hat{\mu})$, respectivamente, onde $\mu^{(0)}$ é a estimativa de máxima verossimilhança de μ sob H_0 .

Podemos definir a seguinte estatística

$$F = \frac{\{D(y; \mu^{(0)}) - D(y; \hat{\mu})\}/q}{D(y; \hat{\mu})/(n - p)},$$

cujas distribuição nula assintótica é $F_{q, (n-p)}$. A vantagem de utilizá-la é que a mesma não depende de ϕ e é invariante sob reparametrização. Como essa estatística pode ser obtida diretamente de funções desvio, é muito conveniente para uso prático. Através do comando *anova*, o S-Plus fornece uma tabela *ANODEV* para os ajustes colocados como objetos (ajustes de um MLG).

1.3 Estimação dos parâmetros de regressão e de dispersão

A função escore total e a matriz de informação total de Fisher para o parâmetro β são, respectivamente, dadas por

$$U(\beta) = \frac{\partial L(\beta; y)}{\partial \beta} = \phi X' W^{\frac{1}{2}} V^{-\frac{1}{2}} (y - \mu)$$

e

$$K(\beta) = E\left\{ -\frac{\partial^2 L(\beta; y)}{\partial \beta \partial \beta'} \right\} = \phi X' W X$$

onde X é a matriz modelo, $W = \text{diag}(w_1, \dots, w_n)$ com

$$w_i = \left(\frac{d\mu}{d\eta} \right)_i^2 \frac{1}{V_i},$$

e $V = \text{diag}(V_1, \dots, V_n)$.

O processo iterativo de Newton-Raphson para obter a estimativa de máxima verossimilhança de β é definido expandindo-se a função escore $U(\beta)$ em série de Taylor em torno de um valor inicial $\beta^{(0)}$ de forma que:

$$U(\beta) \cong U(\beta^{(0)}) + U'(\beta^{(0)})(\beta - \beta^{(0)}),$$

onde $U'(\beta)$ corresponde à primeira derivada de $U(\beta)$ com respeito a β . Assim, como $U(\hat{\beta}) = 0$, chega-se ao processo iterativo:

$$\beta^{(m+1)} = \beta^{(m)} + [-U'(\beta^{(m)})]^{-1}U(\beta^{(m)}),$$

$m = 0, 1, \dots$. Como a matriz $-U'(\beta)$ pode não ser positiva definida, a aplicação do método scoring de Fisher substituindo a matriz $-U'(\beta)$ pelo correspondente valor esperado, pode ser mais conveniente. Isso resulta no seguinte processo iterativo:

$$\beta^{(m+1)} = \beta^{(m)} + K(\beta^{(m)})^{-1}U(\beta^{(m)}),$$

$m = 0, 1, \dots$. Tal processo pode ser reescrito como um processo iterativo de mínimos quadrados ponderados

$$\beta^{(m+1)} = (X'W^{(m)}X)^{-1}X'W^{(m)}z^{(m)}, \quad (1.3)$$

$m = 0, 1, \dots$, onde $z = \eta + W^{-\frac{1}{2}}V^{-\frac{1}{2}}(y - \mu)$. Observe que z faz o papel de uma variável dependente modificada, enquanto que W é uma matriz de pesos que muda a cada passo do procedimento iterativo. A convergência de (1.3) ocorre em geral em um número finito de passos, independentemente dos valores iniciais utilizados (Wedderburn, 1976). É usual iniciar (1.3) com $\eta_i^{(0)} = g(y_i)$, para $i = 1, \dots, n$. Lembramos que, para o modelo linear tradicional, não é preciso recorrer ao processo iterativo (1.3) para a obtenção da estimativa irrestrita de máxima verossimilhança. Nesse caso, $\hat{\beta}$ assume a forma fechada $\hat{\beta} = (X'X)^{-1}X'y$.

Sob condições de regularidade (Sen e Singer, 1993, Cap. 7), mostra-se que $\hat{\beta}$ é um estimador consistente e eficiente de β e que

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{D} N(0, \Sigma^{-1}(\beta)),$$

onde

$$\Sigma(\beta) = \lim_{n \rightarrow \infty} \frac{K(\beta)}{n},$$

sendo $\Sigma(\beta)$ uma matriz positiva definida. Outra referência importante sobre as propriedades assintóticas dos estimadores de máxima verossimilhança dos MLGs é Fahrmeir e Kaufmann (1985).

A estimação do parâmetro ϕ , quando o mesmo é desconhecido, pode ser vista em Cordeiro e McCullagh (1991). No caso em que ϕ é desconhecido, podemos observar que os parâmetros β e ϕ são ortogonais, isto é,

$$E[(\partial L(\beta, \phi; y)/\partial \beta)(\partial L(\beta, \phi; y)/\partial \phi)] = 0.$$

Logo, os estimadores de máxima verossimilhança de ϕ e β são assintoticamente independentes. Derivando o logaritmo da função de verossimilhança com relação ao parâmetro ϕ e igualando a zero, temos a seguinte solução:

$$\sum_{i=1}^n c'(y_i, \hat{\phi}) = \frac{1}{2} D(y; \hat{\mu}) - \sum_{i=1}^n (y_i \tilde{\theta}_i - \beta(\tilde{\theta}_i)),$$

onde $D(y, \hat{\mu})$ denota o desvio do modelo sob investigação e $c'(y_i, \phi) = \partial c(y_i, \phi)/\partial \phi$. As estimativas de máxima verossimilhança para ϕ nos casos normal e normal inverso são dadas por $\hat{\phi} = n/D(y; \hat{\mu})$. Para o caso da distribuição gama, a estimativa de máxima verossimilhança de ϕ vem da equação

$$2n(\log \hat{\phi} - \psi(\hat{\phi})) = D(y; \hat{\mu}),$$

onde $\psi(\phi) = \Gamma'(\phi)/\Gamma(\phi)$ é a função digama. Expandindo-se em série de Taylor $\psi(\phi) - \log(\phi)$ e ignorando termos de ordem ϕ^{-2} , chega-se à seguinte solução:

$$\hat{\phi} = \frac{1 + (1 + 2\bar{D}/3)^{1/2}}{2\bar{D}},$$

onde $\bar{D} = D(y; \hat{\mu})/n$. Essa estimativa não é consistente quando a suposição de distribuição gama é falsa. Nesse caso, um estimador mais usado, que é consistente, é baseado na estatística de Pearson e é dado por

$$\hat{\phi}^{-1} = \sum_{i=1}^n \{(y_i - \hat{\mu}_i)/\hat{\mu}_i\}^2 / (n - p).$$

O S-Plus fornece a estimativa $\hat{\phi}^{-1} = (n - p)D^{-1}(y; \hat{\mu})$, que não é consistente.

1.4 Testes da razão de verossimilhança e escore em MLGs

Nesta seção, assumimos, que ϕ é conhecido. Na Seção 2.2.2, será considerado o caso de dispersão desconhecida.

Consideremos o vetor $y = (y_1, \dots, y_n)'$ representando n observações independentes e cujo logaritmo da função de verossimilhança total $L(\beta)$ depende de um parâmetro desconhecido $\beta = (\beta_1', \beta_2')'$ com p componentes, sendo $\beta_1 = (\beta_1, \dots, \beta_q)'$ o vetor de parâmetros de interesse e $\beta_2 = (\beta_{q+1}, \dots, \beta_p)'$ o vetor de dimensão $(p - q)$ de parâmetros de perturbação. Assume-se que a função $L(\beta)$ seja regular com respeito às derivadas em relação aos componentes de β até quarta ordem. Consideremos a correspondente partição da matriz modelo $X = (X_1, X_2)$. O interesse é testar a hipótese nula $H_0 : \beta_1 = \beta_1^{(0)}$ versus a alternativa $H_1 : \beta_1 \neq \beta_1^{(0)}$, onde $\beta_1^{(0)}$ é um vetor especificado de dimensão q ($q \leq p$).

A estatística da razão de verossimilhança para o teste de H_0 pode ser escrita da seguinte forma

$$LR = 2\{L(\hat{\beta}_1, \hat{\beta}_2) - L(\beta_1^{(0)}, \tilde{\beta}_2)\},$$

onde L é o logaritmo da função de verossimilhança total, definido na Seção 1.2 e é dado por $L = L(\beta) = \sum_{i=1}^n \log f(y_i, \beta)$, $\hat{\beta}$ é o estimador de máxima verossimilhança irrestrito de β e $\tilde{\beta}_2$ é o estimador de máxima verossimilhança de β_2 sob H_0 , isto é, sob o modelo com componente sistemática

$$\eta = \eta_1^{(0)} + \eta_2,$$

com

$$\eta_1^{(0)} = \sum_{j=1}^q x_j \beta_{1j}^{(0)} \text{ e } \eta_2 = \sum_{j=q+1}^p x_j \beta_{2j}.$$

A quantidade $\eta_1^{(0)}$ desempenha o papel de um *offset* (parte conhecida no preditor linear), que deve ser especificado diretamente na fórmula da função *glm*. Assintoticamente e sob a hipótese nula, temos que $LR \sim \chi_q^2$.

A função escore dividida da mesma maneira que β pode ser escrita como:

$$U(\beta) = (U_1'(\beta), U_2'(\beta))',$$

onde $U_1(\beta) = \phi X_1' W^{\frac{1}{2}} V^{-\frac{1}{2}}(y - \mu)$, $U_2(\beta) = \phi X_2' W^{\frac{1}{2}} V^{-\frac{1}{2}}(y - \mu)$ e a matriz de informação total de Fisher, supondo ϕ conhecido, é dada por

$$K(\beta) = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}$$

que supomos ser positiva definida com

$$K_{11} = K_{11}(\beta) = \phi X_1' W X_1, \quad K_{22} = K_{22}(\beta) = \phi X_2' W X_2,$$

$$K_{12} = K_{12}(\beta) = K_{21} = K_{21}(\beta) = \phi X_1' W X_2.$$

Utilizando a fórmula de inversa de matrizes simétricas divididas (Rao, 1973, pg. 33) e após alguma álgebra, chega-se a

$$K^{-1} = \begin{pmatrix} K^{11} & K^{12} \\ K^{21} & K^{22} \end{pmatrix}, \quad \text{com}$$

$$K^{11} = K^{11}(\beta) = \phi^{-1}(R'WR)^{-1}, \quad K^{22} = K^{22}(\beta) = \phi^{-1}[(X_2'WX_2)^{-1} + C(R'WR)^{-1}C'],$$

$$K^{12} = K^{12}(\beta) = K^{21} = K^{21}(\beta) = -\phi^{-1}(R'WR)^{-1}C',$$

onde K^{22} é a matriz de covariância assintótica de $\hat{\beta}_2$, $R = X_1 - X_2C$ com $C = (X_2'WX_2)^{-1}X_2'WX_1$ representando uma matriz $n \times q$ cujas colunas são vetores dos coeficientes da regressão linear das colunas de X_1 sobre X_2 com peso W .

A estatística escore, originalmente sugerida por Rao (1947), é dada por

$$S_R = U_1(\tilde{\beta})' K^{11}(\tilde{\beta}) U_1(\tilde{\beta}),$$

onde $\tilde{\beta} = (\beta_1^{(0)'}, \tilde{\beta}_2')'$ e $\tilde{\beta}_2$ é o estimador de máxima verossimilhança de β_2 sob H_0 . Assim, substituindo $U_1(\beta)$ e $K^{11}(\beta)$ pelas expressões dadas anteriormente, temos que a estatística escore fica dada por

$$S_R = \tilde{S}' \tilde{W}^{\frac{1}{2}} X_1 (\tilde{R}' \tilde{W} \tilde{R})^{-1} X_1' \tilde{W}^{\frac{1}{2}} \tilde{S},$$

onde $S = (S_1, \dots, S_n)'$, com $S_\ell = \phi^{\frac{1}{2}}(y_\ell - \mu_\ell) V_\ell^{-\frac{1}{2}}$ que representa o vetor de resíduos de Pearson, com as quantidades $\tilde{S}$, $\tilde{W}^{\frac{1}{2}}$ e $\tilde{R}$ sendo avaliadas em $\tilde{\beta}$. Assintoticamente e sob a hipótese nula, temos que $S_R \sim \chi_q^2$.

O teste da razão de verossimilhança envolve estimação dos parâmetros sob as hipóteses nula e alternativa, enquanto o teste escore requer estimação somente sob a hipótese nula. No caso em que a hipótese nula for $H_0 : \beta = \beta^{(0)}$, o teste escore não requer estimação alguma. Nos casos em que a estimação sob a hipótese alternativa é complicada, o teste escore pode apresentar vantagem computacional em relação ao outro. Pregibon (1982) mostrou que a estatística escore pode ser calculada no GLIM, como a diferença entre as somas de quadrados dos resíduos de dois modelos de regressão linear ponderada com a mesma facilidade com que se calcula a estatística da razão de verossimilhança. Para calcular a estatística da razão de verossimilhança no S-Plus, basta fazer a diferença dos desvios correspondentes aos modelos sob H_0 e H_1 . Por outro lado, para calcular a estatística escore, é necessário apenas fazer o ajuste do modelo sob H_0 e calcular $\tilde{W}$, $\tilde{R}$, $\tilde{S}_\ell$, chamando no S-Plus os elementos diagonais de $\tilde{W}$ por W , $\tilde{R}$ por R , $\tilde{S}_\ell$ por r_p e as colunas da matriz modelo X sob restrição por imp . Os passos para o cálculo de S_R são dados abaixo, para um modelo de Poisson

```

X1<— X[,imp]
X2<— X[,-imp]
Form<— y ~ -1+X2
fit<— glm(formula=Form,family=poisson)
rp<— resid(fit,type="pearson")
w<— fitted(fit)
W<— diag(w)
E<— sqrt(W)
S<— sqrt(phi)*rp
S<— matrix(S, n, 1)
C<— solve(t(X2)%*%W%*%X2)%*%t(X2)%*%W%*%X1
R<— X1-X2%*%C
T<— solve(t(R)%*%W%*%R)
SR<— t(S)%*%E%*%X1%*%T%*%t(X1)%*%E%*%S.

```

CAPÍTULO 2

Correções de Bartlett e tipo-Bartlett

2.1 Introdução

Em problemas regulares, as estatísticas da razão de verossimilhança (LR) e escore (S_R) têm, sob a hipótese nula, uma distribuição χ^2_q aproximadamente, em amostras grandes, ou seja, χ^2 com q graus de liberdade, onde q é a diferença entre as dimensões dos espaços paramétricos sob as hipóteses alternativa e nula. Os testes são usualmente baseados na comparação das estatísticas com o valor crítico obtido da distribuição χ^2 de referência para um nível de significância nominal fixado.

Em geral, há grande dificuldade em se determinar as distribuições exatas das estatísticas LR e S_R . Esta é a razão pela qual os testes têm sido construídos com base em resultados assintóticos. No entanto, as aproximações resultantes podem não ser boas para amostras pequenas ou mesmo de tamanho moderado. Para melhorar a qualidade da aproximação das distribuições das estatísticas LR e S_R por qui-quadrado são utilizadas as chamadas correções de Bartlett e tipo-Bartlett.

A idéia de modificar a estatística da razão de verossimilhança por um fator de correção, visando a produzir uma estatística modificada com o primeiro momento igual ao da distribuição χ^2 de referência, é devida a Bartlett (1937). Sob a hipótese nula, a média da estatística LR pode ser expandida, em problemas regulares, como $E(LR) = q\{1 + d + O(n^{-2})\}$, onde d é uma constante de ordem n^{-1} que pode ser consistentemente estimada sob H_0 e n é o tamanho da amostra. Assim, a média da estatística modificada $LR^* = LR/(1 + d)$ ou, equivalentemente, $LR_1^* = LR(1 - d)$ é dada por $q + O(n^{-2})$ e é ‘mais próxima’ daquela da distribuição χ^2_q do que a média de LR . Os fatores $1/(1+d)$ e $(1-d)$ ficaram conhecidos como correções de Bartlett. Deve-se destacar que tais fatores não dependem do valor da estatística da razão de verossimilhança, mas podem depender de parâmetros desconhecidos e,

neste caso, estes devem ser substituídos por suas respectivas estimativas de máxima verossimilhança sob a hipótese nula.

Lawley (1956) desenvolveu um método geral de obtenção de d , e mostrou que a estatística da razão de verossimilhança modificada tem todos os momentos concordando com os respectivos da distribuição χ^2 de referência, ignorando os termos de ordem n^{-2} .

Hayakawa (1977) derivou uma expansão assintótica para a distribuição de LR sob a hipótese nula e mostrou que, se a hipótese nula for simples, a estatística LR^* tem distribuição χ^2 até ordem n^{-1} . Entretanto, para hipóteses compostas, só seria possível obter o fator de correção se um determinado coeficiente da expansão fosse nulo. Chesher e Smith (1995) notam um erro na fórmula de tal coeficiente e mostram que, após uma correção nesta fórmula, o valor zero é sempre obtido.

Cordeiro (1982) propôs fatores de correção para a estatística da razão de verossimilhança do teste de adequacidade em modelos multinomiais e também desenvolveu estudos de simulação para a distribuição multinomial mostrando que as taxas de rejeição da estatística corrigida LR^* estão mais próximas dos correspondentes níveis nominais da distribuição χ^2 de referência do que as taxas de rejeição da estatística não corrigida.

Cordeiro (1983 e 1987) obtém uma estatística da razão de verossimilhança modificada nos MLGs quando o fator de escala é conhecido e desconhecido, respectivamente, calculando o termo de ordem n^{-1} do valor esperado do desvio, e mostra que o uso do fator de correção representa um grande aperfeiçoamento no teste de adequação do MLG. Referências sobre artigos que tratam de correções de Bartlett em outros modelos podem ser encontradas no trabalho de Cribari-Neto e Cordeiro (1997).

Harris (1985), seguindo as idéias de Hayakawa (1977), desenvolve uma expansão assintótica até ordem n^{-1} para a distribuição da estatística escore. Como pensava-se que a estatística escore não admitisse um fator de correção simples como o da razão de verossimilhança, Harris propôs um teste escore aperfeiçoado baseado na própria estatística escore e em uma expansão, até ordem n^{-1} , para os seus quantis.

Cordeiro e Ferrari (1991) definem uma estatística escore modificada S_R^* que tem distribuição qui-quadrado até ordem n^{-1} . A estatística modificada é dada por S_R multiplicada por um fator de correção, denominado correção tipo-Bartlett, que envolve um polinômio de segundo grau na própria estatística e cujos coeficientes são, em geral, funções de parâmetros desconhecidos que devem ser substituídos por suas respectivas estimativas de máxima verossimilhança sob a hipótese nula.

Cordeiro, Ferrari e Paula (1993) obtiveram correções tipo-Bartlett para estatísticas escore em MLGs para o caso em que o parâmetro de dispersão é conhecido. Estes resultados foram generalizados por Cribari-Neto e Ferrari (1995) que obtiveram correções para o caso de dispersão desconhecida.

No artigo de Cribari-Neto e Cordeiro (1997) podem ser encontradas referências de outros trabalhos que tratam de correções tipo-Bartlett para a estatística escore em modelos de regressão.

É conveniente salientar que as expansões assintóticas de LR e S_R não são válidas se a distribuição dos dados for discreta. Não se pode garantir, nestes casos, que o fator de Bartlett e a multiplicação da estatística escore pelo fator de correção tipo-Bartlett venham a melhorar a aproximação de sua distribuição por qui-quadrado. Fridenberg e Jensen (1989) mostram, via simulação, que em situações particulares, os desempenhos dos testes modificados chegam a ser piores que os dos não modificados. No entanto, Cordeiro (1982) apresenta estudos de simulação que mostram que testes da razão de verossimilhança modificados têm um melhor desempenho que os correspondentes testes não modificados em alguns modelos multinomiais e de Poisson. Estudos de simulação em modelos discretos serão apresentados no Capítulo 4.

2.2 Correção de Bartlett para a estatística da razão de verossimilhança

No Capítulo 1, definimos a classe dos modelos lineares generalizados (MLGs) supondo o parâmetro de dispersão ϕ^{-1} conhecido e igual para todas as observações e apresentamos testes da razão de verossimilhança e escore para a hipótese nula

$H_0 : \beta_1 = \beta_1^{(0)}$ a ser testada contra a alternativa $H_1 : \beta_1 \neq \beta_1^{(0)}$. Assumimos, por enquanto, que ϕ seja conhecido.

Lawley (1956) obteve, sob condições gerais de regularidade, uma expansão de $L(\hat{\beta}_1, \hat{\beta}_2)$ em série de Taylor sob a hipótese nula até termos de ordem n^{-1} envolvendo derivadas até de quarta ordem do logaritmo da função de verossimilhança. Demonstrou, após um cálculo de álgebra bastante complicado, que

$$2E\{L(\hat{\beta}_1, \hat{\beta}_2) - L(\beta_1, \beta_2)\} = p + \epsilon_{p,\beta} + O(n^{-2}),$$

onde o termo $\epsilon_{p,\beta}$, dado em seu artigo, é de ordem n^{-1} . Este termo depende de cumulantes de derivadas do logaritmo da função de verossimilhança e de suas derivadas, sendo avaliado no parâmetro verdadeiro. Esta expressão tem sido usada para obter o valor esperado do desvio até ordem n^{-1} e o fator de correção de Bartlett para MLGs. Lawley mostrou também que

$$2E(L(\beta_1^{(0)}, \tilde{\beta}_2) - L(\beta_1, \beta_2)) = p - q + \epsilon_{p-q,\beta} + O(n^{-2}).$$

Podemos reescrever a estatística da razão de verossimilhança LR da seguinte forma:

$$LR = 2[\{L(\hat{\beta}_1, \hat{\beta}_2) - L(\beta_1, \beta_2)\} - \{L(\beta_1^{(0)}, \tilde{\beta}_2) - L(\beta_1, \beta_2)\}].$$

O termo $2(L(\hat{\beta}_1, \hat{\beta}_2) - L(\beta_1, \beta_2))$ representa o desvio avaliado no parâmetro verdadeiro. O valor esperado de LR fica, então, dado por

$$\begin{aligned} E(LR) &= 2E[\{L(\hat{\beta}_1, \hat{\beta}_2) - L(\beta_1, \beta_2)\} - \{L(\beta_1^{(0)}, \tilde{\beta}_2) - L(\beta_1, \beta_2)\}] \\ &= q + \epsilon_{p,\beta} - \epsilon_{p-q,\beta} + O(n^{-2}), \\ &= q \left(1 + \frac{\epsilon_{p,\beta} - \epsilon_{p-q,\beta}}{q} \right) + O(n^{-2}). \end{aligned}$$

Assim, podemos melhorar a aproximação da média da estatística LR pela média da distribuição χ_q^2 substituindo LR pela estatística modificada LR^* dada por $LR^* = LR/(1 + d)$ onde o fator de correção de Bartlett $1/(1 + d)$ é determinado através de

$$d = \frac{\epsilon_{p,\beta} - \epsilon_{p-q,\beta}}{q}. \quad (2.1)$$

Hayakawa (1977) (ver correção de Chesher e Smith, 1995) mostra que, sob certas condições de regularidade, a estatística modificada LR^* tem distribuição χ_q^2 até ordem n^{-1} sob H_0 . Um teste da razão de verossimilhança aperfeiçoado compara, então, a estatística modificada LR^* com a distribuição χ_q^2 de referência.

No caso do teste da hipótese nula simples $H_0 : \beta = \beta^{(0)}$, a quantidade d dada em (2.1) que determina o fator de correção de Bartlett se reduz a $d = \epsilon_{p,\beta}/p$.

2.2.1 Correção de Bartlett em MLGs com dispersão conhecida

Os fatores de correção de Bartlett dependem da quantidade $\epsilon_{p,\beta}$, que é uma função aparentemente complicada dos cumulantes conjuntos de derivadas do logaritmo da função de verossimilhança total. Dependendo do modelo associado aos dados, pode ser muito difícil obter tal quantidade. Então, é importante expressar $\epsilon_{p,\beta}$ em forma simples para testes da razão de verossimilhança em MLGs.

Considere um MLG definido por (1.1) e (1.2) e suponha que ϕ é conhecido. Cordeiro (1983) demonstra que, para um MLG, o termo $\epsilon_{p,\beta}$ é dado por

$$\epsilon_{p,\beta} = \frac{1}{\phi} \left\{ \frac{1}{4} \text{tr}(J Z_d^2) - \frac{1}{3} 1' G Z^{(3)} (F + G) 1 + \frac{1}{12} 1' F (2Z^{(3)} + 3Z_d Z Z_d) F 1 \right\}, \quad (2.2)$$

onde $Z = \{z_{ij}\} = X(X'WX)^{-1}X'$ é uma matriz de dimensão $n \times n$ de posto p , $Z_d = \text{diag}\{z_{11}, \dots, z_{nn}\}$ é uma matriz diagonal formada pelos correspondentes elementos da diagonal da matriz Z e $1 = (1 \ 1 \dots 1)'$ é um vetor $n \times 1$ de 1's. Denotamos $Z^{(3)} = Z^{(2)} \otimes Z$, $Z^{(2)} = Z \otimes Z$, onde $\otimes$ denota o produto de Hadamard (Rao, 1973, p. 30), ou seja, o elemento (i,j) de $Z^{(2)}$ é z_{ij}^2 . Na equação (2.2) foi utilizada também a seguinte notação: $F = \text{diag}\{f_1, \dots, f_n\}$, $G = \text{diag}\{g_1, \dots, g_n\}$ e $J = \text{diag}\{j_1, \dots, j_n\}$, matrizes diagonais de dimensão $n \times n$ cujos elementos são definidos por

$$f = \frac{1}{V} \frac{d\mu}{d\eta} \frac{d^2\mu}{d\eta^2},$$

$$g = \frac{1}{V} \frac{d\mu}{d\eta} \frac{d^2\mu}{d\eta^2} - \frac{1}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^3,$$

$$j = \frac{1}{V} \frac{d^2 \mu}{d\eta^2} \left[\frac{d^2 \mu}{d\eta^2} - 4w \frac{dV}{d\mu} \right] + w^2 \left[\frac{2}{V} \left(\frac{dV}{d\mu} \right)^2 - \frac{d^2 V}{d\mu^2} \right],$$

onde

$$w = \left(\frac{d\mu}{d\eta} \right)^2 \frac{1}{V}.$$

A vantagem de expressarmos $\epsilon_{p,\beta}$ em notação matricial é poder obter fatores de correção de Bartlett em forma fechada para vários testes de hipóteses de interesse prático. Além disso, $\epsilon_{p,\beta}$ pode ser calculado numericamente através do S-Plus e ou mesmo do GLIM usando, neste caso, macros especiais que incluem aquelas para multiplicação de matrizes.

É importante observar que a matriz $\phi^{-1}Z$ é a matriz de covariância assintótica de $X\hat{\beta}$. As matrizes W , F , G e J dependem, apenas, das funções de ligação e de variância e de suas duas primeiras derivadas. Todas as matrizes podem ser calculadas facilmente através de operações simples de vetores e matrizes. Uma vez calculadas as matrizes acima, o cálculo de $\epsilon_{p,\beta}$ é imediato.

A fórmula de $\epsilon_{p,\beta}$, em geral, depende, fortemente, da estrutura de covariância assintótica $K^{-1} = (X'WX)^{-1}$ e depende da distribuição dos dados apenas através da função de variância e de suas duas primeiras derivadas. O termo $\epsilon_{p,\beta}$ é diretamente proporcional ao parâmetro de dispersão ϕ tornando-se desprezível se ϕ^{-1} é muito pequeno.

A fórmula de $\epsilon_{p-q,\beta}$ é definida analogamente à de $\epsilon_{p,\beta}$ dada em (2.2) com Z e Z_d substituídos por Z_2 e Z_{2d} , onde $Z_2 = X_2(X_2'WX_2)^{-1}X_2'$ e $Z_{2d} = \text{diag}\{z_{211} \cdots z_{2nn}\}$. A matriz $\phi^{-1}Z_2$ é a matriz de covariância assintótica de $X_2\hat{\beta}_2$. Se denotarmos $\epsilon_{p,\beta} = h(\phi, X, \mu)$, temos que $\epsilon_{p-q,\beta} = h(\phi, X_2, \mu)$. O fator de correção de Bartlett para o teste da razão de verossimilhança é obtido de (2.1) com as quantidades $\epsilon_{p,\beta}$ e $\epsilon_{p-q,\beta}$ deduzidas de (2.2).

2.2.2 Correção de Bartlett em MLGs com dispersão desconhecida

Nesta seção, eliminamos a suposição de que ϕ é conhecido, sendo agora um parâmetro de perturbação e apresentamos o fator de correção de Bartlett obtido

por Cordeiro (1987) para a estatística da razão de verossimilhança para as hipóteses descritas na Seção 2.2.

Consideremos n variáveis aleatórias independentes $y_1, \dots, y_n$ com função densidade (1.1) com $a(y, \phi) = d_1(\phi) + d_2(y)$, isto é,

$$f(y; \theta_i, \phi) = \exp\{\phi[y\theta_i - b(\theta_i) + c(y)] + d_1(\phi) + d_2(y)\}, \quad i = 1, \dots, n$$

que corresponde a um modelo da família exponencial de distribuições na forma natural com parâmetros canônicos ϕ e $\phi\theta$. Supomos que $d_1(\cdot)$ tem as quatro primeiras derivadas. Seguindo a parametrização usada na Tabela 1.1 temos que para a distribuição gama $d_1(\phi) = \phi \log \phi - \log \Gamma(\phi)$ e $d_1(\phi) = 1/2 \log \phi$ para as distribuições normal e normal inversa.

Agora, o logaritmo da função de verossimilhança $L(\beta, \phi)$ é uma função de β e ϕ dado o vetor de observações $y = (y_1, \dots, y_n)'$. A matriz de informação de Fisher para (β, ϕ) é dada por

$$K(\beta, \phi) = \begin{pmatrix} K_\beta & 0 \\ 0 & K_\phi \end{pmatrix} = \text{diag}\{\phi X'WX, -nd_1''(\phi)\},$$

implicando que os parâmetros β e ϕ são globalmente ortogonais (Cox e Reid, 1987). Exemplos de modelos com dois parâmetros globalmente ortogonais incluem o modelo normal com média μ e variância ϕ^{-1} , o modelo gama com média μ e índice ϕ e o modelo normal inverso com média μ e parâmetro de precisão ϕ .

Sejam

$$t_p(\mu, y) = 2 \sum_{i=1}^n \{v(y_i) + c(y_i)\} - D_p(\mu, y),$$

onde $v(\mu) = xq(\mu) - b(q(\mu))$, e

$$D_p(\mu, y) = 2 \sum_{i=1}^n \{v(y_i) - v(\mu_i) + (\mu_i - y_i)q(\mu_i)\}.$$

Usando esta notação, a estatística da razão de verossimilhança, LR , cuja distribuição assintótica é χ_q^2 sob H_0 , pode ser reescrita da seguinte forma

$$LR = \hat{\phi} t_p(\hat{\mu}, y) - \tilde{\phi} t_{p-q}(\tilde{\mu}, y) + 2n\{d_1(\hat{\phi}) - d_1(\tilde{\phi})\},$$

onde $\hat{\phi}$, $\hat{\mu}$, $\tilde{\phi}$ e $\tilde{\mu}$ são estimadores de máxima verossimilhança irrestritos e restritos de ϕ e $\mu = d^{-1}(\eta)$, respectivamente. Para a distribuição *normal*: $t_p(\mu, y) = -D_p(\mu, y)$ e $v(y_i) = y_i(\tilde{\theta}_i) - b(\tilde{\theta}_i)$, com $\tilde{\theta}_i = y_i$ e $b(\tilde{\theta}_i) = y_i^2/2$. Para a distribuição *normal inversa*: $t_p(\mu, y) = -D_p(\mu, y)$ e $v(y_i) = y_i(\tilde{\theta}_i) - b(\tilde{\theta}_i)$, com $\tilde{\theta}_i = -1/(2y_i^2)$ e $b(\tilde{\theta}_i) = -1/y_i$. Para a distribuição *gama*: $t_p(\mu, y) = -2n - D_p(\mu, y)$ e $v(y_i) = y_i(\tilde{\theta}_i) - b(\tilde{\theta}_i)$, com $\tilde{\theta}_i = -1/y_i$ e $b(\tilde{\theta}_i) = \log(y_i)$.

Novamente, através dos resultados de Lawley (1956) e utilizando a notação introduzida na Seção 1.2, encontramos, até ordem n^{-1} e sob H_0 ,

$$E(LR) = q + \epsilon_{p,\beta} - \epsilon_{p-q,\beta} + \gamma,$$

onde $\epsilon_{p,\beta}$ e $\epsilon_{p-q,\beta}$ são obtidos de (2.2) e γ depende da distribuição assumida para os dados e da estrutura do modelo considerado somente através do posto da matriz modelo correspondente às hipóteses H_0 e H_1 , respectivamente, e do parâmetro ϕ e é dada por

$$\gamma = \frac{2d_3q - (2p - q - 2)d_2q}{4nd_2^2},$$

onde $d_2 = \phi^2 d_1'''(\phi)$ e $d_3 = \phi^3 d_1''''(\phi)$.

Para os modelos normal e normal inverso, temos que

$$\gamma = \frac{4q + (2p - q - 2)q}{2n}$$

e para o modelo gama,

$$\gamma = -\frac{2q(1 + \phi^2 d_1''''(\phi)) + (1 - \phi\psi'(\phi))(2p - q - 2)q}{4n\phi(1 - \phi\psi'(\phi))^2}.$$

Assim, o fator de correção de Bartlett para testar H_0 é dado por

$$d = \frac{\epsilon_{p,\beta} - \epsilon_{p-q,\beta}}{q} + \frac{2d_3 - (2p - q - 2)d_2}{4nd_2^2}$$

A expressão do fator de correção de Bartlett pode ser interpretada como uma combinação linear de duas partes. O primeiro termo constitui o fator de correção de Bartlett para o teste de $H_0 : \beta_1 = \beta_1^{(0)}$ quando o parâmetro de dispersão é conhecido. O termo restante representa a contribuição ao fator de correção de Bartlett devido ao fato de ϕ ser desconhecido.

Nos Capítulos 3 e 4, iremos comparar os testes da razão de verossimilhança baseados na estatística LR e nas versões modificadas $LR^* = LR/(1 + d)$ e $LR_1^* = LR(1 - d)$ através de simulações.

2.3 Correção tipo-Bartlett para a estatística escore

Uma expansão assintótica sob a hipótese nula da distribuição da estatística escore S_R para o teste discutido na Seção 2.2 foi obtida por Harris (1985).

Cordeiro e Ferrari (1991) mostram que sempre é possível encontrar uma estatística escore aperfeiçoada tendo distribuição χ^2 até ordem n^{-1} sob a hipótese nula. Sob condições gerais de regularidade, a estatística escore pode ser modificada por um fator de correção que envolve constantes de ordem n^{-1} e um polinômio de segundo grau na própria estatística produzindo uma estatística ajustada com distribuição χ^2 até ordem n^{-1} segundo a hipótese nula. Podemos definir a estatística escore modificada da seguinte forma:

$$S_R^* = S_R \{1 - (c + bS_R + aS_R^2)\},$$

onde a , b e c , de ordem n^{-1} , são dados por

$$a = \frac{A_3}{12q(q+2)(q+4)}, \quad b = \frac{A_2 - 2A_3}{12q(q+2)}, \quad c = \frac{A_1 - A_2 + A_3}{12q}$$

sendo A_1 , A_2 e A_3 funções de cumulantes conjuntos de derivadas do logaritmo da função de verossimilhança.

Portanto, uma estatística escore modificada tendo distribuição χ^2 até ordem n^{-1} é dada por:

$$S_R^* = S_R \{1 - b(S_R, A_1, A_2, A_3, q)\}, \quad (2.3)$$

onde

$$b(x, A_1, A_2, A_3, q) = \frac{A_1 - A_2 + A_3}{12q} + \frac{A_2 - 2A_3}{12q(q+2)}x + \frac{A_3}{12q(q+2)(q+4)}x^2.$$

O fator multiplicativo entre chaves é um tipo de correção de Bartlett envolvendo cumulantes de derivadas do logaritmo da função de verossimilhança como função da própria estatística S_R e é um polinômio de segundo grau em S_R .

Uma vez que os A 's podem ser calculados pela fórmula de Harris (1985), um teste escore aperfeiçoado pode ser conduzido utilizando-se a estatística S_R^* e a distribuição χ_q^2 como referência.

Em geral, S_R^* é uma função de parâmetros desconhecidos. Neste caso, podemos usar a estatística $\hat{S}_R^*$ com os parâmetros substituídos pelos respectivos estimadores de máxima verossimilhança sob a hipótese nula. Sob H_0 e até ordem n^{-1} , as distribuições de S_R^* e $\hat{S}_R^*$ coincidem (Cordeiro e Ferrari, 1991). Assim, até esta ordem, $\hat{S}_R^*$ tem também distribuição χ_q^2 .

A forma da estatística escore modificada em (2.3) está fortemente relacionada com os quantis modificados de S_R obtidos por Harris (1985). Ele mostrou que, desprezando termos de ordem inferior a n^{-1} , $P(S_R \leq z_\alpha) = P(\chi_q^2 \leq x_\alpha) = 1 - \alpha$, com

$$z_\alpha = x_\alpha \{1 + b(x_\alpha, A_1, A_2, A_3, q)\}.$$

Comparando as formas funcionais dos fatores de correção de S_R e de seus quantis vemos que existe uma conexão estreita entre S_R^* e a expressão assintótica dos quantis de S_R . Desprezando termos de ordem inferior a n^{-1} , temos $P(S_R^* \leq x_\alpha) = P(S_R \leq z_\alpha) = P(\chi_q^2 \leq x_\alpha) = 1 - \alpha$, sob H_0 . Um teste escore aperfeiçoado pode ser conduzido, então, das seguintes maneiras: obtém-se a estatística escore aperfeiçoada dada em (2.3) e utiliza-se a distribuição χ_q^2 como referência, ou obtém-se a estatística escore usual e utilizam-se os quantis modificados z_α como referência.

2.3.1 Correção tipo-Bartlett em MLGs com dispersão conhecida

As correções tipo-Bartlett apresentadas na seção anterior dependem das quantidades A_1 , A_2 e A_3 . Cordeiro, Ferrari e Paula (1993) exibem os A 's em notação matricial para testes escore em MLGs supondo que ϕ é conhecido para a hipótese nula $H_0 : \beta_1 = \beta_1^{(0)}$ a ser testada contra a alternativa $H_1 : \beta_1 \neq \beta_1^{(0)}$. Cribari-Neto em sua tese de doutorado (1994) mostra como podemos obter os A 's através da implementação de um programa no S-Plus.

Cordeiro, Ferrari e Paula (1993) mostram que os A 's aqui denotados por $A_{1,\beta}$,

$A_{2,\beta}$, $A_{3,\beta}$ são dados por

$$\begin{aligned} A_{1,\beta} = & \phi^{-1}[3 \mathbf{1}' F Z_{2d}(Z - Z_2)Z_{2d}F + 6 \mathbf{1}' F Z_{2d}Z_2(Z - Z_2)_d(F - G) \mathbf{1} \\ & - 6 \mathbf{1}' F[Z_2^2 \odot (Z - Z_2)](2G - F) \mathbf{1} - 6 \mathbf{1}' H(Z - Z_2)_d Z_{2d} \mathbf{1}], \end{aligned} \quad (2.4)$$

$$\begin{aligned} A_{2,\beta} = & \phi^{-1}[-3 \mathbf{1}' (F - G)(Z - Z_2)_d Z_2(Z - Z_2)_d(F - G) \mathbf{1} \\ & - 6 \mathbf{1}' F Z_{2d}(Z - Z_2)(Z - Z_2)_d(F - G) \mathbf{1} \\ & - 6 \mathbf{1}' (F - G)[(Z - Z_2)^{(2)} \otimes Z_2](F - G) \mathbf{1} + 3 \mathbf{1}' B(Z - Z_2)_d^{(2)} \mathbf{1}], \end{aligned} \quad (2.5)$$

$$\begin{aligned} A_{3,\beta} = & \phi^{-1}[3 \mathbf{1}' (F - G)(Z - Z_2)_d(Z - Z_2)(Z - Z_2)_d(F - G) \mathbf{1} \\ & + 2 \mathbf{1}' (F - G)(Z - Z_2)^{(3)}(F - G) \mathbf{1}], \end{aligned} \quad (2.6)$$

onde $B = \text{diag}\{b_1, \dots, b_n\}$ e $H = \text{diag}\{h_1, \dots, h_n\}$ são matrizes diagonais de dimensão $n \times n$ com elementos definidos por

$$b = \frac{1}{V^3} \left(\frac{d\mu}{d\eta} \right)^4 \left[\left(\frac{dV}{d\mu} \right)^2 + V \frac{d^2 V}{d\mu^2} \right]$$

e

$$h = \frac{1}{V^2} \frac{dV}{d\mu} \left(\frac{d\mu}{d\eta} \right)^2 \frac{d^2 \mu}{d\eta^2} + \frac{1}{V^2} \frac{d^2 V}{d\mu^2} \left(\frac{d\mu}{d\eta} \right)^4,$$

respectivamente, F , G , Z e Z_2 foram definidas na Seção 2.2.1. Observe que as expressões de $A_{1,\beta}$, $A_{2,\beta}$ e $A_{3,\beta}$ são funções das matrizes Z_2 , $Z - Z_2$ e das médias desconhecidas e são ponderadas pelo parâmetro de dispersão. Os A 's dependem da função de ligação e de variância com suas duas primeiras derivadas. O termo $A_{3,\beta}$ também depende das duas primeiras derivadas da função de ligação, porém, somente da primeira derivada da função de variância.

O cálculo de $A_{1,\beta}$, $A_{2,\beta}$ e $A_{3,\beta}$ é feito a partir das expressões (2.4) – (2.6) que envolvem somente operações com vetores e matrizes, desde que as médias desconhecidas tenham sido estimadas sob H_0 e que as matrizes Z_2 e $Z - Z_2$ tenham sido obtidas. Para MLGs com formas fechadas para as matrizes Z_2 e $Z - Z_2$, é possível obter formas fechadas para as expressões dos A 's.

Cordeiro, Ferrari e Paula (1993) mostraram que $Z - Z_2 = R(R'WR)^{-1}R'$ onde

R está definida na Seção 1.4. Numericamente, pode ser mais fácil avaliar $Z - Z_2$ a partir desta expressão pois esta envolve somente a inversa de uma matriz $q \times q$ enquanto que Z e Z_2 envolvem inversas de matrizes $p \times p$ e $(p - q) \times (p - q)$, ($p \geq q$).

A estatística escore aperfeiçoada para o teste discutido na Seção 2.2 é obtida de (2.3) com os A 's dados em (2.4) – (2.6).

2.3.2 Correção tipo-Bartlett com dispersão desconhecida

Nesta seção, eliminamos a suposição de que ϕ é conhecido, sendo agora um parâmetro de perturbação e apresentamos um fator de correção tipo-Bartlett à estatística escore para as hipóteses descritas na seção anterior. A expressão de S_R coincide com o caso de ϕ conhecido, sendo ϕ substituído por seu estimador de máxima verossimilhança sob H_0 .

Quando ϕ é desconhecido, Cribari-Neto e Ferrari (1995) mostram que

$$A_1 = A_{1,\beta} + A_{1,\beta\phi}, \quad A_2 = A_{2,\beta} + A_{2,\beta\phi} \quad \text{e} \quad A_3 = A_{3,\beta},$$

onde $A_{1,\beta}$, $A_{2,\beta}$ e $A_{3,\beta}$ coincidem com as expressões em (2.4)-(2.6), respectivamente,

$$A_{1,\beta\phi} = \frac{6q(d_3 + (2 - p + q)d_2)}{nd_2^2} \quad (2.7)$$

e

$$A_{2,\beta\phi} = \frac{3q(q + 2)}{nd_2}, \quad (2.8)$$

onde d_2 e d_3 foram definidos na Seção 2.2. Assim, as expressões A_1 e A_2 são iguais aos A 's correspondentes ao caso em que o parâmetro ϕ é conhecido, mais termos adicionais que dependem somente do número total dos parâmetros (p), do número de restrições (q), do parâmetro de dispersão (ϕ) e da função $d_1(\cdot)$ através da segunda e da terceira derivadas e representam efeitos provenientes do fato de ϕ ser desconhecido. Em particular, estes termos não dependem da matriz modelo X , exceto pelo seu posto. Entretanto, as expressões para A_3 para os casos de ϕ conhecido e desconhecido coincidem.

Os coeficientes A_1 , A_2 e A_3 são usados para modificar a estatística escore nos

MLGs quando o parâmetro de dispersão é desconhecido e são dados por (2.4)+(2.7), (2.5)+(2.8) e (2.6), respectivamente, com ϕ substituído por seu estimador de máxima verossimilhança sob H_0 . Para os modelos normal e normal inverso, temos que

$$A_{1,\beta\phi} = \frac{12q(p-q)}{n} \quad \text{e} \quad A_{2,\beta\phi} = -\frac{6q(q+2)}{n}.$$

Já, para o modelo gama, temos

$$A_{1,\beta\phi} = -\frac{6q[(1 + \phi^2\psi''(\phi)) - (1 - \phi\psi'(\phi))(2 - p + q)]}{n\phi(1 - \phi\psi'(\phi))^2}$$

e

$$A_{2,\beta\phi} = \frac{3q(q+2)}{n\phi(1 - \phi\psi'(\phi))},$$

onde $\psi'(\cdot)$ e $\psi''(\cdot)$ são as funções trigama e tetragama, respectivamente. Tabelas que fornecem os valores de $\psi'(x)$ e $\psi''(x)$ para $1 \leq x \leq 2$ são encontradas no livro de Abramowitz e Stegun (1980; pgs. 267-271). A relação de recorrência $\psi^{(n)}(x+1) = \psi^{(n)}(x) + (-1)^n n! x^{-n-1}$ pode ser utilizada para avaliar $\psi(x)$ e suas derivadas fora deste intervalo. Os sistemas computacionais *MATHEMATICA* e *MAPLE V* também avaliam estas funções.

2.3.3 Formas alternativas de correção tipo-Bartlett

A correção tipo-Bartlett para a estatística escore é usualmente definida como $S_R^* = S_R(1 - B)$, onde B é um polinômio na estatística S_R . Entretanto, existem outras definições alternativas para a correção tipo-Bartlett que são equivalentes, tais como

$$S_{R1}^* = \frac{S_R}{(1 + B)} \quad \text{e} \quad S_{R2}^* = S_R \exp(-B)$$

quando termos de ordem menor do que n^{-1} são ignorados. Observe que a última estatística tem a vantagem de sempre ser não negativa.

As estatísticas modificadas S_R^* , S_{R1}^* e S_{R2}^* podem ser funções não monótonas em S_R . Cordeiro e Ferrari (1996) obtiveram uma estatística escore modificada, S_{R3}^* digamos, que é assintoticamente equivalente a $S_R^* = S_R\{1 - (c + b S_R + a S_R^2)\}$ e

tem a propriedade de ser monótona em S_R . Esta estatística é dada por

$$S_{R3}^* = \begin{cases} \sqrt{\frac{\pi}{3a}} \exp\left(\frac{b^2}{3a} - c\right) \left\{ \Phi\left(\sqrt{6a}S_R + \sqrt{\frac{2}{3a}}b\right) - \Phi\left(\sqrt{\frac{2}{3a}}b\right) \right\}, & \text{se } a > 0, \\ \frac{1}{2b} \exp(-c) \{1 - \exp(-2bS_R)\}, & \text{se } a = 0 \text{ e } b \neq 0. \end{cases} \quad (2.9)$$

Note que se $a = 0$ e $b = 0$, então $S_R^* = S_R\{1 - c\}$, que é uma função monótona em S_R . Ferrari (1991) mostra que $a \geq 0$ sempre.

Nos Capítulos 3 e 4, iremos comparar as diversas formas alternativas da estatística escore modificada através de simulações.

2.4 Alguns casos especiais

Para um modelo linear $\eta = \mu$ com qualquer distribuição em (1.1), temos que:

$$w = \frac{1}{V},$$

$$f = 0,$$

$$g = -\frac{1}{V^2} \frac{dV}{d\mu},$$

$$j = \frac{2}{V^3} \left(\frac{dV}{d\mu}\right)^2 - \frac{1}{V^2} \left(\frac{d^2V}{d\mu^2}\right),$$

$$h = \frac{1}{V^2} \frac{d^2V}{d\mu^2},$$

$$b = \frac{1}{V^3} \left[\left(\frac{dV}{d\mu}\right)^2 + V \frac{d^2V}{d\mu^2} \right].$$

Neste caso, o terceiro termo em $\epsilon_{p,\beta}$ é zero e o termo $A_{1,\beta}$ se reduz à sua última parcela, embora $A_{2,\beta}$ e $A_{3,\beta}$ não tenham uma maior redução.

A Tabela 2.1 apresenta as funções w , g , j , h e b para modelos lineares.

Tabela 2.1 *Modelos lineares - funções w , g , j , h e b*

Modelo	w	g	j	h	b
$N(\mu, \phi^{-1})$	1	0	0	0	0
$P(\mu)$	$\frac{1}{\mu}$	$-\frac{1}{\mu^2}$	$\frac{2}{\mu^3}$	0	$\frac{1}{\mu^3}$
$B(n, \mu)$	$\frac{1}{\mu(1-\mu)}$	$\frac{2\mu-1}{\mu^2(1-\mu)^2}$	$\frac{1-3\mu(1-\mu)}{\mu^3(1-\mu)^3}$	$-\frac{2}{\mu^2(1-\mu)^2}$	$\frac{1-6\mu(1-\mu)}{\mu^3(1-\mu)^3}$
$G(\mu, \phi)$	$\frac{1}{\mu^2}$	$-\frac{2}{\mu^3}$	$\frac{6}{\mu^4}$	$\frac{2}{\mu^4}$	$\frac{6}{\mu^4}$
$N^-(\mu, \phi)$	$\frac{1}{\mu^3}$	$-\frac{3}{\mu^4}$	$\frac{12}{\mu^5}$	$\frac{6}{\mu^5}$	$\frac{15}{\mu^5}$

Nota: Em todos os modelos, $f = 0$.

Para modelos canônicos, para os quais $\eta = \theta$, temos que:

$$w = V = \frac{d\mu}{d\eta},$$

$$f = V \frac{dV}{d\mu} = \frac{d^2\mu}{d\eta^2},$$

$$g = 0,$$

$$j = -V \left(\frac{dV}{d\mu} \right)^2 - V^2 \frac{d^2V}{d\mu^2} = -\frac{d^3\mu}{d\eta^3},$$

$$h = b = V \left(\frac{dV}{d\mu} \right)^2 + V^2 \frac{d^2V}{d\mu^2} = \frac{d^3\mu}{d\eta^3},$$

Neste caso, somente o segundo termo de $\epsilon_{p,\beta}$ é zero e os A 's se simplificam muito pouco.

A Tabela 2.2 apresenta as funções w , g , j , h e b para modelos canônicos.

Tabela 2.2 Modelos canônicos - funções w, f, j, h e b

Função	$N(\mu, \phi^{-1})$	$P(\mu)$	$B(n, \mu)$	$G(\mu, \phi)$	$N^-(\mu, \phi)$
w	1	μ	$\mu(1 - \mu)$	μ^2	μ^3
f	0	μ	$\mu(1 - \mu)(1 - 2\mu)$	$2\mu^3$	$3\mu^5$
j	0	$-\mu$	$-\mu(1 - \mu)(1 + 6\mu^2 - 6\mu)$	$-6\mu^4$	$-15\mu^7$
h	0	μ	$\mu(1 - \mu)(1 + 6\mu^2 - 6\mu)$	$6\mu^4$	$15\mu^7$
b	0	μ	$\mu(1 - \mu)(1 + 6\mu^2 - 6\mu)$	$6\mu^4$	$15\mu^7$

Nota: Em todos os modelos, $g = 0$.

No caso de uma hipótese nula $H_0 : \beta = \beta^{(0)}$ ($q=p$), tem-se que $Z_2 = 0$ o que conduz a $A_{1,\beta} = 0$ enquanto os outros A 's reduzem-se a

$$A_{2,\beta} = 3\phi^{-1}1'BZ_d^{(2)}1, \quad A_{3,\beta} = \phi^{-1}1'(F - G)(3Z_dZ + 2Z^{(3)})(F - G)1$$

e $\epsilon_{p,\beta}$ é obtido de (2.2), enquanto que $\epsilon_{p-q,\beta} = 0$.

2.5 Modelo de regressão linear simples

Nesta seção, estamos interessados em um MLG definido por (1.1) com parte sistemática na forma linear simples

$$g(\mu_i) = \alpha + \beta x_i, \quad i = 1, \dots, n,$$

onde x_i denota o valor da variável explicativa para a i -ésima observação.

A estatística da razão de verossimilhança para testar $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$ é dada por

$$LR = 2\{L(\hat{\alpha}, \hat{\beta}) - L(\hat{\alpha}, 0)\}.$$

O fator de correção de Bartlett para este teste é obtido através de $\epsilon_{p,\beta}$ definido

em (2.1), sendo dado por $d = \epsilon_{2,\beta} - \epsilon_{1,\beta}$, onde os termos $\epsilon_{1,\beta}$ e $\epsilon_{2,\beta}$ são desenvolvidos através de (2.2) sob o modelo especificado por H_0 e o modelo irrestrito, respectivamente.

Como $\epsilon_{1,\beta}$ e $\epsilon_{2,\beta}$ são avaliados sob a hipótese nula, as matrizes diagonais W, F, G e J são proporcionais à matriz identidade I de ordem n , isto é, $W = wI$, $F = fI$, $G = gI$, e $J = jI$, com os escalares w, f, g e j definidos na Seção 2.2.1.

Sob H_0 , obtemos $X_2 = 1$ sendo $1 = (1, \dots, 1)'$ é um vetor de dimensão $n \times 1$ de 1's e a matriz Z_2 é dada por

$$Z_2 = \frac{1}{\sum_{i=1}^n w_i} 11',$$

o que implica em

$$\epsilon_{1,\beta} = \frac{1}{12n\phi w^3} (3wj - 4g(f + g) + 5f^2).$$

Definimos o a -ésimo momento central amostral da variável x por

$$s_a = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^a, \quad \text{para } a=2,3,4, \quad \text{com} \quad \bar{x} = \sum_{i=1}^n \frac{x_i}{n}.$$

Sob o modelo irrestrito, temos que $X = (1 \ x)$ com $x = (x_1, \dots, x_n)'$ e o elemento (ℓ, m) da matriz Z , avaliado no ponto $\beta = 0$, tem a forma

$$z_{\ell m} = \frac{1}{nws_2} (s_2 + (x_\ell - \bar{x})(x_m - \bar{x})).$$

Temos, então, que

$$\epsilon_{2,\beta} = \frac{1}{12n\phi w^3} (18jw - 16g(f + g) + 20f^2 + 3jw\gamma_{2x} + \{5f^2 - 4g(f + g)\}\gamma_{1x}^2),$$

onde $\gamma_{1x} = s_3/s_2^{3/2}$ e $\gamma_{2x} = (s_4 - 3s_2^2)/s_2^2$ são as medidas usuais de assimetria e curtose amostrais da covariada x .

O fator de correção de Bartlett para o teste da hipótese H_0 , no caso de ϕ ser conhecido, é dado por

$$d = \frac{1}{12n\phi w^3} (c_1 + c_2\gamma_{2x} + c_3\gamma_{1x}^2),$$

onde $c_1 = 3(5jw - 4g(f + g) + 5f^2)$, $c_2 = 3jw$ e $c_3 = 5f^2 - 4g(f + g)$.

A equação acima sugere, claramente, que a aproximação da distribuição de LR pela distribuição χ_1^2 é particularmente sensível a mudanças nas medidas amostrais de assimetria e curtose de x .

A estatística de escore é dada por

$$S_R = \frac{\phi}{\tilde{V}} \frac{\{\sum_{i=1}^n x_i(y_i - \bar{y})\}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{n\phi s^2}{\tilde{V} s_2}, \quad (2.10)$$

onde $s = n^{-1}(\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y})$ e $\tilde{V} = V(\bar{y})$.

Os A 's seguem imediatamente das equações (2.4)-(2.6) como funções de n, ϕ e μ . Os A 's não dependem da ligação e têm formas bastante simples dadas por

$$\begin{aligned} A_{1,\beta} &= -\frac{6}{n\phi} \left(\frac{d^2 V}{d\mu^2} \right), \\ A_{2,\beta} &= -\frac{3}{n\phi V} \left[3 \left(\frac{dV}{d\mu} \right)^2 - \left\{ \left(\frac{dV}{d\mu} \right)^2 + V \frac{d^2 V}{d\mu^2} \right\} \frac{s_4}{s_2^2} \right] \\ A_{3,\beta} &= \frac{5}{n\phi V} \left(\frac{dV}{d\mu} \right)^2 \frac{s_3^2}{s_2^3}, \end{aligned}$$

sendo que as quantidades que dependem do parâmetro β devem ser avaliadas sob a hipótese H_0 .

Denotamos por γ_{1y} e γ_{2y} as medidas usuais de assimetria e curtose da distribuição de y , isto é, o terceiro e quarto cumulantes padronizados da distribuição de y . Os A 's podem ser escritos explicitamente em termos destas quantidades e das medidas de assimetria e curtose amostrais da covariada x da seguinte forma:

$$A_{1,\beta} = \frac{6}{n}(\gamma_{1y}^2 - \gamma_{2y}), \quad A_{2,\beta} = \frac{3}{n}\{\gamma_{2y}(\gamma_{2x} + 3) - 3\gamma_{1y}^2\} \quad \text{e} \quad A_{3,\beta} = \frac{5}{n}\gamma_{1y}^2\gamma_{1x}^2.$$

É claro que, se y é normalmente distribuído, $\gamma_{1y} = \gamma_{2y} = 0$ e os A 's são iguais a zero. Entretanto, se este não for o caso, a aproximação da estatística escore pela distribuição χ_1^2 é particularmente sensível a mudanças nas medidas de amostrais de assimetria e curtose de x .

Exemplo 2.1

Para apresentar uma ilustração numérica de testes da razão de verossimilhança e escore aperfeiçoados, consideremos os dados da Tabela 2.3 analisados por Cox e Snell (1981, pgs. 148-150) relativos a tempos de sobrevivência y de 17 pacientes sofrendo de leucemia. Eles consideraram o modelo de regressão exponencial com estrutura log-linear $\log \mu_i = \alpha + \beta x_i$, onde μ_i é o tempo médio de sobrevivência desde o diagnóstico e x_i é o logaritmo na base 10 do número de glóbulos brancos no sangue. Estamos interessados em testar a hipótese nula $H_0 : \beta = 0$ contra a hipótese alternativa $H_0 : \beta \neq 0$. Cordeiro, Ferrari e Paula (1993) realizaram este teste de hipóteses, ao nível de significância de 1%, utilizando a estatística escore e sua versão modificada S_R^* .

Tabela 2.3 *Tempo de sobrevivência em semanas e $\log_{10}$ do número de glóbulos brancos para 17 pacientes com leucemia*

x	y	x	y	x	y
3,36	65	4,00	121	4,54	22
2,88	156	4,23	4	5,00	1
3,63	100	3,73	39	5,00	1
3,41	134	3,85	143	4,72	5
3,78	16	3,97	56	5,00	65
4,02	108	4,51	26		

Para este teste, obtemos $S_R = 5,655$, $S_R^* = 5,913$, $S_{R1}^* = 5,925$, $S_{R2}^* = 5,919$, $S_{R3}^* = 5,920$, $LR = 6,826$, $LR^* = 6,743$ e $LR_1^* = 6,742$ que conduzem aos respectivos níveis descritivos 1,74%, 1,50%, 1,49%, 1,50%, 1,50%, 0,89%, 0,94%

e 0,94%. Como se pode notar, para este conjunto de dados as correções não têm grande efeito.

Para avaliar o efeito das correções em amostras pequenas, consideramos agora 10 pacientes sofrendo leucemia, selecionados entre os 17 pacientes, através de uma amostragem aleatória sem reposição. Os pacientes selecionados foram: 12, 11, 4, 5, 15, 13, 7, 9, 14 e 16. Nesta sub-amostra, obtemos $S_R = 6,444$, $S_R^* = 7,375$, $S_{R1}^* = 7,533$, $S_{R2}^* = 7,446$, $S_{R3}^* = 7,453$, $LR = 12,741$, $LR^* = 12,345$ e $LR_1^* = 12,333$ que conduzem aos respectivos níveis descritivos 1,11%, 0,66%, 0,61%, 0,64%, 0,63%, 0,04%, 0,04% e 0,04%.

Podemos observar que a estatística da razão de verossimilhança e suas versões modificadas, levam à rejeição de H_0 ao nível de significância de 1%. No entanto, a estatística score não leva à rejeição de H_0 a este nível, embora tenha um nível descritivo em torno de 1%, enquanto que a rejeição de H_0 é obtida se o teste for baseado nas versões modificadas da estatística score. Observamos ainda que os efeitos das correções para a estatística score foram bem mais significativas nesta amostra ($n=10$) do que na amostra original ($n=17$). Por outro lado, em ambas as amostras, os efeitos das correções para a estatística da razão de verossimilhança foram pequenos.

2.6 Modelo de classificação a um fator

Consideremos $p \geq 2$ populações cada uma tendo densidade da forma (1.1) e suponhamos que amostras aleatórias independentes de tamanhos $n_1, \dots, n_p$ ($n_i \geq 1$, $i = 1, \dots, p$) são retiradas destas populações. Suponhamos, também, que ϕ é conhecido e igual para todas as populações. Assumimos a estrutura linear dada por

$$\eta_i = \beta + \beta_i, \quad i = 1, \dots, p$$

com $\beta_p = 0$, onde β é uma média geral e β_i é o efeito da i -ésima população na variável resposta, sendo η relacionado com $E(y) = \mu$ segundo alguma função de ligação. Denotamos as médias amostrais por $\bar{y}_1, \dots, \bar{y}_p$, o número total de observações por $n = \sum_{i=1}^p n_i$ e a média geral por $\bar{y} = \sum_{i=1}^p n_i \bar{y}_i / n$.

A estatística da razão de verossimilhança para testar a homogeneidade das médias das p populações, isto é, testar $H_0 : \beta_1 = \dots = \beta_{p-1} = 0$ contra a alternativa $H_1 : \beta_i$ não é constante, tem a seguinte forma:

$$LR = 2\{L(\hat{\beta}, \hat{\beta}_1, \dots, \hat{\beta}_{p-1}) - L(\tilde{\beta}, 0, \dots, 0)\} = 2\{L(\hat{\mu}_1, \dots, \hat{\mu}_p) - L(\tilde{\mu})\}.$$

Para obtermos o termo $\epsilon_{p,\beta}$ e os A 's precisamos encontrar as matrizes Z , Z_2 e $Z - Z_2$ que são funções das matrizes X , X_1 , X_2 e W , de dimensões $n \times p$, $n \times (p-1)$, $n \times 1$ e $n \times n$, respectivamente. Como $\epsilon_{1,\beta}$ e $\epsilon_{p,\beta}$ são avaliados sob H_0 , as matrizes diagonais W , F , G e J são proporcionais à matriz identidade de dimensão $n \times n$, ou seja, $W = w I_n$, $F = f I_n$, $G = g I_n$ e $J = j I_n$, onde o escalar w é definido na Seção 1.3 e os escalares f , g e j são definidos nas Seções 2.2 e 2.3. Para este caso, obtemos

$$\epsilon_{1,\beta} = \frac{1}{12n\phi} \left[\frac{2}{V} \left(\frac{dV}{d\mu} \right)^2 - 3 \frac{d^2V}{d\mu^2} \right].$$

As matrizes X , $X'WX$ e $X_2'WX_2$ para o modelo de classificação a um fator assumem as formas

$$X = \begin{pmatrix} 1_{n_1 \times 1} & 0_{n_1 \times 1} & \dots & 0_{n_1 \times 1} & 1_{n_1 \times 1} \\ 0_{n_2 \times 1} & 1_{n_2 \times 1} & \dots & 0_{n_2 \times 1} & 1_{n_2 \times 1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{n_{p-1} \times 1} & 0_{n_{p-1} \times 1} & \dots & 1_{n_{p-1} \times 1} & 1_{n_{p-1} \times 1} \\ 0_{n_p \times 1} & 0_{n_p \times 1} & \dots & 0_{n_p \times 1} & 1_{n_p \times 1} \end{pmatrix},$$

$$X'WX = \begin{pmatrix} n_1 & 0 & \dots & 0 & n_1 \\ 0 & n_2 & \dots & 0 & n_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & n_{p-1} & n_{p-1} \\ n_1 & n_2 & \dots & n_{p-1} & n \end{pmatrix}$$

e $X_2'WX_2 = nw$. A inversa de $X'WX$ segue da fórmula da inversa de uma matriz dividida (Rao, 1973, pg. 33). Obtemos $Z_2 = (nw)^{-1}$ e

$$Z - Z_2 = \begin{pmatrix} (n - n_1)n_1^{-1}1_{n_1 n_1} & -1_{n_1 n_2} & \dots & -1_{n_1 n_p} \\ -1_{n_2 n_1} & (n - n_2)n_2^{-1}1_{n_2 n_2} & \dots & -1_{n_2 n_p} \\ \vdots & \vdots & \ddots & \vdots \\ -1_{n_p n_1} & -1_{n_p n_2} & \dots & (n - n_p)n_p^{-1}1_{n_p n_p} \end{pmatrix}$$

onde 1_{qs} é uma matriz de $q \times s$ de 1's. A matriz de informação para este modelo, de posto p , tem a forma fechada $K = \phi X'WX = \text{diag}\{\phi n_i w_i\}$. A matriz de covariância assintótica Z tem dimensão $n \times n$ e elemento típico $\delta_{ij}(n_i \phi w_i)^{-1}$ onde $\delta_{ij} = 1$ se as observações i e j pertencem à mesma população e zero caso contrário. Pode-se deduzir da fórmula da matriz Z ,

$$\begin{aligned} Z^{(3)} &= \text{diag}\{\delta_{ij} n_i^{-3} \phi^{-1} w_i^{-3}\}, \\ \text{tr}(JZ_d^2) &= \sum_{i=1}^p n_i^{-1} \phi^{-1} j_i w_i^{-2}, \\ 1' F Z^{(3)} F 1 &= \sum_{i=1}^p n_i^{-1} \phi^{-1} f_i^2 w_i^{-3} e \\ 1' G Z^{(3)} (F + G) 1' &= \sum_{i=1}^p n_i^{-1} \phi^{-1} g_i (f_i + g_i) w_i^{-3}. \end{aligned}$$

Obtemos, então

$$\epsilon_{p,\beta} = \frac{1}{12\phi} \sum_{i=1}^p \frac{1}{n_i} \left[\frac{2}{V} \left(\frac{dV}{d\mu} \right)^2 - 3 \frac{d^2 V}{d\mu^2} \right]_i.$$

Observe que tanto a estatística da razão de verossimilhança LR , para testar H_0 , quanto as expressões de $\epsilon_{1,\beta}$ e $\epsilon_{p,\beta}$ não dependem da ligação. O fator de correção de Bartlett para o teste da hipótese H_1 no caso de ϕ ser conhecido, é dado por

$$d = \frac{1}{12\phi(p-1)} \left[\frac{2}{V} \left(\frac{dV}{d\mu} \right)^2 - 3 \frac{d^2 V}{d\mu^2} \right] \left(\sum_{i=1}^p \frac{1}{n_i} - \frac{1}{n} \right).$$

A estatística score é dada por

$$S_R = \frac{\phi}{\hat{V}} \sum_{i=1}^p n_i (\bar{y}_i - \bar{y})^2,$$

onde $\hat{V}$ é o estimador de máxima verossimilhança de V sob H_0 . Assintoticamente, S_R tem distribuição χ_{p-1}^2 sob a hipótese nula.

De (2.4)-(2.6) e das matrizes Z e $Z - Z_2$ acima, temos as seguintes expressões para os A 's

$$A_{1,\beta} = -6 \frac{(p-1)}{n\phi} \frac{d^2 V}{d\mu^2},$$

$$A_{2,\beta} = \frac{3}{n\phi} \left[n \left\{ \frac{1}{V} \left(\frac{dV}{d\mu} \right)^2 + \frac{d^2V}{d\mu^2} \right\} \sum_{i=1}^p n_i^{-1} - (p^2 + 2p - 2) \frac{1}{V} \left(\frac{dV}{d\mu} \right)^2 - (2p - 1) \left(\frac{d^2V}{d\mu^2} \right) \right],$$

e

$$A_{3,\beta} = \frac{1}{n\phi} \frac{1}{V} \left(\frac{dV}{d\mu} \right)^2 \left\{ 5n \sum_{i=1}^p \frac{1}{n_i} - (3p^2 + 6p - 4) \right\}.$$

É conveniente observar que S_R e os A 's não dependem da ligação.

Apresentamos na Tabela 2.4, $A_{1,\beta}$, $A_{2,\beta}$, $A_{3,\beta}$ e $\epsilon_{p,\beta}$ para o modelo de classificação a um fator considerado com 2 populações, das quais foram retiradas amostras de tamanhos $n_1 = n_2 = n/2$. Para todos os modelos supomos que o parâmetro de dispersão é conhecido e igual para as duas populações.

Tabela 2.4 Quantidades $A_{1,\beta}$, $A_{2,\beta}$, $A_{3,\beta}$ e $\epsilon_{p,\beta}$ para um modelo de classificação a um fator com 2 populações

Modelo	$A_{1,\beta}$	$A_{2,\beta}$	$\epsilon_{p,\beta}$
$N(\mu, \phi^{-1})$	0	0	0
$P(\mu)$	0	$-\frac{6}{n\mu}$	$\sum_{i=1}^2 \frac{n_i^2 - 1}{6n_i\mu}$
$B(n, \mu)$	$\frac{12}{n}$	$-\frac{6}{n\mu(1-\mu)} + \frac{18}{n}$	$\sum_{i=1}^2 \frac{1 - \mu(1-\mu)}{6n_i\mu(1-\mu)}$
$G(\mu, \phi)$	$-\frac{12}{n\phi}$	$-\frac{18}{n\phi}$	$\sum_{i=1}^2 \frac{1}{6n_i\phi}$
$N^-(\mu, \phi)$	$-\frac{36\mu}{n\phi}$	$-\frac{36\mu}{n\phi}$	0

Nota: Em todos os modelos, $A_{3,\beta} = 0$.

Exemplo 2.2

A Tabela 2.5 (Lawless 1982, pg. 201) mostra os resultados de um experimento conduzido para avaliar o desempenho de cinco tipos de turbina ($p=5$) de alta velocidade. Foram considerados dez motores de cada tipo na análise e foi observado para cada um o tempo (em milhões de ciclos) até a perda da velocidade. Chhikara e Folks (1989, pg. 94) consideraram o modelo normal inverso, para analisar os dados da Tabela 2.5. Estamos interessados em testar a hipótese nula $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ contra a hipótese alternativa $H_1 : \beta_i$ não é constante. Para analisar estes dados, consideramos o modelo normal inverso com ligação canônica e ϕ desconhecido.

Tabela 2.5 *Tempo até a perda de velocidade de 5 tipos de motores*

Tipo I	Tipo II	Tipo III	Tipo IV	Tipo V
3,03	3,19	3,46	5,88	6,43
5,53	4,26	5,22	6,74	9,97
5,60	4,47	5,69	6,90	10,39
9,30	4,53	6,54	6,98	13,55
9,92	4,67	9,16	7,21	14,45
12,51	4,69	9,40	8,14	14,72
12,95	5,78	10,19	8,59	16,81
15,21	6,79	10,71	9,80	18,39
16,04	9,37	12,58	12,28	20,84
16,84	12,75	13,41	25,46	21,51

Para este teste, obtemos $S_R = 13,241$, $S_R^* = 13,945$, $S_{R1}^* = 13,984$, $S_{R2}^* = 13,964$, $S_{R3}^* = 13,983$, $LR = 17,050$, $LR^* = 15,787$ e $LR_1^* = 15,686$ que conduzem aos respectivos níveis descritivos de 1,01%, 0,75%, 0,74%, 0,74%, 0,74%, 0,19%, 0,33% e 0,35%.

Podemos observar que para a estatística da razão de verossimilhança e suas versões modificadas levam a rejeição de H_0 ao nível de significância de 1%. Por outro lado, a estatística escore não leva à rejeição de H_0 a este nível, embora tenha um nível descritivo próximo de 1%, enquanto que a rejeição de H_0 é obtida se o teste for baseado nas versões modificadas da estatística escore.

Para avaliar o efeito das correções em amostras pequenas, tomamos uma sub-amostra da amostra original, considerando 5 observações tomadas ao acaso, e sem reposição, de cada tipo de motor. Os motores selecionados foram, para o tipo I : 6, 7, 4, 3 e 1, tipo II: 1, 10, 3, 2 e 7, tipo III: 5, 6, 8, 4 e 2, tipo IV: 4, 7, 9, 10 e 6 e para o tipo V: 2, 3, 7, 1 e 8. Nesta sub-amostra obtemos $S_R = 12,525$, $S_R^* = 13,754$, $S_{R1}^* = 13,888$, $S_{R2}^* = 13,817$, $S_{R3}^* = 13,911$, $LR = 18,369$, $LR^* = 15,836$ e $LR_1^* = 15,430$ que conduzem aos respectivos níveis descritivos de 1,39%, 0,81%, 0,77%, 0,79%, 0,76%, 0,11%, 0,33% e 0,39%.

Podemos observar que os efeitos das correções foram bem mais significativos nesta amostra ($n_i=5$) do que na amostra original ($n_i=10$). É interessante notar também que, em ambas as amostras, a diferença entre os valores das estatísticas escore e da razão de verossimilhança (ambas sem correção) é bem maior do que as diferenças entre os valores das estatísticas corrigidas. Observamos ainda que os valores das estatísticas corrigidas na amostra original são muito próximos dos valores das correspondentes estatísticas na sub-amostra.

2.7 Modelo de regressão com dois parâmetros

Consideremos um MLG com componente sistemática $\eta_i = \alpha u_i + \beta x_i$, $i = 1, \dots, n$, onde u_i e x_i denotam os valores de duas variáveis auxiliares associadas à i -ésima observação. A estatística da razão de verossimilhança para testar $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$ é dada por

$$LR = 2\{L(\hat{\alpha}, \hat{\beta}) - L(\tilde{\alpha}, 0)\}.$$

O fator de correção de Bartlett para este teste definido em (2.1) é obtido através de $\epsilon_{p,\beta}$ que é calculado em (2.2), observando que o elemento (ℓ, m) da matriz Z é

$$z_{\ell m} = \Delta^{-1} \left[u_{\ell} u_m \left(\sum_{\ell=1}^n w_{\ell} x_{\ell}^2 \right) + x_{\ell} x_m \left(\sum_{\ell=1}^n w_{\ell} u_{\ell}^2 \right) - (x_{\ell} u_m + x_m u_{\ell}) \sum_{\ell=1}^n w_{\ell} u_{\ell} x_{\ell} \right],$$

onde

$$\Delta = \left(\sum_{\ell=1}^n w_{\ell} u_{\ell}^2 \right) \left(\sum_{\ell=1}^n w_{\ell} x_{\ell}^2 \right) - \left(\sum_{\ell=1}^n w_{\ell} u_{\ell} x_{\ell} \right)^2.$$

A estatística score para testar $H_0 : \beta = \beta^{(0)}$ contra $\beta \neq \beta^{(0)}$ pode ser obtida de (2.10) observando que

$$\begin{aligned} r_{\ell} &= x_{\ell} - u_{\ell} \left(\sum_{\ell=1}^n w_{\ell} x_{\ell} u_{\ell} \right) \left(\sum_{\ell=1}^n w_{\ell} u_{\ell}^2 \right)^{-1} \\ \sum_{\ell=1}^n w_{\ell} r_{\ell}^2 &= \left(\sum_{\ell=1}^n w_{\ell} u_{\ell}^2 \right)^{-1} \left\{ \left(\sum_{\ell=1}^n w_{\ell} x_{\ell}^2 \right) \left(\sum_{\ell=1}^n w_{\ell} u_{\ell}^2 \right) - \left(\sum_{\ell=1}^n w_{\ell} x_{\ell} u_{\ell} \right)^2 \right\}, \end{aligned}$$

conduzindo a

$$S_R = \phi \frac{\left(\sum_{\ell=1}^n \tilde{w}_{\ell} u_{\ell}^2 \right) \left\{ (\tilde{w}_{\ell} / \tilde{V}_{\ell})^{1/2} x_{\ell} (y_{\ell} - \tilde{\mu}_{\ell}) \right\}^2}{\left(\sum_{\ell=1}^n \tilde{w}_{\ell} x_{\ell}^2 \right) \left(\sum_{\ell=1}^n \tilde{w}_{\ell} u_{\ell}^2 \right) \left(\sum_{\ell=1}^n \tilde{w}_{\ell} x_{\ell} u_{\ell} \right)^2}.$$

Temos que o elemento (ℓ, m) da matriz Z_2 é $u_{\ell} u_m / \sum_{\ell=1}^n w_{\ell} u_{\ell}^2$. Substituindo a expressão de $z_{2\ell m}$ chega-se a

$$\begin{aligned}
 A_{1,\beta} &= \frac{1}{\phi \left(\sum_{\ell=1}^n w_{\ell} r_{\ell}^2 \right) \left(\sum_{\ell=1}^n w_{\ell} u_{\ell}^2 \right)^2} \left[3 \left(\sum_{\ell=1}^n f_{\ell} u_{\ell}^2 r_{\ell} \right)^2 + 6 \left(\sum_{\ell=1}^n f_{\ell} u_{\ell}^3 \right) \left\{ \sum_{\ell=1}^n (f - g)_{\ell} u_{\ell} r_{\ell}^2 \right\} \right. \\
 &\quad \left. - 6 \left(\sum_{\ell=1}^n f_{\ell} u_{\ell}^2 r_{\ell} \right) \left(\sum_{\ell=1}^n (2g - f)_{\ell} u_{\ell}^2 r_{\ell} \right) - 6 \left(\sum_{\ell=1}^n h_{\ell} u_{\ell}^2 r_{\ell}^2 \right) \left(\sum_{\ell=1}^n w_{\ell} u_{\ell}^2 \right) \right], \\
 A_{2,\beta} &= \frac{1}{\phi \left(\sum_{\ell=1}^n w_{\ell} r_{\ell}^2 \right)^2 \sum_{\ell=1}^n w_{\ell} u_{\ell}^2} \left[-9 \left(\sum_{\ell=1}^n (f - g)_{\ell} u_{\ell} r_{\ell}^2 \right)^2 - 6 \left(\sum_{\ell=1}^n f_{\ell} u_{\ell}^2 r_{\ell} \right) \left\{ \sum_{\ell=1}^n (f - g)_{\ell} r_{\ell}^3 \right\} \right. \\
 &\quad \left. + 3 \left(\sum_{\ell=1}^n b_{\ell} r_{\ell}^4 \right) \left(\sum_{\ell=1}^n w_{\ell} u_{\ell}^2 \right) \right], \\
 A_{3,\beta} &= \frac{5}{\phi \left(\sum_{\ell=1}^n w_{\ell} r_{\ell}^2 \right)^3} \left(\sum_{\ell=1}^n (f - g)_{\ell} r_{\ell}^3 \right)^2.
 \end{aligned}$$

Exemplo 2.3

Na Tabela 2.6 são apresentadas as siglas de 40 estados norte-americanos juntamente com as seguintes variáveis: *taxa* (taxa do combustível no estado), *licença* (porcentagem de motoristas licenciados), *renda* (renda per capita), *estrada* (ajuda federal para as estradas) e *consumo* (consumo de combustível por habitante). As observações foram obtidas a partir de uma seleção aleatória dos dados de Gray (1989).

Para analisar estes dados, suponha que o consumo segue uma distribuição normal de variância constante σ^2 e média $\mu_{\ell} = \alpha + \beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \beta_3 x_{\ell 3} + \beta_4 x_{\ell 4}$, para $\ell = 1, \dots, 40$. Estamos interessados em testar a seguinte hipótese nula $H_0 : \beta_1 = \beta_3 = 0$ e consideramos ϕ desconhecido.

Tabela 2.6 *Informações sobre taxa, licença, renda e estrada obtidas de 40 estados norte-americanos*

<i>Sigla</i>	x_1	x_2	x_3	x_4	y	<i>Sigla</i>	x_1	x_2	x_3	x_4	y
VT	9,00	0,580	3865	1586	561	MA	7,50	0,529	4870	2351	414
RI	8,00	0,544	4399	431	410	CT	10,00	0,571	5342	1333	457
NY	8,00	0,451	5319	11868	344	NJ	8,00	0,553	5126	2138	467
IN	8,00	0,530	4391	5939	580	IL	7,50	0,525	5126	14186	471
MI	7,00	0,574	4817	6930	525	WI	7,00	0,545	4207	6580	508
MN	7,00	0,608	4332	8159	566	IA	7,00	0,586	4318	10340	635
MO	7,00	0,572	4206	8508	603	ND	7,00	0,540	3718	4725	714
KS	7,00	0,663	4593	7834	649	DE	8,00	0,602	4983	602	540
MD	9,00	0,511	4897	2449	464	VA	9,00	0,517	4258	4686	547
WV	8,50	0,551	4574	2619	460	NC	9,00	0,544	3721	4746	566
SC	8,00	0,548	3448	5399	577	GA	7,50	0,579	3846	9061	631
FL	8,00	0,563	4188	5975	574	KY	9,00	0,493	3601	4650	534
MS	8,00	0,578	3063	6524	577	AR	7,50	0,547	3357	4121	628
LA	8,00	0,487	3528	3495	487	OK	6,58	0,629	3802	7834	644
TX	5,00	0,566	4045	17782	640	MT	7,00	0,586	3897	6385	704
ID	8,50	0,663	3635	3274	648	WY	7,00	0,672	4345	3905	968
CO	7,00	0,626	4449	4639	587	NM	7,00	0,563	3656	3985	699
AZ	7,00	0,603	4300	3635	632	UT	7,00	0,508	3745	2611	591
NV	6,00	0,672	5215	2302	782	WA	9,00	0,571	4476	3942	510
OR	7,00	0,623	4296	4083	610	CA	7,00	0,593	5002	9794	524

Nota: x_1 , x_2 , x_3 , x_4 e y correspondem, respectivamente, a *Taxa*, *Licença*, *Renda*, *Estrada* e *Consumo*.

Para este teste, obtemos $S_R = 5,724$, $S_R^* = 5,418$, $S_{R1}^* = 5,434$, $S_{R2}^* = 5,426$, $S_{R3}^* = 5,431$, $LR = 6,177$, $LR^* = 5,491$ e $LR_1^* = 5,405$ que conduzem aos respectivos níveis descritivos de 5,72%, 6,66%, 6,61%, 6,63%, 6,62%, 6,42%, 4,56%, 6,42% e 6,70%.

Podemos observar que a estatística score e suas versões modificadas, não levam à rejeição de H_0 ao nível de significância de 5%. No entanto, a estatística da razão de verossimilhança leva à rejeição de H_0 a este nível, enquanto que a não rejeição

de H_0 também é obtida se o teste for baseado nas versões modificadas da estatística da razão de verossimilhança.

Como estamos ajustando o modelo normal, calculamos o teste exato F que conduz ao valor 2,923. O nível descritivo obtido da distribuição $F_{2,35}$ vale 6,70% indicando que H_0 não é rejeitada ao nível de significância de 5%, que concorda com a decisão dos testes modificados da razão de verossimilhança e escore. Notamos ainda que os níveis descritivos dos testes corrigidos são bem próximos do nível descritivo do teste exato. O mesmo não ocorre com os níveis dos testes não corrigidos.

CAPÍTULO 3

Simulações em modelos contínuos

Neste capítulo, apresentamos alguns resultados de simulação para avaliar a eficácia das correções de Bartlett e tipo-Bartlett para testes da razão de verossimilhança e escore em MLGs contínuos.

Comparamos os desempenhos dos nove testes, isto é, o da razão de verossimilhança e escore com suas respectivas versões modificadas. Estes desempenhos são avaliados em função da proximidade das probabilidades de rejeição da hipótese nula, sendo esta verdadeira (probabilidade do erro tipo I), aos respectivos níveis nominais dos testes. As simulações obtidas foram feitas para as distribuições normal e gama para o caso de ϕ ser desconhecido.

Para todas as simulações o preditor linear é

$$\eta_i = \beta_0 + \sum_{j=1}^9 \beta_j x_{ij}, \quad i = 1, \dots, n$$

e a hipótese nula considerada é $H_0 : \beta_5 = \beta_6 = 0$ ($q=2$). A variável resposta foi gerada assumindo que $\beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_7 = \beta_8 = \beta_9 = 0,05$.

Dez mil amostras foram geradas para diversos tamanhos de amostra ($n=20, 30, 40$ e 50). O valor assumido para ϕ no modelo normal foi $1,0$ e para o modelo gama os índices $\phi=1,0$ e $\phi=2,0$ foram considerados.

Um cuidado foi tomado na escolha dos valores das covariadas de modo que as medidas amostrais de assimetria e curtose das covariadas fossem estabilizadas para diferentes tamanhos de amostra. Para $n=10$, as nove covariadas foram tomadas como amostras aleatórias das seguintes distribuições: $U(0,1)$, $F(2,5)$, Cauchy, $N(0,1)$, t_3 , $LN(0,1)$, χ_3^2 , $F(3,3)$ e $N(0,2)$. Para outros tamanhos de amostra, as covariadas foram replicadas. Ou seja, se para $n=10$, os valores de uma covariada são $x_1, x_2, \dots, x_{10}$, então, para $n=20$ temos $x_1, x_1, x_2, x_2, x_3, x_3, \dots, x_{10}, x_{10}$ e similar-

mente para $n=30, 40$ e 50 . Os valores das covariadas foram mantidos fixos para as 10.000 amostras.

Na Tabela 3.1 apresentamos os resultados de simulação obtidos supondo que a variável resposta tem distribuição normal com função de ligação identidade. Para cada tamanho da amostra e cada nível considerado, calculamos as taxas de rejeição de cada teste, isto é, estimamos, via simulação $P(LR \geq x_\alpha)$, $P(LR^* \geq x_\alpha)$, $P(LR_1^* \geq x_\alpha)$, $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$, $P(S_{R1}^* \geq x_\alpha)$, $P(S_{R2}^* \geq x_\alpha)$, $P(S_{R3}^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$, onde z_α é dado na Seção 2.3. Vale recordar que $LR^* = LR/(1 + d)$, $LR_1^* = LR(1 - d)$, $S_R^* = S_R(1 - B)$, $S_{R1}^* = S_R/(1 + B)$, $S_{R2}^* = S_R \exp(-B)$ onde $B = c + bS_R + aS_R^2$ e S_{R3}^* é a estatística modificada monótona em S_R dada em (2.9).

A análise desta tabela mostra que tanto a estatística da razão de verossimilhança (LR) quanto a estatística escore (S_R) não têm uma boa aproximação por χ^2 e apresentam taxas de rejeição bem maiores do que os níveis nominais mesmo quando o número total de observações é razoavelmente grande ($n=50$). Além disto, existe forte evidência de que a aproximação por χ^2 para os testes modificados é melhor do que para os testes da razão de verossimilhança e escore usuais.

Existe também evidência de que, entre os testes da razão de verossimilhança modificados, o teste baseado em LR_1^* tem melhor desempenho do que aquele baseado em LR^* . Já, entre os testes escore modificados, o teste baseados em S_R^* é o que apresentou taxas de rejeição mais próximas dos níveis nominais.

Claramente, conforme cresce o tamanho da amostra, as taxas de rejeição de todos os testes vão se aproximando dos respectivos níveis nominais e como era de se esperar, as correções vão sendo cada vez menos necessárias.

Outro estudo de simulação foi realizado com o objetivo de analisar a influência do número de parâmetros de perturbação no desempenho dos testes. Consideramos nove modelos, a saber:

$$i) \quad \eta = \beta_5 x_5 + \beta_6 x_6 \quad (p - q = 0)$$

- ii) $\eta = \beta_0 + \beta_5 x_5 + \beta_6 x_6 \quad (p - q = 1)$
- iii) $\eta = \beta_0 + \beta_1 x_1 + \beta_5 x_5 + \beta_6 x_6 \quad (p - q = 2)$
- iv) $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_5 x_5 + \beta_6 x_6 \quad (p - q = 3)$
- $\vdots$
- ix) $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \cdots + \beta_9 x_9 \quad (p - q = 8).$

A hipótese nula a ser testada nos nove modelos é $H_0 : \beta_5 = \beta_6 = 0$. Dez mil amostras foram geradas para o tamanho de amostra $n=30$. Os resultados destas simulações são apresentados na Tabela 3.2 e nas Figuras 3.1 a 3.3 para o modelo normal com ligação identidade.

Em todas as figuras, $S_R(*)$ indica o teste baseado na estatística S_R e nos quantis modificados. Analisando as figuras e a Tabela 3.2, podemos observar claramente que o desempenho do teste score é bem melhor do que o da razão de verossimilhança (considerando ambos sem correção). Notamos ainda que o número de parâmetros de perturbação ($p-q+1$) tem um impacto considerável sobre a aproximação por χ^2 para as distribuições destas duas estatísticas. Para $p-q=8$, por exemplo, ambos os testes são muito anti-conservativos apresentando taxas de rejeição bem maiores do que os níveis nominais. Este fato é mais marcante no caso da estatística LR do que no caso da estatística score. É interessante notar que o impacto do número de parâmetros de perturbação é bem menos marcante no caso dos testes corrigidos. Para estes testes, as taxas de rejeição permanecem mais estáveis em relação aos respectivos níveis nominais do que as taxas dos testes não corrigidos. Em poucas palavras, podemos dizer que, na grande maioria dos casos, as estatísticas modificadas produzem melhores testes do que as não modificadas. Entre as estatísticas da razão de verossimilhança modificadas, aquela que produz melhores testes é LR_1^* . Já, entre as estatísticas score modificadas, uma ordenação entre as diversas estatísticas não é tão evidente mas parece que a estatística S_R^* tende a ter um melhor desempenho do que as demais.

Tabela 3.1 *Tamanhos dos testes - $p=10$ - modelo normal*

n	$tamanhos$ $nominais(\%)$	LR			S_R					
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
20	1,0	9,5	3,0	1,0	4,3	1,1	1,5	1,2	1,3	1,1
	5,0	21,6	10,1	4,7	15,9	5,9	7,8	6,9	7,1	6,7
	10,0	30,7	17,2	9,5	25,8	11,4	14,5	13,7	13,7	13,6
30	1,0	5,0	1,6	1,0	2,6	1,1	1,3	1,2	1,3	1,1
	5,0	13,8	7,6	5,4	11,1	6,1	7,0	6,6	6,6	6,6
	10,0	22,0	13,0	10,4	19,1	11,0	12,4	11,8	11,8	12,0
40	1,0	3,5	1,5	1,1	2,2	1,1	1,3	1,2	1,3	1,1
	5,0	10,8	6,2	5,3	8,9	5,5	6,0	5,8	5,8	5,8
	10,0	18,0	11,7	10,1	16,2	10,4	11,3	11,0	11,0	11,1
50	1,0	2,7	1,2	1,1	1,7	1,1	1,2	1,1	1,2	1,1
	5,0	9,4	5,8	5,2	8,1	5,3	5,6	5,5	5,5	5,5
	10,0	16,2	11,4	10,4	14,7	10,6	11,1	10,9	10,9	10,9

Nota: Os números (1), (2), (3), (4), (5), (6), (7), (8) e (9) correspondem a $P(LR \geq x_\alpha)$, $P(LR^* \geq x_\alpha)$, $P(LR_1^* \geq x_\alpha)$, $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$, $P(S_{R1}^* \geq x_\alpha)$, $P(S_{R2}^* \geq x_\alpha)$, $P(S_{R3}^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ respectivamente.

Tabela 3.2 *Tamanhos dos testes - $n=30$ - modelo normal*

$p - q$	tamanhos nominais(%)	LR					S_R			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0	1,0	1,3	1,0	1,0	0,7	0,9	0,9	0,9	0,9	0,6
1	1,0	1,5	1,0	1,0	0,7	0,9	0,9	0,9	0,9	0,9
2	1,0	1,8	1,1	1,0	0,8	0,9	0,9	0,9	0,9	0,9
3	1,0	2,1	1,1	1,0	2,6	1,1	1,3	1,2	1,3	1,1
4	1,0	2,3	1,0	1,0	1,1	1,0	0,9	0,9	0,9	0,9
5	1,0	2,6	1,2	1,0	1,3	1,0	1,0	1,0	1,0	0,9
6	1,0	3,4	1,4	1,0	1,8	1,0	1,0	1,0	1,0	1,0
7	1,0	3,7	1,6	1,0	2,3	1,1	1,1	1,1	1,1	1,0
8	1,0	5,0	1,6	1,0	2,6	1,1	1,3	1,2	1,3	1,1
0	5,0	5,9	4,8	4,8	4,8	4,8	4,8	4,7	4,8	3,8
1	5,0	6,8	4,9	4,8	4,8	4,8	4,8	4,8	4,8	4,8
2	5,0	7,4	5,0	4,7	5,3	4,7	4,7	4,7	4,8	4,7
3	5,0	7,7	5,0	4,7	5,8	4,7	4,7	4,7	4,7	4,7
4	5,0	8,7	5,4	4,7	6,5	4,8	5,0	4,9	4,9	4,8
5	5,0	10,0	5,6	4,7	7,5	4,9	5,1	5,0	5,1	5,0
6	5,0	11,0	6,1	5,0	8,5	5,3	5,6	5,4	5,5	5,4
7	5,0	12,0	6,2	4,6	9,3	5,4	5,8	5,7	5,7	5,7
8	5,0	14,0	7,6	5,4	11,1	6,1	7,0	6,6	6,6	6,6
0	10,0	11,6	10,0	10,0	10,6	10,0	10,0	10,2	10,2	0,8
1	10,0	12,3	10,0	10,0	10,6	10,0	10,0	10,0	10,0	10,0
2	10,0	13,5	10,0	10,1	11,5	10,1	10,2	10,2	10,2	10,1
3	10,0	14,1	10,0	9,4	11,9	10,0	10,0	10,0	10,0	10,0
4	10,0	15,3	10,1	9,6	12,8	10,0	10,0	10,0	10,0	10,0
5	10,0	16,4	10,1	9,9	14,4	10,2	10,7	10,5	10,5	10,5
6	10,0	18,3	10,1	9,7	15,9	10,2	10,9	10,7	10,7	10,7
7	10,0	19,5	12,0	10,0	17,3	10,4	11,4	10,9	10,9	11,0
8	10,0	22,0	13,0	10,4	19,1	11,0	12,4	11,7	11,8	12,0

Nota: Os números (1), (2), (3), (4), (5), (6), (7), (8) e (9) correspondem a $P(LR \geq x_\alpha)$, $P(LR^* \geq x_\alpha)$, $P(LR_1^* \geq x_\alpha)$, $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$, $P(S_{R1}^* \geq x_\alpha)$, $P(S_{R2}^* \geq x_\alpha)$, $P(S_{R3}^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ respectivamente.

Figura 3.1 : Tamanhos dos Testes - $n=30$
Modelo Normal - $\alpha=1\%$

Figura 3.2 : Tamanhos dos Testes - $n=30$
Modelo Normal - $\alpha=5\%$

Figura 3.3 : Tamanhos dos Testes - $n=30$
Modelo Normal - $\alpha=10\%$

Os resultados de simulação referentes ao modelo gama com ligação recíproca ($\mu = 1/\eta$), são apresentados nas Tabelas 3.3 a 3.6 e nas Figuras 3.4 a 3.9. É importante notar que o caso (i), isto é, aquele em que $\eta = \beta_5 x_5 + \beta_6 x_6$ não foi considerado pois, sob $H_0 : \beta_5 = \beta_6 = 0$, temos $\eta = 0$ e não podemos calcular $\mu = 1/\eta$.

Das Tabelas 3.3 e 3.5 podemos observar que, na grande maioria dos casos, as taxas de rejeição das estatísticas modificadas estão mais próximas dos correspondentes tamanhos nominais do que as taxas de rejeição das não modificadas. Além disso, as taxas de rejeição das estatísticas não modificadas são, em geral, superiores aos tamanhos nominais correspondentes. À medida que cresce o tamanho da amostra ou o valor de ϕ , as taxas de rejeição das estatísticas modificadas e não modificadas se aproximam dos correspondentes tamanhos nominais.

Analisando as Figuras 3.4 a 3.9 e as Tabelas 3.4 e 3.6 nota-se que, da mesma forma que para o modelo normal, o teste escore tem um desempenho bem melhor que o da razão de verossimilhança. É de se destacar também, no caso em que $\phi = 2, 0$, o excelente desempenho do teste baseado em LR_1^* , muito melhor que o daquele baseado em LR ou mesmo em LR^* . As versões corrigidas do teste escore, para $\phi = 2, 0$, também produzem, em geral, testes com taxas de rejeição bem mais próximas dos níveis nominais do que o teste escore usual especialmente se o número de parâmetros de perturbação é grande.

Para o caso em que $\phi = 1, 0$, novamente os testes da razão de verossimilhança modificados têm desempenho melhor do que o correspondente teste não modificado. No entanto, entre as diversas versões dos testes escore não modificados, em alguns casos, o teste $S_R(*)$, baseado em S_R e nos quantis modificados, apresenta desempenho bem pior do que o teste escore sem correção.

Tabela 3.3 *Tamanhos dos testes - $p=10$ - modelo gama com índice $\phi=2,0$*

n	$tamanhos$ $nominais(\%)$	LR			S_R					
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
20	1,0	11,9	3,2	1,0	4,1	1,2	1,5	1,4	1,4	1,6
	5,0	24,6	11,6	4,8	16,2	6,3	8,0	7,3	7,4	7,4
	10,0	34,0	18,8	9,8	26,4	12,3	15,5	14,2	14,3	14,6
30	1,0	5,6	1,7	1,0	2,1	1,1	1,1	1,1	1,1	1,0
	5,0	14,7	7,5	5,1	9,8	5,6	6,1	5,9	6,0	5,7
	10,0	23,1	13,2	9,7	18,4	10,5	11,6	11,1	11,2	11,0
40	1,0	3,6	1,4	1,0	1,5	1,0	1,0	1,0	1,0	1,0
	5,0	11,7	6,2	4,8	8,1	5,4	5,6	5,5	5,5	5,4
	10,0	18,5	12,1	9,7	15,5	10,3	11,0	10,7	10,7	10,6
50	1,0	2,8	1,2	1,0	1,3	1,0	1,0	1,0	1,0	1,0
	5,0	9,9	5,6	5,0	7,3	6,5	5,3	5,2	5,2	5,0
	10,0	17,0	11,5	10,1	14,5	10,5	10,9	10,7	10,8	10,7

Nota: Os números (1), (2), (3), (4), (5), (6), (7), (8) e (9) correspondem a $P(LR \geq x_\alpha)$, $P(LR^* \geq x_\alpha)$, $P(LR_1^* \geq x_\alpha)$, $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$, $P(S_{R1}^* \geq x_\alpha)$, $P(S_{R2}^* \geq x_\alpha)$, $P(S_{R3}^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ respectivamente.

Tabela 3.4 *Tamanhos dos testes - $n=30$ - modelo gama com índice $\phi=2,0$*

$p - q$	tamanhos nominais(%)	LR			S_R					
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	1,0	1,6	1,1	1,0	1,0	1,1	1,1	1,1	1,1	1,1
2	1,0	1,8	1,0	0,9	0,9	1,0	1,0	1,0	1,0	1,0
3	1,0	2,1	1,1	1,1	1,1	0,9	0,9	0,9	1,0	1,0
4	1,0	2,7	1,2	1,1	1,2	1,0	1,0	1,0	1,0	1,1
5	1,0	2,9	1,3	1,0	1,3	1,0	1,0	1,0	1,1	1,1
6	1,0	3,4	1,3	1,0	1,2	0,9	0,9	0,9	0,9	1,0
7	1,0	4,3	1,7	1,1	1,6	1,1	1,1	1,1	1,1	1,1
8	1,0	5,6	1,7	0,8	2,1	1,1	1,1	1,1	1,1	1,0
1	5,0	6,4	5,0	4,9	4,5	4,8	4,8	4,8	4,8	4,8
2	5,0	7,0	5,2	5,1	4,9	4,9	4,9	4,9	4,9	4,9
3	5,0	7,7	5,1	4,8	5,8	4,8	4,8	4,8	4,8	4,8
4	5,0	8,9	5,3	4,8	6,5	4,9	5,0	5,0	5,0	5,0
5	5,0	10,2	5,8	5,1	7,3	5,1	5,2	5,1	5,2	5,1
6	5,0	10,9	6,0	4,9	7,2	4,9	5,1	5,0	5,1	5,0
7	5,0	12,7	6,7	5,0	8,3	5,3	5,6	5,5	5,6	5,3
8	5,0	14,7	7,5	5,1	9,8	5,6	6,1	5,9	6,0	5,7
1	10,0	12,0	10,0	9,7	9,7	9,9	9,9	9,9	9,9	9,9
2	10,0	13,3	10,2	9,9	10,5	10,0	10,0	10,0	10,0	9,9
3	10,0	13,6	10,2	9,7	11,8	10,0	10,1	10,0	10,0	10,0
4	10,0	16,1	10,9	9,8	13,2	10,0	10,2	10,1	10,1	10,1
5	10,0	17,0	11,3	10,1	13,6	10,4	10,7	10,6	10,6	10,6
6	10,0	18,1	11,5	9,8	14,4	10,2	10,5	10,4	10,5	10,3
7	10,0	20,1	12,5	10,0	16,0	10,7	11,3	11,0	11,1	10,9
8	10,0	23,1	13,2	9,7	18,4	10,5	11,6	11,1	11,2	11,0

Nota: Os números (1), (2), (3), (4), (5), (6), (7), (8) e (9) correspondem a $P(LR \geq x_\alpha)$, $P(LR^* \geq x_\alpha)$, $P(LR_1^* \geq x_\alpha)$, $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$, $P(S_{R1}^* \geq x_\alpha)$, $P(S_{R2}^* \geq x_\alpha)$, $P(S_{R3}^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ respectivamente.

Tabela 3.5 *Tamanhos dos testes - $p=10$ - modelo gama com índice $\phi=1.0$*

n	$tamanhos$ $nominais(\%)$	LR			S_R					
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
20	1,0	14,7	2,4	5,4	5,6	3,2	1,0	3,3	1,8	2,1
	5,0	27,3	9,0	7,6	11,3	6,8	4,8	7,1	5,8	9,3
	10,0	37,3	16,0	10,9	21,7	11,4	10,5	12,4	11,1	14,4
30	1,0	6,9	1,3	2,1	2,4	1,7	1,0	1,8	1,2	1,4
	5,0	17,6	6,8	5,1	8,6	5,4	4,8	5,6	5,1	6,0
	10,0	25,2	13,1	10,0	17,2	10,4	10,6	11,1	10,6	11,4
40	1,0	4,1	1,2	1,2	1,1	1,2	1,0	1,2	1,0	1,3
	5,0	12,7	6,2	4,9	7,3	5,2	5,0	5,3	5,1	5,2
	10,0	20,4	11,7	9,7	15,2	10,1	10,4	10,5	10,3	10,3
50	1,0	3,5	1,3	1,1	1,4	1,2	1,1	1,2	1,2	1,2
	5,0	10,6	5,9	5,0	7,4	5,4	5,5	5,5	5,5	5,4
	10,0	17,9	11,3	10,0	13,9	10,3	10,6	10,5	10,6	10,4

Nota: Os números (1), (2), (3), (4), (5), (6), (7), (8) e (9) correspondem a $P(LR \geq x_\alpha)$, $P(LR^* \geq x_\alpha)$, $P(LR_1^* \geq x_\alpha)$, $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$, $P(S_{R1}^* \geq x_\alpha)$, $P(S_{R2}^* \geq x_\alpha)$, $P(S_{R3}^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ respectivamente.

Tabela 3.6 *Tamanhos dos testes - $n=30$ - modelo gama com índice $\phi=1,0$*

$p - q$	tamanhos nominais(%)	LR					S_R			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	1,0	1,6	1,3	1,2	1,4	0,7	1,1	0,9	1,2	1,3
2	1,0	1,5	0,9	0,9	0,9	1,1	1,1	1,1	1,1	1,2
3	1,0	2,6	1,1	1,7	1,0	0,8	0,8	0,8	0,8	1,7
4	1,0	3,2	1,3	1,3	1,4	1,2	1,2	1,2	1,2	1,4
5	1,0	4,2	1,2	2,6	1,1	1,0	1,0	1,0	1,0	2,9
6	1,0	4,5	1,1	2,3	0,7	0,9	0,7	0,9	0,8	2,5
7	1,0	5,2	1,3	2,1	0,9	1,0	0,7	1,0	0,8	2,3
8	1,0	6,9	1,3	2,1	2,4	1,7	0,8	1,7	1,2	1,4
1	5,0	6,6	5,3	5,2	5,0	5,4	5,6	5,5	5,6	5,6
2	5,0	6,9	5,0	4,9	5,0	5,3	5,3	5,3	5,3	5,4
3	5,0	8,0	5,2	5,6	5,8	5,0	5,0	5,0	5,0	5,9
4	5,0	9,6	6,1	5,5	7,0	5,4	5,5	5,5	5,5	5,6
5	5,0	10,6	5,1	6,1	6,3	4,9	4,9	4,9	4,9	6,7
6	5,0	11,9	5,7	6,1	6,1	4,8	4,7	4,8	4,8	6,3
7	5,0	13,4	6,3	5,7	6,7	5,1	5,0	5,2	5,1	6,1
8	5,0	17,1	6,8	5,1	8,6	5,4	4,8,4	5,6	5,1	6,0
1	10,0	12,2	10,3	10,1	10,0	10,6	10,8	10,7	10,7	10,8
2	10,0	12,6	10,0	10,0	10,0	10,3	10,3	10,3	10,4	10,4
3	10,0	14,1	10,2	10,3	11,7	10,1	10,2	10,1	10,2	11,0
4	10,0	16,4	11,3	10,4	13,5	10,4	10,7	10,6	10,7	10,7
5	10,0	17,3	10,3	10,6	12,3	10,0	10,0	10,0	10,0	11,5
6	10,0	19,1	11,0	10,8	12,8	10,0	10,0	10,0	10,0	11,3
7	10,0	21,0	11,65	10,4	13,9	10,0	10,0	10,0	10,0	10,8
8	10,0	25,2	13,1	10,0	17,2	10,4	10,6	11,1	10,6	11,4

Nota: Os números (1), (2), (3), (4), (5), (6), (7), (8) e (9) correspondem a $P(LR \geq x_\alpha)$, $P(LR^* \geq x_\alpha)$, $P(LR_1^* \geq x_\alpha)$, $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$, $P(S_{R1}^* \geq x_\alpha)$, $P(S_{R2}^* \geq x_\alpha)$, $P(S_{R3}^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ respectivamente.

Figura 3.4 : Tamanhos dos Testes - $n=30$
Modelo Gama - $\phi = 2$ e $\alpha = 1\%$

Figura 3.5 : Tamanhos dos Testes - $n=30$
Modelo Gama - $\phi=2$ e $\alpha=5\%$

Figura 3.6 : Tamanhos dos Testes - $n=30$
Modelo Gama - $\phi=2$ e $\alpha=10\%$

Figura 3.8 : Tamanhos dos Testes - $n=30$
Modelo Gama - $\phi=1$ e $\alpha=5\%$

Figura 3.9 : Tamanhos dos Testes - $n=30$
Modelo Gama - $\phi=1$ e $\alpha=10\%$

CAPÍTULO 4

Simulações em modelos discretos

Neste capítulo, apresentaremos alguns resultados de simulação para avaliar a eficácia de correções de Bartlett e tipo-Bartlett para testes da razão de verossimilhança e escore em MLGs discretos.

Sabemos que as correções de Bartlett e tipo-Bartlett são utilizadas para melhorar a qualidade das aproximações das estatísticas da razão de verossimilhança e escore para a distribuição χ^2 . Para os modelos contínuos, sob condições adequadas de regularidade, os resultados teóricos garantem que as distribuições das estatísticas corrigidas são χ^2 até ordem n^{-1} . Este resultado não é necessariamente válido para modelos discretos. Claro está, pelos resultados desenvolvidos no Capítulo 2, que as fórmulas para as correções podem ser obtidas tanto para modelos contínuos quanto discretos. No entanto, pelo que foi exposto acima, os resultados teóricos, relativos aos erros de aproximações por χ^2 , somente são garantidos no caso contínuo (Frydenberg e Jensen, 1989).

Diversos artigos são encontrados na literatura estatística tratando de fórmulas para correções de testes em casos discretos levando, então, à falsa impressão de que as aproximações são válidas também nestes casos.

Cordeiro (1982) faz vários estudos de simulação envolvendo distribuições multinomial e de Poisson que mostram que os testes modificados, baseados em LR^* , apresentam taxas de rejeição da hipótese nula bem mais próximas dos respectivos níveis nominais do que os testes não modificados, baseados em LR . Por outro lado, Frydenberg e Jensen, (1989) mostram, através de simulação, que os desempenhos dos testes modificados chegam a ser piores que os dos não modificados, em situações especiais. Nosso objetivo aqui, é, então, investigar, através de um estudo de simulação, o desempenho das estatísticas da razão de verossimilhança e escore

modificadas por fatores de Bartlett e tipo-Bartlett nos MLGs discretos.

As simulações feitas para o caso discreto são totalmente análogas àquelas realizadas para o caso contínuo (ver Capítulo 3). Os resultados para a distribuição de Poisson com ligação canônica, são apresentados nas Tabelas 4.1 e 4.2 e nas Figuras 4.1 a 4.3.

A Tabela 4.1 nos indica que há um ganho com o uso da correção de Bartlett e tipo-Bartlett. Os testes modificados apresentam um melhor desempenho em relação aos testes não modificados. No caso da estatística da razão de verossimilhança, a correção parece ser mais necessária e eficaz do que no caso da estatística escore uma vez que o teste escore apresenta taxas de rejeição razoavelmente próximas dos níveis nominais, mesmo para n pequeno.

Analisando as Figuras 4.1 a 4.3 e a Tabela 4.2, podemos notar claramente que o desempenho do teste escore é bem melhor do que o da razão de verossimilhança. Além disso, é evidente também que os desempenhos dos testes da razão de verossimilhança modificados são bem melhores do que o do não modificado especialmente se $p-q$ é grande. Por outro lado, em algumas situações, alguns testes escore modificados apresentaram taxas de rejeição mais distantes dos níveis nominais do que o teste escore usual. Deve-se levar em conta, no entanto, que as diferenças entre as taxas de rejeição, em diversas situações, são muito pequenas e, assim, é difícil fazer distinção entre os testes, com base em suas taxas de rejeição, uma vez que estas são estimadas via simulação (que envolve, evidentemente, erros aleatórios).

Para a segunda simulação no caso discreto, consideramos o modelo binomial com parâmetro $m=5$ e função de ligação canônica. Os resultados desta distribuição são apresentados nas Tabelas 4.3 e 4.4 e nas Figuras 4.4 a 4.6. Uma observação a ser feita com respeito a esta simulação é que para valores de $\hat{w} = \hat{\mu}(1 - \hat{\mu})$ muito próximos de zero ou um, não é possível calcular a estatística escore. Para solucionar este problema, consideramos os valores dos pesos $\hat{w}$ em uma determinada

faixa ($10^{-5} < \hat{w} < 10^5$), com a condição de que as amostras que tivessem pesos fora desta seriam excluídas, de forma que se fosse permitido fazer o cálculo da estatística *escore*. A proporção de amostras excluídas nas simulações foi próxima de 0,35% e em apenas três casos ocorreram em torno de 0,75%.

Na Tabela 4.3, nota-se que há um ganho com o uso da correção de Bartlett e tipo-Bartlett e os testes modificados apresentam um melhor desempenho em relação aos testes não modificados. Podemos observar que como estamos trabalhando com $m=5$ e com amostras de tamanhos $n=20, 30, 40$ e 50 (que correspondem a $m \times n = 100, 150, 200$ e 250 ensaios de Bernoulli), os testes sem correção já apresentam taxas de rejeição razoavelmente próxima dos níveis nominais.

Analisando a Tabela 4.4 e as Figuras 4.4 a 4.6, podemos notar que o teste *escore* tende a ser melhor que o da razão de verossimilhança. Notamos também que se o número de parâmetros de perturbação não é muito grande (digamos, $p-q \leq 5$), ambas as estatísticas não corrigidas têm boa aproximação por χ^2 e, nesta situação, as correções têm efeito insignificante. Se $p - q \geq 6$ a aproximação por χ^2 já não é tão boa e as correções produzem estatísticas que apresentam taxas de rejeição mais próximas dos níveis nominais.

Tabela 4.1 *Tamanhos dos testes - $p=10$ - modelo de Poisson*

n	<i>tamanhos nominais(%)</i>	LR					S_R			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
20	1,0	2,1	1,0	1,0	1,0	1,1	1,0	1,1	1,3	1,3
	5,0	8,9	4,8	4,2	5,5	4,7	4,6	4,7	5,0	5,0
	10,0	16,7	10,5	9,0	11,7	9,3	9,3	9,5	9,6	11,1
30	1,0	1,9	1,3	1,2	1,2	1,1	1,1	1,1	1,1	1,2
	5,0	8,1	5,7	5,5	6,1	5,2	5,2	5,2	5,3	5,4
	10,0	14,3	10,9	10,5	11,7	10,2	10,3	10,2	10,3	10,4
40	1,0	1,4	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
	5,0	6,3	5,2	5,0	5,0	4,5	4,5	4,5	4,5	4,5
	10,0	12,2	10,3	10,2	10,9	9,8	9,9	9,8	9,8	9,8
50	1,0	1,2	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1
	5,0	5,6	5,1	5,1	5,4	5,2	5,2	5,2	5,2	5,2
	10,0	10,8	10,1	10,0	10,5	10,1	10,1	10,1	10,1	10,1

Nota: Os números (1), (2), (3), (4), (5), (6), (7), (8) e (9) correspondem a $P(LR \geq x_\alpha)$, $P(LR^* \geq x_\alpha)$, $P(LR_1^* \geq x_\alpha)$, $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$, $P(S_{R1}^* \geq x_\alpha)$, $P(S_{R2}^* \geq x_\alpha)$, $P(S_{R3}^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ respectivamente.

Tabela 4.2 *Tamanhos dos testes - $n=20$ - modelo de Poisson*

$p - q$	tamanhos nominais(%)	LR					S_R			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0	1,0	1,1	1,0	1,0	1,2	0,6	0,9	0,7	0,9	1,0
1	1,0	1,3	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1
2	1,0	1,3	1,0	1,1	0,9	1,0	1,0	1,0	1,0	1,0
3	1,0	1,0	0,9	0,9	0,9	1,0	1,0	1,0	1,0	1,0
4	1,0	1,5	1,2	1,2	1,2	1,2	1,2	1,2	1,2	1,2
5	1,0	1,6	1,2	1,2	1,0	1,1	1,1	1,1	1,1	1,1
6	1,0	2,0	1,4	1,3	0,9	1,0	1,0	1,0	1,0	1,0
7	1,0	2,5	1,5	1,4	0,9	1,1	1,0	1,1	1,1	1,1
8	1,0	2,2	1,0	1,0	1,0	1,1	1,0	1,1	1,3	1,3
0	5,0	5,4	4,9	4,9	5,3	5,5	5,5	5,5	5,5	5,5
1	5,0	6,1	5,4	5,4	4,6	5,0	5,1	5,0	5,1	5,0
2	5,0	6,0	5,2	5,2	4,8	5,0	5,0	5,0	5,0	5,0
3	5,0	6,1	5,4	5,4	5,0	5,1	5,1	5,1	5,1	5,1
4	5,0	6,0	5,1	5,1	5,0	5,0	5,0	5,0	5,0	5,0
5	5,0	6,6	5,4	5,3	5,1	5,0	5,0	5,0	5,0	5,0
6	5,0	7,6	5,9	5,8	5,7	5,6	5,6	5,6	5,6	5,6
7	5,0	8,5	6,0	5,8	5,4	5,1	5,1	5,2	5,2	5,5
8	5,0	8,9	4,8	4,2	5,5	4,7	4,6	4,7	5,0	5,0
0	10,0	10,8	10,1	10,1	9,4	10,0	10,0	10,0	10,0	10,0
1	10,0	11,5	10,5	10,5	9,6	10,0	10,0	10,0	10,0	10,0
2	10,0	11,4	10,5	10,4	9,7	9,9	9,8	9,8	9,8	9,8
3	10,0	11,6	10,3	10,3	9,7	9,9	9,9	9,9	10,0	9,9
4	10,0	11,3	10,0	10,2	10,2	9,9	9,9	9,9	9,9	9,9
5	10,0	11,9	10,1	10,0	10,3	9,8	9,8	9,8	9,8	9,8
6	10,0	13,3	11,0	10,7	11,0	10,3	10,4	10,4	10,4	10,3
7	10,0	14,6	11,4	10,9	11,1	10,3	10,2	10,3	10,3	10,5
8	10,0	16,7	10,5	9,0	11,7	9,3	9,3	9,5	9,6	11,1

Nota: Os números (1), (2), (3), (4), (5), (6), (7), (8) e (9) correspondem a $P(LR \geq x_\alpha)$, $P(LR^* \geq x_\alpha)$, $P(LR_1^* \geq x_\alpha)$, $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$, $P(S_{R1}^* \geq x_\alpha)$, $P(S_{R2}^* \geq x_\alpha)$, $P(S_{R3}^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ respectivamente.

Tabela 4.3 *Tamanhos dos testes - $p=10$ - modelo binomial com $m=5$*

n	tamanhos nominais(%)	LR					S_R			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
20	1,0	1,9	1,2	1,2	1,4	1,1	1,1	1,2	1,2	1,3
	5,0	7,3	5,5	5,2	6,5	5,2	5,2	5,2	5,2	5,4
	10,0	13,3	10,3	10,0	12,2	9,9	10,0	10,0	10,1	10,2
30	1,0	1,5	1,1	1,1	1,3	1,1	1,1	1,1	1,1	1,2
	5,0	6,5	5,4	5,4	6,0	5,4	5,4	5,4	5,4	5,4
	10,0	12,1	10,4	10,3	11,8	10,3	10,3	10,3	10,3	10,3
40	1,0	1,1	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
	5,0	6,3	5,3	5,3	6,0	5,3	5,3	5,3	5,3	5,3
	10,0	11,8	10,7	10,7	11,5	10,7	10,7	10,7	10,7	10,7
50	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
	5,0	5,3	4,6	4,6	5,1	4,6	4,7	4,7	4,7	4,7
	10,0	11,0	10,2	10,1	10,8	10,1	10,1	10,1	10,1	10,1

Nota: Os números (1), (2), (3), (4), (5), (6), (7), (8) e (9) correspondem a $P(LR \geq x_\alpha)$, $P(LR^* \geq x_\alpha)$, $P(LR_1^* \geq x_\alpha)$, $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$, $P(S_{R1}^* \geq x_\alpha)$, $P(S_{R2}^* \geq x_\alpha)$, $P(S_{R3}^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ respectivamente.

Tabela 4.4 *Tamanhos dos testes - $n=30$ - modelo binomial com $m=5$*

$p - q$	tamanhos nominais(%)	LR					S_R			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0	1,0	1,1	1,0	1,0	0,9	1,0	1,0	1,0	1,0	1,0
1	1,0	1,1	1,0	1,0	1,0	1,0	1,0	1,0	1,1	1,0
2	1,0	1,1	1,1	1,1	1,0	1,1	1,1	1,1	1,1	1,1
3	1,0	1,1	1,0	1,0	0,9	0,9	0,9	0,9	0,9	0,9
4	1,0	1,0	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9
5	1,0	1,1	0,9	0,9	1,0	0,9	0,9	0,9	0,9	0,9
6	1,0	1,4	1,2	1,2	1,3	1,2	1,2	1,2	1,2	1,2
7	1,0	1,3	1,0	1,0	1,1	1,0	1,0	1,0	1,0	1,0
8	1,0	1,5	1,1	1,1	1,3	1,1	1,1	1,1	1,1	1,2
0	5,0	5,5	5,2	5,2	5,1	5,2	5,2	5,2	5,2	5,2
1	5,0	5,1	4,8	4,8	4,8	4,8	4,8	4,8	4,9	4,8
2	5,0	5,5	5,1	5,1	5,1	5,1	5,1	5,1	4,8	5,1
3	5,0	5,1	4,7	4,7	4,9	4,7	4,7	4,7	4,2	4,7
4	5,0	5,4	4,8	4,8	5,1	4,8	4,8	4,8	4,8	4,8
5	5,0	5,8	5,1	5,1	5,4	5,1	5,1	5,1	5,1	5,1
6	5,0	6,1	5,3	5,3	5,8	5,3	5,3	5,3	5,3	5,3
7	5,0	6,0	5,0	4,9	5,7	4,9	4,9	4,9	4,9	4,9
8	5,0	6,5	5,3	5,3	6,0	5,4	5,4	5,4	5,4	5,4
0	10,0	10,6	10,3	10,3	10,2	10,3	10,3	10,3	10,3	10,3
1	10,0	10,5	9,9	9,8	10,0	9,9	9,9	9,9	9,5	9,9
2	10,0	10,5	9,9	9,9	10,1	9,9	9,9	9,9	9,8	9,9
3	10,0	10,5	9,7	9,7	10,0	9,7	9,7	9,7	9,6	9,7
4	10,0	10,7	10,0	10,0	10,4	10,0	10,0	10,0	10,0	10,0
5	10,0	10,7	9,9	9,9	10,3	9,9	9,9	9,9	9,9	9,9
6	10,0	11,2	10,3	10,2	11,0	10,2	10,3	10,2	10,2	10,2
7	10,0	11,7	10,3	10,3	11,2	10,3	10,3	10,3	10,3	10,3
8	10,0	12,1	10,4	10,3	11,8	10,3	10,3	10,3	10,3	10,3

Nota: Os números (1), (2), (3), (4), (5), (6), (7), (8) e (9) correspondem a $P(LR \geq x_\alpha)$, $P(LR^* \geq x_\alpha)$, $P(LR_1^* \geq x_\alpha)$, $P(S_R \geq x_\alpha)$, $P(S_R^* \geq x_\alpha)$, $P(S_{R1}^* \geq x_\alpha)$, $P(S_{R2}^* \geq x_\alpha)$, $P(S_{R3}^* \geq x_\alpha)$ e $P(S_R \geq z_\alpha)$ respectivamente.

Figura 4.1 : Tamanhos dos Testes - $n=20$
Modelo de Poisson - $\alpha=1\%$

Figura 4.2 : Tamanhos dos Testes - $n=20$
Modelo de Poisson - $\alpha=5\%$

Figura 4.3 : Tamanhos dos Testes - $n=20$
Modelo de Poisson - $\alpha=10\%$

Figura 4.4 : Tamanhos dos Testes - $n=30$
Modelo Binomial - $m=5$ e $\alpha=1\%$

Figura 4.5 : Tamanhos dos Testes - $n=30$
Modelo Binomial - $m=5$ e $\alpha=5\%$

**Figura 4.6 : Tamanhos dos Testes - $n=30$
Modelo Binomial - $m=5$ e $\alpha=10\%$**

Conclusões

Nesta dissertação, revisamos alguns resultados teóricos acerca dos modelos lineares generalizados e da obtenção das correções de Bartlett e tipo-Bartlett nestes modelos. Estudamos diversos casos particulares e aplicamos os resultados a conjuntos de dados reais. Além disto, apresentamos estudos de simulação para averiguar o efeito das correções em modelos lineares generalizados contínuos e discretos.

Nos modelos contínuos que foram simulados e para os tamanhos de amostra considerados, os resultados de simulação mostraram que, em geral, o teste escore tende a ter desempenho melhor do que o da razão de verossimilhança. Este é um resultado interessante pois o cálculo da estatística escore é mais simples do que o da estatística da razão de verossimilhança, uma vez que a primeira envolve estimação somente sob o modelo restrito (sob H_0), enquanto que a segunda envolve estimação também sob o modelo irrestrito. Nos modelos contínuos, as simulações mostraram que todas as correções são muito eficazes, produzindo testes com tamanhos bem mais próximos dos respectivos níveis nominais do que os testes sem correção. A importância das correções fica ainda mais evidente no caso em que o tamanho da amostra não é grande ou o número de parâmetros de perturbação é considerável. Nos modelos discretos, as correções também se mostraram eficazes. No entanto, na maioria dos casos simulados, os testes não corrigidos já apresentaram tamanhos próximos dos níveis nominais e, portanto, as correções não apresentaram efeito muito marcante.

Deve ser ressaltado que os estudos de simulação desenvolvidos nesta dissertação procuraram avaliar o efeito das correções nos tamanhos dos testes. Pelo exposto acima, todas as correções apresentaram bom efeito sob este enfoque. No entanto, não foram feitos estudos dos poderes dos testes. Ferrari (1991) mostra que os

testes corrigidos têm o mesmo poder que os correspondentes testes não corrigidos até ordem $n^{-1/2}$. No entanto, é possível que, em amostras pequenas ou de tamanho moderado, alguns dos testes corrigidos tenham poderes muito diferentes dos demais.

Para concluir, acrescentamos os estudos de simulação que poderão ser realizados no futuro:

- (i) comparação dos poderes dos testes corrigidos e não corrigidos;
- (ii) análise da influência dos valores das variáveis auxiliares no desempenho dos testes;
- (iii) análise exhaustiva do efeito das correções em modelos discretos.

Referências

- Abramowitz, M. e Stegun, I.A. (1980). *Handbook of Mathematical Functions*. New York: Dover Publications, Inc. 1046p.
- Aitkin, M., Anderson, D., Francis, B. e Hinde, J. (1990). *Statistical Modelling in GLIM*. Oxford: Clarendon Press. 374p.
- Bartlett, M.S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society A*, **160**, 268-282.
- Chambers, J. H. e Hastie, J. T. (1992) *Statistical Models in S*. California: Wadsworth & Brooks/Cole Advanced Books & Software Pacific Grove.
- Chesher, A. e Smith, R. (1995). Bartlett corrections to likelihood ratio tests. *Biometrika*, **82**, 433-436.
- Chhikara, R.S. e Folks, J.L. (1989). *The Inverse Gaussian Distribution: Theory, Methodology and Applications*. New York: Marcel Dekker.
- Cordeiro, G.M.(1982). *Improved likelihood ratio statistics for generalized linear models*. London. 227p. Tese (Doutorado) - Imperial College of Science and Technology - University of London.
- Cordeiro, G.M.(1983). Improved likelihood ratio statistics for generalized linear models. *Journal of the Royal Statistical Society B*, **45**, 404-13.
- Cordeiro, G.M.(1986). *Modelos Lineares Generalizados*. VII SINAPE, Simpósio Nacional de Probabilidade e Estatística. Campinas, São Paulo. 286p.
- Cordeiro, G.M.(1987). On the corrections to the likelihood ratio statistic. *Biometrika*, **74**, 265-74.
- Cordeiro, G.M. e Ferrari, S.L.P. (1991). A modified score test statistic having chisquared distribution to the order n^{-1} . *Biometrika*, **78**, 573-82.
- Cordeiro, G.M. e Ferrari, S.L.P. (1996). A method of moments for finding Bartlett-

- type corrections. Relatório Técnico, RT - MAE 9608. Departamento de Estatística - Universidade de São Paulo.
- Cordeiro, G.M., Ferrari, S.L.P. e Paula, G.A. (1993). Improved score tests for generalized linear models. *Journal of the Royal Statistical Society B*, **55**, 661-74.
- Cordeiro, G.M. e McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society B* **53**, 629-643.
- Cox, D.R. e Reid, N. (1987). Parameter orthonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society, B*, **40**(1): 1-39.
- Cox, D.R. e Snell, E.J. (1981) *Applied statistics*. London, Chapman & Hall, 189p.
- Cribari-Neto, F. (1994). *Finite-sample corrections to asymptotically chi-squared econometric criteria*. 124p. Tese (Doutorado). University of Illinois at Urbana-Champaign - The Graduate College.
- Cribari-Neto, F. e Cordeiro, G.M. (1997). On Bartlett and Bartlett-type corrections. *Econometric Reviews*, **15**, 339-367.
- Cribari-Neto, F. e Ferrari, S.L.P. (1995). Second order asymptotics for score tests in generalized linear models. *Biometrika*, **82**, 426-32.
- Dobson, A.J (1990). *An introduction to generalized linear models*, 2nd ed London, Chapman & Hall. 174p.
- Dudewicz, E.J. e Mishra S.N. (1988). *Modern Mathematical*. 1nd ed. New York, John Wiley. 838p. (Wiley Series in Probability and Mathematical Statistics).
- Fahrmeir, L. e Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics* **13**, 342-368.
- Ferrari, S.L.P. (1991) *Aperfeiçoamento de testes escore, aplicações e extensões*. 185p. Tese (Doutorado). Instituto de Matemática e Estatística - Universidade de São Paulo.
- Frydenberg, M. e Jensen, J.L. (1989). Is the "improved likelihood statistic " really improved in discrete case? *Biometrika*, **76**, 655-61.
- Gray, J.R. (1989). On the use of regression diagnostics. *The Statistician* **38**(2):

97-105.

Harris, P. (1985). An asymptotic expansion for the null distributions of the efficient score statistic. *Biometrika*, **72**(3): 653-9.

Hayakawa, T. (1977). The likelihood ratio criterion and the asymptotic expansion of its distribution. *Annals of the Institute of Statistical Mathematics A*, **28**(3): 309-28.

IMSL Library Reference Manual. (1982) 9th ed. Houston. IMSL Inc.

Jørgensen, B. (1987). Small dispersion asymptotics. REBRAPE: Revista Brasileira de Probabilidade e Estatística, **1**(1): 59-90.

Lawless, J.F. (1982). *Statistical models and methods for lifetime data*. New York, John Wiley. 580p. (Wiley Series in Probability and Mathematical Statistics).

Lawley, D.N. (1956). A generalized method for approximating to the distribution of the likelihood ratio criteria. *Biometrika*, **43**, 295-303.

McCullagh, P. e Nelder, J.A. (1989) *Generalized Linear Models*, 2nd ed. London, Chapman & Hall. 511p.

Nelder, J.A. e Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society A*, **135**, 370-384.

Payne, C.D., (1986). *The GLIM Manual: Release 3.77*. Oxford, NAG.

Pregibon, D. (1982). Score tests in GLIM with applications. In: Gilchrist, R., ed. GLIM 82: Proceedings of the international conference on Generalized Linear Models, 1, London, 1982. Berlin, Springer. 87-97. (Lecture Notes in Statistics, 14).

Rao, C.R. (1947). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society*, **44**: 50-7.

Rao, C.R. (1973). *Linear Statistical Inference and its Applications*. 2nd ed. New York, John Wiley. 625p. (Wiley Series in Probability and Mathematical Statistics).

Sen, P.K. e Singer, J.M. (1993). *Large Sample Methods in Statistics: An Introduction with Applications*. New York: Chapman & Hall.

- Wedderburn, R. W. M. (1976). On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika* 63, 27-32.