

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE ESTATÍSTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

CORREÇÃO TIPO-BARTLETT À ESTATÍSTICA
GRADIENTE NOS MODELOS NÃO LINEARES DA
FAMÍLIA EXPONENCIAL

TELMA DE SOUZA LÔBO

DISSERTAÇÃO DE MESTRADO



Recife
2016

TELMA DE SOUZA LÔBO

CORREÇÃO TIPO-BARTLETT À ESTATÍSTICA
GRADIENTE NOS MODELOS NÃO LINEARES DA
FAMÍLIA EXPONENCIAL

ORIENTADORA: PROFA. DRA. AUDREY HELEN MARIZ DE AQUINO CYSNEIROS

Área de Concentração: Estatística Matemática

Dissertação apresentada ao programa de Pós-graduação em Estatística do Departamento de Estatística da Universidade Federal de Pernambuco, como requisito parcial para obtenção do grau de **Mestre em Estatística**.

Recife
2016

Catálogo na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

L799c Lôbo, Telma de Souza.
 Correção tipo-Bartlett à estatística gradiente nos modelos não lineares da
 família exponencial / Telma de Souza Lôbo. – 2016.
 69 f.: il., fig., tab.

 Orientadora: Audrey Helen Mariz de Aquino Cysneiros.
 Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN,
 Estatística, Recife, 2016.
 Inclui referências e apêndices.

 1. Estatística matemática. 2. Correção tipo Bartlett. I. Cysneiros, Audrey
 Helen Mariz de Aquino (orientadora). II. Título.

 519.5 CDD (23. ed.) UFPE- MEI 2016-038

TELMA DE SOUZA LÔBO

**CORREÇÃO TIPO-BARTLETT À ESTATÍSTICA GRADIENTE NOS MODELOS
NÃO LINEARES DA FAMÍLIA EXPONENCIAL**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Aprovada em: 22 de fevereiro de 2016.

BANCA EXAMINADORA

Prof.^a Dra. Audrey Helen Mariz de Aquino Cysneiros
UFPE

Prof. Dr. Manoel Raimundo de Sena Júnior (Examinador Interno)
UFPE

Prof.^a Dra. Renata Maria Cardoso Rodrigues de Souza (Examinador Externo)
UFPE

Dedico esse trabalho

À minha mãe e a todos os colaboradores.

Agradecimentos

Primeiramente a Deus, onde busquei força e sabedoria.

Em especial à minha mãe pelo apoio, carinho, amor e dedicação em todos momentos de necessidade.

Às minhas irmãs Iara, Cíntia e Josiane pela compreensão dos momentos em que não pude estar presente.

ÀS minhas amigas de Ouro Preto e Josiane Barros pelas palavras de motivação.

Aos meus amigos que colaboraram cedendo seus computadores para minhas simulações, Ewerton, Gaston e Rafael.

Às minhas companheiras de apartamento Roberta, Lilia, Andréa e Moniky pela amizade e momentos de descontração.

Aos meus amigos de turma por todo companherismo e amizade, em especial, Jonas, Pedro, Hildemar e Rodrigo Alves.

A todos os professores que fizeram parte da minha formação acadêmica.

Em especial, à minha orientadora, Profa. Audrey Helen Mariz de Aquino Cysneiros, pela orientação e a confiança em mim depositada.

Aos demais funcionários do Departamento de Pós-Graduação em Estatística por seus trabalhos realizados.

À Capes, pelo apoio financeiro.

“Todas as vitórias ocultam uma abdicação.”

Simone de Beauvoir

Resumo

Nesta dissertação, tratamos de refinamentos para testes de hipóteses em modelos de regressão não lineares heteroscedásticos. Na primeira parte da dissertação, reunimos resultados importantes da literatura sobre correção de Bartlett à razão de verossimilhanças perfiladas modificadas (CYSNEIROS; FERRARI, 2006) e razão de verossimilhanças bootstrap (ROCKE, 1989) nos modelos de regressão não lineares da família exponencial com parâmetros de dispersão que não são constantes para todas as observações, respectivamente. Na segunda parte, obtemos um fator de correção tipo-Bartlett para a estatística gradiente (TERRELL, 2002), nesta classe de modelos. Na terceira parte, desenvolvemos estudos de simulação Monte Carlo para avaliar e comparar numericamente os desempenhos dos testes em amostras finitas. Finalmente, realizamos uma aplicação empírica.

Palavras-chave: Correção de Bartlett. Correção tipo-Bartlett. Estatística gradiente. Estatística da razão de verossimilhanças. Modelo não linear heteroscedástico. Verossimilhança perfilada.

Abstract

In this dissertation, we deal with refinements for hypothesis tests in nonlinear heteroskedastic regression models. Firstly, we describe important results in the literature on Bartlett corrections to the modified profile likelihood ratio and likelihood ratio bootstrap to the class of the exponential family nonlinear models with dispersion parameters that are not constant over the observations, respectively. Secondly, we obtain a Bartlett-type correction to the gradient statistic in this class of models. Thirdly, the finite-sample performances of the various tests considered in this dissertation are evaluated and compared using Monte Carlo simulations. Finally, we present one empirical application.

Keywords: Bartlett correction. Bartlett-type correction. Bootstrap. Gradient. Likelihood ratio. Modified profile likelihood. Score. Nonlinear heteroskedastic regression model.

Lista de Figuras

2.1	Distorção do tamanho dos testes LR, LR _m , LR _{Boot} , S, G - Modelo normal não linear com $n = 35$, $q = 2$ e diferentes valores de p	30
2.2	Discrepâncias relativas de quantis das estatísticas LR, LR _m , S, G - Modelo normal não linear com $n = 35$, $p = 5$ e $q = 2$	30
3.1	Discrepâncias relativas de quantis das estatísticas LR, LR _m , S, G, LR _m [*] , LR _{Boot} [*] , G [*] - Modelo normal não linear com $n = 35$, $p = 5$ e $q = 2$	47
3.2	(a) Gráfico dos resíduos padronizados contra percentis normais com envelope simulado; (b) Gráfico de dispersão do peso das lentes dos olhos contra a idade; (c) Gráfico dos resíduos studentizados t_{s_i} contra os valores ajustados e (d) Gráfico dos resíduos studentizados t_{s_i} contra x para o modelo homoscedástico - Coelhos europeus na Austrália.	49
3.3	(a) Gráfico dos resíduos padronizados contra percentis normais com envelope simulado; (b) Gráfico de dispersão da quantidade absorvida de cálcio radioativo contra o tempo; (c) Gráfico dos resíduos studentizados t_{s_i} contra os valores ajustados e (d) Gráfico dos resíduos studentizados t_{s_i} contra x para o modelo homoscedástico - Solução de cálcio radioativo.	51

Lista de Tabelas

2.1	Quantidades d_ℓ e $s(\phi_\ell)$ para os modelos normal, normal inversa e gama.	18
2.2	Tamanhos empíricos dos testes LR, LR_m , LR_{Boot} , S e G - Modelo normal não linear para $n = 35$, $q = 2$ e diferentes valores de p	28
2.3	Tamanhos empíricos dos testes LR, LR_m , LR_{Boot} , S e G - Modelo normal não linear para $p = 5$, $q = 2$ e diferentes tamanhos amostrais.	29
2.4	Poder dos testes LR, LR_m , LR_{Boot} , S e G - Modelo normal não linear para $n = 35$, $p = 5$, $q = 2$ e $\alpha = 10, 0\%$	29
3.1	Tamanhos empíricos dos testes LR, LR_m , LR_{Boot} , S, G, LR_m^* , LR_{Boot}^* e G^* - Modelo normal não linear para $n = 35$, $q = 2$ e diferentes valores de p	42
3.2	Tamanhos empíricos dos testes LR, LR_m , LR_{Boot} , S, G, LR_m^* , LR_{Boot}^* e G^* - Modelo normal não linear para $n = 35$, $p = 5$ e diferentes valores de q	43
3.3	Tamanhos empíricos dos testes LR, LR_m , LR_{Boot} , S, G, LR_m^* , LR_{Boot}^* e G^* - Modelo normal não linear para $p = 5$, $q = 2$ e diferentes tamanhos amostrais.	44
3.4	Poder dos testes LR_m , LR_{Boot} , S, LR_m^* , LR_{Boot}^* e G^* - Modelo normal não linear para $n = 35$, $p = 5$, $q = 2$ e $\alpha = 10, 0\%$	45
3.5	Quantis das estatísticas LR, LR_m , S, G, LR_m^* , LR_{Boot}^* e G^* - Modelo normal não linear para $p = 5$, $q = 2$ e diferentes tamanhos amostrais.	46
3.6	Média e variância das estatísticas LR, LR_m , LR_{Boot} , S, G, LR_m^* , LR_{Boot}^* e G^* - Modelo normal não linear para $p = 5$ e $n = 35$ e diferentes valores de q	48
D.1	Pesos das lentes dos olhos de coelhos europeus (y), em miligramas e idade (x), em dias, numa amostra de 71 observações.	68
E.1	Quantidade absorvida de cálcio radioativo (y), em nmoles/mg, e tempo (x), em minutos, numa amostra de 27 observações.	69

Sumário

1	Introdução	13
1.1	Introdução	13
1.2	Organização da dissertação	14
2	Modelos não lineares da família exponencial	16
2.1	Introdução	16
2.2	Definição do modelo	17
2.3	Alguns aspectos inferenciais	18
2.4	Testes de hipóteses	21
2.4.1	Teste da razão de verossimilhanças	22
2.4.2	Teste escore	22
2.4.3	Teste gradiente	22
2.4.4	Teste da razão de verossimilhanças perfiladas modificadas	23
2.4.5	Teste da razão de verossimilhanças bootstrap	24
2.5	Avaliação numérica	25
2.6	Comentários	31
3	Aperfeiçoamento de testes em MNLFE	32
3.1	Introdução	32
3.2	Cumulantes conjuntos e identidades de Bartlett	33
3.3	Estatística da razão de verossimilhanças corrigida	34
3.4	Estatística da razão de verossimilhanças perfiladas modificadas corrigida	34
3.5	Estatística da razão de verossimilhanças bootstrap corrigida	35
3.6	Estatística gradiente corrigida	36
3.7	Avaliação numérica	38
3.8	Aplicações	49
3.8.1	Exemplo 1	49
3.8.2	Exemplo 2	50
3.9	Comentários	51

4	Considerações finais	53
	Referências	55
A	Cálculo dos cumulantes nos modelos não lineares da família exponencial com dispersão variável	58
B	Obtenção das quantidades A_1, A_2 e A_3 para a correção tipo-Bartlett em modelos não lineares da família exponencial com dispersão variável	63
C	Multiplicação de “array”	66
D	Coelhos Europeus na Austrália	68
E	Solução de Cálcio Radioativo	69

Capítulo 1

Introdução

1.1 Introdução

Os modelos não lineares da família exponencial (MNLFE) (CORDEIRO; PAULA, 1989) são extensões naturais dos modelos lineares generalizados (GLM) e dos modelos de regressão não lineares normais. Esses modelos são menos restritivos permitindo uma flexibilidade na função de ligação que relaciona a média e as componentes sistemáticas e são compostos por funções conhecidas e parâmetros desconhecidos. Wei (1998) em seu livro apresenta uma discussão bem detalhada dos MNLFE enfatizando na análise de diagnóstico e medidas de influência. Nesta dissertação abordamos os MNLFE com dispersão variável, ou seja, a dispersão não são constantes para todas observações. Os testes mais comuns para verificar a presença de heteroscedasticidade dos dados são baseados nas estatísticas: razão de verossimilhanças (LR), escore (S), wald (W) e gradiente (G) recentemente proposta por Terrell (2002). A estatística gradiente surge como uma alternativa para os testes de hipóteses baseados nas estatísticas usuais por sua simplicidade de cálculo que não exige a determinação da matriz de informação, matriz observada e nem de suas inversas. Em grandes amostras e sob a hipótese nula, as estatísticas LR, S, W e G são equivalentes.

Lemonte e Ferrari (2012b) fizeram um estudo de poder local do teste baseado na estatística gradiente comparando com o poder local dos testes baseados nas estatísticas: razão de verossimilhanças, wald e escore, na qual concluíram que nenhum dos testes é uniformemente mais poderoso que o outro. Além disso, outros estudos sobre o poder local dos testes baseados nas estatísticas LR, S, W e G foram realizados para os modelos: modelos não lineares da família exponencial (LEMONTE, 2011), modelos de regressão Birnbaum-Saunders (LEMONTE; FERRARI, 2011), modelos lineares generalizados com dispersão variável (LEMONTE, 2013) e modelos de dispersão variável (LEMONTE; FERRARI, 2012a).

Geralmente a distribuição exata das estatísticas razão de verossimilhanças, escore, wald e gradiente em amostras finitas é desconhecida, então os testes baseados nestas estatísticas são baseados em aproximações assintóticas. Sob a hipótese nula, a distribuição de referência utilizada é a χ^2 . Contudo, em amostras pequenas essas aproximações pela χ^2 pode não ser muito boa. No entanto, faz-se necessário desenvolver técnicas de aperfeiçoamento das estatísticas de

tal forma que melhore a aproximação das distribuições dessas estatísticas pela distribuição χ^2 de referência produzindo inferências mais precisas. Essa linha de pesquisa é denominada teoria assintótica de alta ordem, que tem como objetivo o refinamento de métodos inferenciais a fim de torná-los mais precisos em amostras finitas e principalmente de tamanho pequeno ou moderado.

Um fator de correção foi introduzido por Bartlett (1937), a idéia consiste em modificar a estatística por um fator produzindo uma estatística corrigida com o primeiro momento igual ao da distribuição χ^2 usada como referência. Lawley (1956) obteve uma fórmula geral para a constante que compõe o fator de correção de Bartlett, esta fórmula envolve cumulantes conjuntos do logaritmo da função de verossimilhança e derivadas de cumulantes. Outra técnica que visa o refinamento de métodos inferenciais são denominados correções do tipo-Bartlett. Essa correção produz uma estatística modificada por um fator, que é um polinômio na própria estatística cujos coeficientes são funções dos cumulantes conjuntos do logaritmo da função de verossimilhança. Podemos citar alguns trabalhos que utilizam essas técnicas de refinamentos em modelos lineares e não lineares da família exponencial com dispersão variável, dentre eles: correção de Bartlett às estatísticas da razão de verossimilhanças (CORDEIRO, 1993), escore (CORDEIRO; FERRARI, 1991), razão de verossimilhanças perfiladas modificadas (CYSNEIROS; FERRARI, 2006), razão de verossimilhanças bootstrap (ROCKE, 1989) e correção tipo-Bartlett à estatística escore (CAVALCANTI, 2009).

O objetivo principal desta dissertação é a obtenção de refinamentos de testes de hipóteses em MNLFE. Especificamente:

- Obter uma fórmula para o fator de correção tipo-Bartlett em notação matricial para a estatística gradiente nos MNLFE, para testar a presença de heteroscedasticidade nos dados.
- Comparar os desempenhos dos testes baseados nas estatísticas razão de verossimilhanças, razão de verossimilhanças perfiladas modificadas, razão de verossimilhanças bootstrap e suas versões corrigidas, escore, gradiente e a gradiente corrigida em que utilizamos os resultados obtidos por Vargas, Ferrari e Lemonte (2013), em termos de tamanho e poder em amostras finitas.
- Avaliar o impacto da quantidade de parâmetros de perturbação e interesse no desempenho dos testes supracitados acima.

1.2 Organização da dissertação

A presente dissertação está dividida em quatro capítulos e cinco apêndices. O Capítulo 2 introduzimos brevemente os modelos não lineares da família exponencial (MNLFE) e discutimos alguns aspectos inferenciais, tais como: função escore, informação de Fisher, estimação dos parâmetros e testes de hipóteses baseados nas estatísticas da razão de verossimilhanças, razão de verossimilhanças perfiladas modificadas, escore, gradiente e razão de verossimilhanças bootstrap. Avaliamos numericamente o comportamento desses diferentes testes em amostras finitas através do tamanho e poder. Além disso, analisamos a influência do número de parâmetros de

perturbação no desempenho dos testes e consideramos diferentes tamanhos amostrais. Algumas considerações finais são apresentadas neste capítulo.

No Capítulo 3, apresentamos a principal contribuição dessa dissertação, em que desenvolvemos a partir dos resultados obtidos por Vargas, Ferrari e Lemonte (2013) um fator de correção tipo-Bartlett à estatística gradiente nos MNLFE, em especial no modelo de regressão não linear normal e normal inverso. Apresentamos também, alguns resultados importantes sobre correção de Bartlett às estatísticas da razão de verossimilhanças (CORDEIRO, 1993), razão de verossimilhanças perfiladas modificadas (CYSNEIROS; FERRARI, 2006) e razão de verossimilhanças bootstrap (ROCKE, 1989). Resultados numéricos sobre o comportamento dos testes considerando diferentes números de parâmetros de interesse e perturbação em amostras finitas são apresentados. Além disso, avaliamos o desempenho dos testes em amostras de tamanho pequeno e moderado. Aplicações a dados reais também são apresentadas e são feitos alguns comentários no final deste capítulo.

No Capítulo 4 são apresentadas as conclusões desta dissertação. No Apêndice A destacamos alguns cumulantes conjuntos de derivadas do logaritmo da função de verossimilhanças e suas derivadas de até segunda ordem necessários para o cálculo do fator de correção tipo-Bartlett à estatística gradiente e no Apêndice B mostramos como obtemos a quantidade A_2 necessária para o cálculo do fator de correção tipo-Bartlett, as outras quantidades são obtidas de forma analógica e serão omitidas nesta dissertação.

Para elaboração desta dissertação utilizou-se o sistema tipográfico $\text{\LaTeX}$ desenvolvido por Leslie Lamport na década de 80. Todos os resultados das simulações foram obtidas usando a linguagem matricial de programação $\text{\Ox}$ em sua versão 7.01 para o sistema operacional Windows. Mais detalhes sobre a linguagem $\text{\Ox}$ podem ser obtidos em (DOORNIK, 2009) e em (CRIBARI-NETO; ZARKOS, 1999). As maximizações não lineares para o cálculo das estimativas de máxima verossimilhanças foram realizadas utilizando os algoritmos BFGS e SQP, ver Nocedal e Wright (1999).

Capítulo 2

Modelos não lineares da família exponencial

2.1 Introdução

Os modelos lineares generalizados (MLG) foram definidos por Nelder e Wedderburn (1972). Estes modelos foram introduzidos a partir dos modelos lineares normais, expandindo assim um leque de opções sobre a distribuição da variável resposta. Os MLG são definidos por variáveis aleatórias independentes que pertencem a família exponencial podendo ser estendidos também a uma família mais ampla de distribuições. Através de uma função de ligação relaciona-se a média da variável resposta com uma estrutura nos parâmetros linear ou não linear. Um caso especial abrangido pelo MLG e muito utilizado em pesquisas que envolvem diversas áreas como agronomia, econometria, engenharias, química, entre outras, são os modelos não lineares da família exponencial (MNLFE). Em várias aplicações do conhecimento observamos que a variância da variável resposta não é constante e assim dizemos que ela apresenta um comportamento heteroscedástico. Assim neste estudo iremos abordar os MNLFE com dispersão variável. Definimos, na Seção 2.2, os modelos não lineares da família exponencial. Na Seção 2.3, discutimos alguns aspectos inferenciais para esta classe de modelos. Testes de hipóteses usuais com erro de aproximação por qui-quadrado até de ordem $O(n^{-1})$ ¹ e um teste baseado em reamostragem bootstrap são apresentados na Seção 2.4. Por fim, realizamos estudos de simulação Monte Carlo para avaliar e comparar numericamente os desempenhos dos testes discutidos neste capítulo em amostras finitas de tamanho pequeno e moderado.

¹Se $\{a_n\}_{n \geq 1}$ e $\{b_n\}_{n \geq 1}$ são duas sequências de números reais, dizemos que $\{a_n\}$ é de ordem menor que $\{b_n\}$ e denotamos $\{a_n\} = o(b_n)$ se $\lim_{n \rightarrow \infty} a_n/b_n = 0$. Dizemos que a_n é de ordem no máximo igual a b_n , denotado por $\{a_n\} = O(b_n)$ se existe um número real $M > 0$ tal que, $|a_n/b_n| \leq M$, isto é, a razão $|a_n/b_n|$ é limitada. Observamos que se $\{a_n\} = o(b_n)$, então $\{a_n\} = O(b_n)$ e que quando $b_n \rightarrow 0$ a ordem fornece noção de taxa de convergência de a_n para zero.

2.2 Definição do modelo

Sejam $Y_1, \dots, Y_n$ n variáveis aleatórias independentes em que cada Y_ℓ tem função densidade na família exponencial da forma

$$\pi(y_\ell, \theta_\ell, \phi_\ell) = \exp \left\{ \phi_\ell [y_\ell \theta_\ell - b(\theta_\ell) - c(y_\ell)] - \frac{1}{2} a(y_\ell, \phi_\ell) \right\}, \quad \ell = 1, \dots, n \quad (2.1)$$

em que $a(\cdot, \cdot)$, $b(\cdot)$ e $c(\cdot)$ são funções conhecidas, θ_ℓ e $(\phi_\ell > 0)$ são denominados de parâmetros canônico e precisão, respectivamente. Aqui, assumimos que ϕ_ℓ é desconhecido e denominamos ϕ_ℓ^{-1} como parâmetro de dispersão. Admitimos aqui que $\phi_\ell^{-1} = \sigma^2 m_\ell$, sendo σ^2 uma constante desconhecida finita e estritamente positiva e $m_\ell = m(z_\ell, \delta) > 0$ o ℓ -ésimo elemento da matriz diagonal M com dimensão $n \times n$, z_ℓ é a ℓ -ésima linha da matriz Z com dimensão $n \times q$ de covariadas para modelar as flutuações da variância. A esperança e a variância da variável aleatória Y_ℓ do modelo (2.1) é dado por $\mathbb{E}[Y_\ell] = \mu_\ell = b'(\theta_\ell)$ e $\text{Var}[Y_\ell] = \phi_\ell^{-1} V_\ell$, $\ell = 1, \dots, n$, sendo $V_\ell = V_\ell(\mu_\ell) = d(\mu_\ell)/d(\theta_\ell)$ a função variância e $\theta_\ell = \int V_\ell^{-1} d(\mu_\ell) = q(\mu_\ell)$.

Nesta classe de modelos, a componente sistemática é definida como

$$g(\mu_\ell) = \eta_\ell = f(x_\ell, \beta),$$

em que $g(\cdot)$ é uma função conhecida, monótona e diferenciável, denominada função de ligação relacionando a média da variável resposta μ_ℓ com o preditor não linear η_ℓ , $f(\cdot; \cdot)$ é uma função contínua, diferenciável e não linear em relação as componentes dos parâmetros desconhecidos β 's com $\beta = (\beta_1, \dots, \beta_p)^\top$ sendo um vetor p -dimensional, com $p < n$ e $x_\ell = (x_{\ell 1}, \dots, x_{\ell p})^\top$ um vetor p de covariáveis conhecidas associadas à ℓ -ésima observação. Temos que a matriz de derivadas do preditor não linear em relação aos β 's é denotada por $X^* = X^*(\beta) = \partial \eta / \partial \beta^\top$ com posto p para todo β . Supomos identificabilidade tal que diferentes β 's implicam diferentes η 's, sendo assim a matriz X^* terá dimensão $n \times p$ cujo elementos são funções do vetor de parâmetros β desconhecidos a serem estimados. Em particular, $f(x_\ell, \beta) = x_\ell^\top \beta$ obtemos os modelos lineares generalizados. Para todos estudos nessa dissertação adotamos $g(\mu_\ell) = \mu_\ell$.

Assumindo $a(y_\ell, \phi_\ell) = s(\phi_\ell) + t(y_\ell)$, podemos reescrever (2.1) como

$$\pi(y_\ell, \theta_\ell, \phi_\ell) = \exp \left\{ -\frac{1}{2} [\phi_\ell d(y_\ell) + s(\phi_\ell) + t(y_\ell)] \right\}, \quad \ell = 1, \dots, n \quad (2.2)$$

em que $d(y_\ell) = d_\ell = -2[y_\ell \theta_\ell - b(\theta_\ell) - c(y_\ell)]$. Esta forma corresponde a um modelo da família exponencial na forma natural com os parâmetros canônicos ϕ_ℓ e $\phi_\ell \theta_\ell$. Por exemplo, a Tabela 2.1 apresenta as quantidades $s(\phi_\ell)$ e $d(y_\ell)$ para as distribuições normal, normal inversa e gama pertencentes a classe dos MNLFs. Para estes casos supomos que $s(\phi_\ell)$ tem as quatro primeiras derivadas.

Tabela 2.1: Quantidades d_ℓ e $s(\phi_\ell)$ para os modelos normal, normal inversa e gama.

	Normal	Normal inversa	Gama
d_ℓ	$(y_\ell - \mu_\ell)^2$	$\frac{(y_\ell - \mu_\ell)^2}{\mu_\ell^2 y_\ell}$	$2 \left[\frac{y_\ell}{\mu_\ell} - \log \left(\frac{y_\ell}{\mu_\ell} \right) \right]$
$s(\phi_\ell)$	$-\log(\phi_\ell)$	$-\log(\phi_\ell)$	$-2 [\phi_\ell \log(\phi_\ell) - \log \Gamma(\phi_\ell)]$

2.3 Alguns aspectos inferenciais

Sejam $Y_1, \dots, Y_n$ variáveis aleatórias independentes com função densidade de probabilidade dada em (2.1). O logaritmo da função de verossimilhança $\mathcal{L}$ para o modelo definido por (2.1) e (2.2) com o vetor de parâmetros $\psi = (\boldsymbol{\delta}^\top, \boldsymbol{\beta}^\top, \sigma^\top)^\top$ dadas as observações y_ℓ com $\ell = 1, \dots, n$ é dado por

$$\mathcal{L}(\boldsymbol{\delta}, \boldsymbol{\beta}, \sigma^2; y) = -\frac{1}{2} \sum_{\ell=1}^n \left\{ \frac{1}{\sigma^2 m_\ell} d(y_\ell) + s(\phi_\ell) + t(y_\ell) \right\}, \quad (2.3)$$

assumimos que $\mathcal{L}$ seja regular com respeito às derivadas em relação aos parâmetros até quarta ordem; ver Cox e Hinkley (1974).

A função escore $U = U(\boldsymbol{\delta}, \boldsymbol{\beta}, \sigma^2) = (U_{\boldsymbol{\delta}}^\top, U_{\boldsymbol{\beta}}^\top, U_{\sigma^2})^\top$ apresenta as seguintes componentes:

$$U_{\boldsymbol{\delta}} = \frac{\partial \mathcal{L}(\boldsymbol{\delta}, \boldsymbol{\beta}, \sigma^2; y)}{\partial \delta^a} = \frac{1}{2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \frac{1}{m_\ell} \frac{\partial m_\ell}{\partial \delta^a} \left\{ d_\ell + \dot{s} \left(\frac{1}{\sigma^2 m_\ell} \right) \right\},$$

$$U_{\boldsymbol{\beta}} = \frac{\partial \mathcal{L}(\boldsymbol{\delta}, \boldsymbol{\beta}, \sigma^2; y)}{\partial \beta^i} = \sum_{\ell=1}^n (y_\ell - \mu_\ell) \frac{1}{\sigma^2 m_\ell} \frac{d\theta_\ell}{d\mu_\ell} \frac{d\mu_\ell}{d\eta_\ell} x_{\ell i}^* \quad \text{e}$$

$$U_{\sigma} = \frac{\partial \mathcal{L}(\boldsymbol{\delta}, \boldsymbol{\beta}, \sigma^2; y)}{\partial \sigma^2} = \frac{1}{2\sigma^2} \sum_{\ell=1}^n \frac{1}{\sigma^2 m_\ell} \left\{ d_\ell + \dot{s} \left(\frac{1}{\sigma^2 m_\ell} \right) \right\}.$$

Em notação matricial, obtemos:

$$U = U(\boldsymbol{\delta}, \boldsymbol{\beta}, \sigma^2) = \begin{pmatrix} \frac{1}{2} \Psi^\top \Phi (\mathbf{d} + \dot{S}) \\ X^{*\top} \Psi T V^{-1} (\mathbf{y} - \boldsymbol{\mu}) \\ \frac{1}{2\sigma^2} \mathbf{1}^\top \Psi (\mathbf{d} + \dot{S}) \end{pmatrix},$$

em que Φ, T, V são matrizes diagonais de dimensão $n \times n$, cujos elementos são $\phi_\ell = 1/(\sigma^2 m_\ell)$, $t_\ell = d\mu_\ell/d\eta_\ell$ e $v_\ell = d\mu_\ell/d\theta_\ell$, respectivamente, Ψ é uma matriz de dimensão $n \times p$ com a ℓ -ésima linha dada por $\partial \log m(z_\ell, \boldsymbol{\delta}) / \partial \boldsymbol{\delta}^\top$, $\dot{S} = (\dot{s}(\phi_1), \dots, \dot{s}(\phi_n))^\top$ sendo $\dot{s}(\phi_\ell) = ds(\phi_\ell)/d\phi_\ell$ com $\ell = 1, \dots, n$, $\mathbf{1}$ é um vetor de uns n -dimensional, $\mathbf{d} = (d_1, \dots, d_n)^\top$ e $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^\top$.

Vimos na Tabela 2.1 apresentada na Seção 2.2 a quantidade $s(\phi_\ell)$ para as distribuições normal, normal inversa e gama. Assim é fácil ver que $\dot{S} = (-1/\phi_1, \dots, -1/\phi_n)$ para as distriuições normal e normal inversa. O estimador de máxima verossimilhança (EMV) de ψ é obtido através da solução de sistemas de equações dado por

$$U_{\boldsymbol{\delta}} = \mathbf{0}, \quad U_{\boldsymbol{\beta}} = \mathbf{0} \text{ e } U_{\sigma^2} = 0.$$

Como estamos trabalhando com modelos não lineares, não é possível expressar os estimadores de máxima verossimilhança em uma forma analítica fechada, portanto utilizamos métodos iterativos de otimização não linear como, por exemplo, Newton-Raphson, escore de Fisher, quasi-Newton BFGS (Broyden-Fletcher-Goldfarb-Shanno), SQP (Sequential quadratic programming) e entre outros. Ver Nocedal e Wright (1999). A linguagem de programação Ox disponibiliza no seu pacote de maximização o comando `MAXSQP`. Este comando maximiza funções não lineares sujeitas às restrições não lineares usando uma técnica de programação quadrática sequencial, que, basicamente, consiste em gerar passos (iterações) resolvendo subproblemas quadráticos. Maiores detalhes sobre o comando `MAXSQP` podem ser encontrados em:

<http://www.nuff.ox.ac.uk/Users/Doornik/doc/ox/index.html>.

Pode ser mostrado facilmente que

$$\mathbb{E}[c(y_\ell)] = \theta_\ell b'(\theta) - b(\theta) - \frac{1}{2} \dot{s} \left(\frac{1}{\sigma^2 m_\ell} \right)$$

e através das condições de regularidade que $\mathbb{E}(U_\sigma^2) = 0$. Sendo assim temos

$$\begin{aligned} \mathbb{E}(d_\ell) &= \mathbb{E}\{-2[y_\ell \theta_\ell - b(\theta_\ell) - c(y_\ell)]\} \\ &= -2\{\theta_\ell \mathbb{E}(y_\ell) - b(\theta_\ell) - \mathbb{E}[c(y_\ell)]\} \\ &= -2\theta_\ell b'(\theta) + 2b(\theta_\ell) + 2\left[\theta_\ell b'(\theta) - b(\theta_\ell) - \frac{1}{2} \dot{s} \left(\frac{1}{\sigma^2 m_\ell} \right)\right] \\ &= -\dot{s} \left(\frac{1}{\sigma^2 m_\ell} \right). \end{aligned} \quad (2.4)$$

A matriz de informação total de Fisher para $\psi = (\boldsymbol{\delta}^\top, \boldsymbol{\beta}^\top, \sigma^2)^\top$ na classe dos modelos não lineares da família exponencial é dada por

$$K = -\mathbb{E} \left(\frac{\partial \mathcal{L}}{\partial \psi \partial \psi^\top} \right) = \begin{pmatrix} \frac{1}{2} \Psi^\top \Phi \ddot{S} \Phi \Psi & 0 & \frac{1}{2\sigma^2} \Psi^\top \ddot{S} \Phi^2 \mathbf{1} \\ 0 & X^{*\top} \Phi^{1/2} W \Phi^{1/2} X^* & 0 \\ \frac{1}{2\sigma^2} \mathbf{1}^\top \Phi^2 \ddot{S} \Psi & 0 & \frac{1}{2\sigma^4} \mathbf{1}^\top \Phi^2 \ddot{S} \mathbf{1} \end{pmatrix}, \quad (2.5)$$

em que $\ddot{S}$ e W são matrizes diagonais $n \times n$ cujos respectivos elementos são dados por $\ddot{s}(\phi_\ell) = d^2 s(\phi_\ell) / d\phi_\ell^2$ e $w_\ell = (d\mu_\ell / d\eta_\ell)^2 / v_\ell$, $\ell = 1, \dots, n$ e $\mathbf{0}$ representa uma matriz de zeros ou vetores nulos com dimensões apropriadas.

Podemos notar da matriz de informação de Fisher (2.5) que os parâmetros $(\boldsymbol{\delta}^\top, \sigma^2)$ são

ortogonais a β , isto é, $-\mathbb{E}(\partial\mathcal{L}/\partial\delta^a\partial\beta^i)$ e $-\mathbb{E}(\partial\mathcal{L}/\partial\sigma^2\partial\beta^i)$ são iguais a zero, entretanto δ e σ^2 não são ortogonais. A fim de obter a ortogonalidade entre os parâmetros uma transformação é necessária, assim uma solução para tal nas distribuições normal e normal inversa vem da seguinte igualdade:

$$\sigma^2 = \frac{\gamma}{(\prod_{\ell=1}^n m_\ell)^{1/n}}, \quad (2.6)$$

ver Simonoff e Tsai (1994).

Agora, o nosso vetor de parâmetros que pertence a um espaço paramétrico é dado por $\psi = (\delta, \beta, \gamma)^\top$. Logo, utilizando a transformação em (2.6) o logaritmo da função de verossimilhança do modelo reparametrizado fica dado por

$$\mathcal{L}^*(\psi; y) = -\frac{1}{2} \sum_{\ell=1}^n \left\{ \frac{q_\ell}{\gamma} d(y_\ell) + t(y_\ell) + \log \left(\frac{\gamma}{q_\ell} \right) \right\} \quad (2.7)$$

em que

$$q_\ell = q_\ell(\delta) = \frac{(\prod_{s=1}^n m_s)^{1/n}}{m_\ell}.$$

A função escore total $U = U(\delta, \beta, \gamma) = (U_\delta^\top, U_\beta^\top, U_\gamma)^\top$ para o modelo reparametrizado (2.7) tem componentes dadas por

$$U = U(\delta, \beta, \gamma) = \begin{pmatrix} -\frac{1}{2\gamma} \dot{Q}^\top \mathbf{d} \\ X^{*\top} \Phi T V^{-1} (\mathbf{y} - \boldsymbol{\mu}) \\ \frac{1}{2\gamma} (\Phi \mathbf{d} - \mathbf{1}) \end{pmatrix}.$$

Aqui, Φ é uma matriz diagonal $n \times n$ cujos os seus elementos são dados por $\phi_\ell = q_\ell/\gamma$ e $\dot{Q}$ é uma matriz $n \times q$ cuja a ℓ -ésima linha é dada por $q_{\ell a} = \partial q_\ell / \partial \delta^a$. Então, a matriz de informação de Fisher do modelo reparametrizado (2.7) é dada por

$$K = -\mathbb{E} \left(\frac{\partial \mathcal{L}^*}{\partial \psi \partial \psi^\top} \right) = \begin{pmatrix} \frac{1}{2} \mathbf{1}^\top \ddot{Q} Q^{-1} \mathbf{1} & 0 & 0 \\ 0 & X^{*\top} \Phi^{1/2} W \Phi^{1/2} X^* & 0 \\ 0 & 0 & \frac{n}{2\gamma^2} \end{pmatrix}, \quad (2.8)$$

$\ddot{Q} = \ddot{Q}(\delta)$ é um “array” de dimensão $n \times q \times q$ com elementos dados por $q_{\ell ab} = \partial^2 / (\partial \delta^a \partial \delta^b)$; ver Apêndice C e Q^{-1} a matriz inversa da matriz diagonal $Q = Q(\delta)$ de dimensão $n \times n$ com o ℓ -ésimo elemento dado por $q_\ell = q_\ell(\delta) = (\prod_{s=1}^n m_s)^{1/n} / m_\ell$. Observando a matriz (2.8) e aplicando a transformação dada em (2.6) atingimos a ortogonalidade desejada obtemos $\mathbb{E}(-\partial\mathcal{L}/\partial\delta^a\partial\beta^i) = \mathbb{E}(-\partial\mathcal{L}/\partial\beta^i\partial\gamma) = \mathbb{E}(-\partial\mathcal{L}/\partial\delta^a\partial\gamma) = 0$, ou seja, os parâmetros são todos ortogonais.

Nesta dissertação, abordamos para o vetor de observações $(y_1, \dots, y_n)$ apenas a distribui-

ção normal. Desse modo, do resultado (2.4) e após a reparametrização temos que $\mathbb{E}(d_\ell) = -\dot{s}(q_\ell/\gamma) = -(-\gamma/q_\ell) = \gamma/q_\ell$ e da matriz (2.5) temos $\ddot{S} = \Phi^2$ e $W = I_n$ é a matriz identidade de ordem n .

2.4 Testes de hipóteses

Após a estimação dos parâmetros é comum realizar testes afim de avaliar se as hipóteses feitas sobre os parâmetros apresentam evidências obtidas através de amostras. Nosso interesse neste capítulo é realizar testes usuais para identificar se há presença de heteroscedasticidade na classe dos modelos não lineares da família exponencial. Por isso abordamos a situação em que os parâmetros de dispersão não são constantes para todas as observações, apresentando assim uma estrutura heteroscedástica. Usualmente, em alguns modelos há interesse em testar apenas um subconjunto de parâmetros ao invés de todos os parâmetros, desta forma é conveniente fazer uma partição no vetor de parâmetros. Denominamos *parâmetros de interesse* aqueles que desejamos testar e os demais denominamos *parâmetros de perturbação*. Nesta dissertação o interesse é testar o vetor de parâmetros δ , portanto este será nosso vetor de parâmetros de interesse que será testado pela hipótese nula e os demais β e γ são considerados como parâmetros de perturbação. Realizamos todas estimações e testes baseados no modelo reparametrizado definido em (2.7). Sendo assim, o número de parâmetros de interesse e de perturbação são q e $(p+1)$, respectivamente. Aqui, o nosso interesse está em testar a hipótese nula $\mathcal{H}_0 : \delta = \delta_0$ $\mathcal{H}_1 : \delta \neq \delta_0$, sendo δ_0 um vetor q -dimensional de constantes especificadas tal que $m(z_\ell, \delta) = 1$ para $\ell = 1, \dots, n$.

Considerando a heteroscedasticidade multiplicativa, isto é, $m_\ell = \exp\{z_\ell^\top \delta\}$, obtemos $q_\ell = \exp\left\{-(z_\ell - \bar{z})^\top \delta\right\}$, $q_{\ell a} = -(z_\ell - \bar{z})_a \exp\left\{-(z_\ell - \bar{z})^\top \delta\right\} = -(z_\ell - \bar{z})_a q_\ell = (z_{\ell a} - \bar{z}_a) q_\ell$ e $q_{\ell ab} = (z_\ell - \bar{z})_{ab} q_\ell = (z_{\ell a} - \bar{z}_a)(z_{\ell b} - \bar{z}_b) q_\ell$, com $\bar{z} = (\bar{z}_1, \dots, \bar{z}_q)^\top$ e $\bar{z}_a = n^{-1} \sum_{\ell=1}^n z_{\ell a}$ para $a = 1, \dots, q$. Usando estas quantidades a matriz de informação de Fisher do modelo repametrizado (2.7) se reduz a

$$K = -\mathbb{E} \left(\frac{\partial \mathcal{L}^*}{\partial \psi \partial \psi^\top} \right) = \begin{pmatrix} \frac{1}{2} (Z - \bar{Z}) (Z - \bar{Z})^\top & 0 & 0 \\ 0 & X^{*\top} W \Phi X^* & 0 \\ 0 & 0 & \frac{n}{2\gamma^2} \end{pmatrix} \quad (2.9)$$

em que $(Z - \bar{Z}) = (z_1 - \bar{z}, \dots, z_n - \bar{z})^\top$.

Denota-se $K_\delta = \mathbb{E}\{U(\delta) U(\delta)^\top\} = \frac{1}{2} (Z - \bar{Z}) (Z - \bar{Z})^\top$ e $K_\beta = \mathbb{E}\{U(\beta) U(\beta)^\top\} = X^{*\top} W \Phi X^*$ como sendo as matrizes de informação de Fisher de δ e β , respectivamente.

2.4.1 Teste da razão de verossimilhanças

A estatística da razão de verossimilhanças para testar $\mathcal{H}_0$ no modelo reparametrizado (2.7) pode ser escrita da seguinte forma:

$$\text{LR} = 2\{\mathcal{L}^*(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}, \hat{\gamma}; \mathbf{y}) - \mathcal{L}^*(\boldsymbol{\delta}_0, \tilde{\boldsymbol{\beta}}, \tilde{\gamma}; \mathbf{y})\}, \quad (2.10)$$

em que $\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}, \hat{\gamma}$ são os EMV irrestrito, respectivamente, de $\boldsymbol{\delta}, \boldsymbol{\beta}, \gamma$ e $\tilde{\boldsymbol{\beta}}, \tilde{\gamma}$ são os EMV restritos de $\boldsymbol{\beta}, \gamma$, respectivamente. Sob $\mathcal{H}_0$, a estatística dada em (2.10) tem distribuição assintótica χ_q^2 . Rejeitamos a hipótese nula, ao nível de significância α se, $\text{LR} \geq \chi_{(\alpha, q)}^2$, em que $\chi_{(\alpha, q)}^2$ é o percentil $(1 - \alpha)$ da distribuição qui-quadrado. Sabemos que a aproximação das estatísticas LR, S, LR_m e G pela distribuição χ^2 em amostras pequenas pode conduzir a taxas de rejeição distorcidas e os testes baseados nessas estatísticas tendem a ter comportamento liberal, isto é, tendem a rejeitar a hipótese nula com frequência demasiada ou conservativo, uma vez que nessa situação erro tipo I ocorre com baixa probabilidade.

2.4.2 Teste escore

A estatística escore proposta originalmente por Rao (1947) para testar $\mathcal{H}_0$ é expressa por:

$$S = \tilde{U}_{\boldsymbol{\delta}}^\top K_{\boldsymbol{\delta}}^{-1} \tilde{U}_{\boldsymbol{\delta}},$$

em que $\tilde{U}_{\boldsymbol{\delta}} = -\frac{1}{2\gamma} \dot{Q}^\top \mathbf{d}$, $K_{\boldsymbol{\delta}}^{-1} = \left[\frac{1}{2} (Z - \bar{Z}) (Z - \bar{Z})^\top \right]^{-1}$ a matriz inversa da informação de Fisher para $\boldsymbol{\delta}, \mathbf{y} = (y_1, \dots, y_\ell)^\top$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^\top$, em que $\mu_\ell = \eta_\ell = f(x_\ell, \boldsymbol{\beta})$, para $\ell = 1, \dots, n$. Utilizamos notação til em $\tilde{U}_{\boldsymbol{\delta}}$ para denotar que o EMV do parâmetro de interesse está sendo avaliado sob $\mathcal{H}_0$.

Assintoticamente e sob $\mathcal{H}_0$, temos que a estatística escore tem distribuição χ_q^2 . Rejeitamos a hipótese nula, ao nível de significância α se, $S \geq \chi_{(\alpha, q)}^2$, em que $\chi_{(\alpha, q)}^2$ é o percentil $(1 - \alpha)$ da distribuição qui-quadrado.

2.4.3 Teste gradiente

A estatística gradiente recentemente proposta por Terrell (2002) vem como uma estatística alternativa para a realização de testes de hipóteses, notavelmente mais simples que algumas estatísticas de testes usuais. Assim como a estatística da razão de verossimilhanças proposta originalmente por Neyman e Pearson (1928) e discutida com mais detalhes por Wilks (1938), a estatística Wald proposta por Wald (1943) e a estatística escore proposta por Rao (1973), a estatística gradiente também tem distribuição assintótica χ_q^2 .

Seja um vetor de observações $(y_1, \dots, y_n)$ com função de densidade dada em (2.2) que depende de um vetor de parâmetros e possui logaritmo da função de verossimilhança dada em

(2.7), a estatística gradiente para testar $\mathcal{H}_0$ é dada por:

$$G = \tilde{U}_{\delta}(\hat{\delta} - \delta_0), \quad (2.11)$$

em que $\tilde{U}_{\delta} = -\frac{1}{2\gamma} \dot{Q}^{\top} \mathbf{d}$, sendo $\tilde{\beta}$ e $\tilde{\gamma}$ os EMV restritos de β e γ , respectivamente. Vale ressaltar que a estatística gradiente é atrativa pela sua simplicidade de cálculo e pode ser facilmente computada, além disso apresenta vantagem em problemas complexos quando não é possível calcular, estimar ou inverter a matriz de informação de Fisher. Rejeitamos a hipótese nula, ao nível de significância α se, $G \geq \chi^2_{(\alpha, q)}$, em que $\chi^2_{(\alpha, q)}$ é o percentil $(1 - \alpha)$ da distribuição qui-quadrado.

2.4.4 Teste da razão de verossimilhanças perfiladas modificadas

Uma função de verossimilhança perfilada é utilizada em modelos em que há presença de parâmetros de perturbação e esta função dependerá apenas do vetor de parâmetros de interesse. A verossimilhança perfilada é obtida da verossimilhança original substituindo os parâmetros de perturbação por seus estimadores de máxima verossimilhança para cada valor fixado dos parâmetros de interesse. Assim o logaritmo da função de verossimilhança perfilada para o parâmetro de interesse δ é dado por

$$\mathcal{L}_p(\delta; \mathbf{y}) = \mathcal{L}_p(\delta, \hat{\beta}_{\delta}, \hat{\sigma}_{\delta}^2; \mathbf{y}), \quad (2.12)$$

em que $\hat{\beta}_{\delta}$ e $\hat{\sigma}_{\delta}^2$ são os EMVs de β e σ^2 , respectivamente, para δ fixado. Considerando o modelo reparametrizado podemos reescrever (2.12) como

$$\mathcal{L}_p^*(\delta; \mathbf{y}) = \mathcal{L}_p(\delta, \hat{\beta}_{\delta}, \hat{\gamma}_{\delta}; \mathbf{y}).$$

Considerando um modelo normal não linear reparametrizado o logaritmo da função de verossimilhança perfilada é dado por

$$\mathcal{L}_p^* = \mathcal{L}_p^*(\delta; \mathbf{y}) = -\frac{n}{2} \log \hat{\gamma}_{\delta} - \frac{1}{2\hat{\gamma}_{\delta}} (\mathbf{y} - \hat{\mu})^{\top} Q (\mathbf{y} - \hat{\mu}) - \frac{n}{2} \log(2\pi),$$

sendo $\hat{\gamma}_{\delta}$ o EMV de γ , com δ fixo e $\mathcal{L}_p^*$ denota a função de verossimilhança perfilada do modelo reparametrizado. Esta função apresenta algumas propriedades da função de verossimilhança original, porém pode apresentar vícios na função escore e matriz de informação de Fisher. Além disso a presença de um número de parâmetros de perturbação pode conduzir a uma má qualidade em inferências que se baseiam em resultados assintóticos. Assim, a fim de corrigir este efeito dos parâmetros de perturbação Cox e Reid (1987) definiram uma versão modificada explorando a ortogonalidade entre os parâmetros de interesse e perturbação. Para tal, utilizaram a transformação (2.6) e em seguida um termo de ajuste foi incorporado a função de verossimilhança perfilada. Sendo assim, o logaritmo da função de verossimilhança perfilada modificada para δ é dado por

$$\mathcal{L}_{mp}^* = \mathcal{L}_p^*(\boldsymbol{\delta}; \mathbf{y}) - \frac{1}{2} \log \left[\det \left\{ j^*(\boldsymbol{\delta}, \hat{\boldsymbol{\beta}}_\delta, \hat{\gamma}_\delta) \right\} \right],$$

em que a notação $\mathcal{L}_{mp}^*$ representa o logaritmo da função de verossimilhança perfilada modificada do modelo reparametrizado e j^* é o bloco da matriz de informação observada definida como $J = -\partial U_\delta^\top / \partial \delta = -\partial^2 \mathcal{L}_p^* / \partial^a \delta^b$. Lembrando que tomando a esperança da matriz de informação observada obtemos a matriz de informação de Fisher, isto é, $\mathbb{E}(J) = K$. Para o modelo normal não linear obtemos

$$j^*(\boldsymbol{\delta}, \hat{\boldsymbol{\beta}}_\delta, \hat{\gamma}_\delta) = \begin{pmatrix} \frac{1}{\hat{\gamma}_\delta} \{ X^{*\top} W^{1/2} Q W^{1/2} X^* - [(\mathbf{y} - \boldsymbol{\mu})^\top Q][X^{**}] \} & 0 \\ 0 & \frac{n}{2\hat{\gamma}_\delta^2} \end{pmatrix},$$

em que $W = I_n$ por se tratar de um modelo normal, $X^{**} = X^{**}(\boldsymbol{\delta}) = \partial^2 \boldsymbol{\eta} / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top$ é um “array” de dimensão $n \times p \times p$ e $[\cdot][\cdot]$ representa a multiplicação de uma matriz por um array; ver Wei (1998).

A estatística da razão de verossimilhanças perfiladas modificadas para testar $\mathcal{H}_0$ é dada por:

$$\text{LR}_m = 2\{\mathcal{L}_{mp}^*(\tilde{\boldsymbol{\delta}}; \mathbf{y}) - \mathcal{L}_{mp}^*(\boldsymbol{\delta}_0; \mathbf{y})\},$$

em que $\tilde{\boldsymbol{\delta}}$ é o EMV restrito de $\boldsymbol{\delta}$.

Sob $\mathcal{H}_0$ a estatística LR_m tem distribuição assintoticamente χ_q^2 . Rejeitamos a hipótese nula, ao nível de significância α se, $\text{LR}_m \geq \chi_{(\alpha, q)}^2$, em que $\chi_{(\alpha, q)}^2$ é o percentil $(1 - \alpha)$ da distribuição qui-quadrado. No entanto, esta aproximação pode não ser muito boa quando o tamanho da amostra é pequeno resultando em taxas de rejeição distorcida, geralmente, apresentando comportamento liberal. Por esse motivo se faz necessário uso de técnicas estatísticas para melhorar essa aproximação. Essas técnicas serão apresentadas e avaliadas por simulação Monte Carlo no Capítulo 3.

2.4.5 Teste da razão de verossimilhanças bootstrap

Uma alternativa para melhorar testes baseados nas estatísticas razão de verossimilhança, Wald e Escore é utilizar um esquema de reamostragem pra estimar a distribuição nula da estatística de teste. Essa metodologia é conhecida como bootstrap. A técnica bootstrap introduzida por Efron (1979) e mais detalhada para testes de hipóteses em Efron e Tibshirani (1994) utiliza um valor crítico a partir da distribuição nula da estatística de teste estimada pela reamostragem bootstrap. Essa metodologia consiste em obter a partir da amostra original $\mathbf{y} = (y_1, \dots, y_n)$ um número grande (por exemplo, B) de pseudos-amostras, a qual denominamos de amostras bootstrap $\mathbf{y}^* = (y^{*1}, \dots, y^{*B})^\top$. Estas amostras artificiais são geradas impondo como verdadeira a hipótese nula. Posteriormente computa-se para cada amostra $\mathbf{y}^*$ a estatística de teste denotada por τ^* , assim, obtem-se τ^{*b} , para $b = 1, \dots, B$. As estatísticas de bootstrap são então utilizadas para estimar a distribuição nula de τ^* . O valor crítico bootstrap denotado por $\hat{q}_{(1-\alpha)}$ correspondente ao nível de significância α ($0 < \alpha < 1$) e é obtido como o percentil $(1 - \alpha)$ da distribuição empírica das B realizações da estatística τ^* . Assim, a hipótese nula é rejeitada se $\tau \geq \hat{q}_{(1-\alpha)}$.

Note que o teste da razão de verossimilhanças bootstrap não usa um valor crítico obtido da distribuição nula assintótica, ele utiliza o valor crítico estimado da distribuição nula de τ .

Para o teste da razão de verossimilhança bootstrap a estatística de teste bootstrap é dada por

$$\text{LR}_{\text{Boot}} = 2\{\mathcal{L}^*(\hat{\psi}^{*b}; \mathbf{y}^{*b}) - \mathcal{L}^*(\tilde{\psi}^{*b}; \mathbf{y}^{*b})\}, \quad (2.13)$$

em que $\hat{\psi}^{*b}$ e $\tilde{\psi}^{*b}$ são os EMV irrestrito e restrito de ψ^* , respectivamente, na b -ésima réplica bootstrap, com $b = 1, \dots, B$. Aqui, rejeita-se $\mathcal{H}_0$ se a estatística $\text{LR}_{\text{Boot}} \geq \hat{q}_{(1-\alpha)}$.

2.5 Avaliação numérica

Neste capítulo apresentamos resultados da simulação Monte Carlo comparando os desempenhos dos testes baseados nas estatísticas da razão de verossimilhanças (LR), razão de verossimilhanças perfiladas modificadas (LR_m), razão de verossimilhanças bootstrap (LR_{Boot}), escore (S) e gradiente (G). Avaliamos estes desempenhos através da proximidade das probabilidades de rejeição da hipótese nula, sendo esta verdadeira (erro tipo I) com os respectivos níveis nominais 10, 0%, 5, 0%, 1, 0% e 0, 5% e o poder dos testes levando em consideração a hipótese alternativa para diferentes valores de δ . As hipóteses consideradas foram:

$$\begin{aligned} \mathcal{H}_0: \delta_1 = \delta_2 = \dots = \delta_q = 0 \text{ (modelo homoscedástico)} \\ \text{versus} \\ \mathcal{H}_1: \delta_1 \neq \delta_2 \neq \dots \neq \delta_q \neq 0 \text{ (modelo heteroscedástico).} \end{aligned}$$

Consideramos o seguinte modelo de regressão não-linear para a simulação

$$y_\ell = \beta_1 + \exp\{\beta_2 x_{\ell 2}\} + \sum_{j=3}^p \beta_j x_{\ell j} + u_\ell, \quad \ell = 1, 2, \dots, n,$$

em que $u_\ell \sim \mathcal{N}(0, 1)$ e $\text{Cov}(u_\ell, u_m)$ para todo $\ell \neq m$. As covariadas x_j foram geradas como amostras da aleatórias da distribuição $\mathcal{U}(0, 1)$. A variável resposta foi gerada assumindo $\beta_1 = \beta_2 = \dots = \beta_p = 1$ e $\sigma^2 = 1$. Usamos diferentes valores de p e q a fim de avaliar o desempenho dos testes quando aumentamos o número de parâmetros de interesse e perturbação. Para a modelagem da flutuação da variância construímos a matriz $\mathbf{Z}$ utilizando as colunas 2, 3, $\dots, p$ da matriz $\mathbf{X}$ de covariadas, em que $q < p$ e para o caso, $q \geq k$ as colunas extras foram geradas usando amostras aleatórias independentes da distribuição $\mathcal{U}(0, 1)$. O número de réplicas Monte Carlo e bootstrap foram fixadas em 10000 e 600, respectivamente. Para cada tamanho amostra e nível nominal considerado, estimamos, via simulação, as taxas de rejeição: $P(\text{LR} \geq \chi_{(\alpha, q)}^2)$, $P(\text{LR}_m \geq \chi_{(\alpha, q)}^2)$, $P(S \geq \chi_{(\alpha, q)}^2)$, $P(G \geq \chi_{(\alpha, q)}^2)$, tal que $\chi_{(\alpha, q)}^2$ é o percentil $(1 - \alpha)$ da distribuição qui-quadrado e para o teste baseado na LR_{Boot} , as taxas de rejeição são obtidas calculando a $P(\text{LR}_{\text{Boot}} \geq \hat{q}_{(1-\alpha)})$, em que $\hat{q}_{(1-\alpha)}$ é estimado de acordo com o procedimento descrito na Subseção 2.4.5. As simulações foram realizadas utilizando a linguagem de programação matricial Ox (DOORNIK, 2009).

Nesta seção, inicialmente, analisamos a influência dos parâmetros de perturbação e do tamanho amostral no desempenho dos testes no modelo normal não linear. Em seguida, avaliamos também o poder dos testes, considerando diferentes valores para δ . Obviamente, para $\delta = 0$ temos o tamanho do teste. Todos os resultados apresentados aqui baseiam-se no modelo de regressão não linear repametrizado com logaritmo da função de verossimilhança definida em (2.7).

A Tabela 2.2 apresenta os tamanhos estimados dos testes para a situação em que $n = 35$, $q = 2$ e diferentes valores de p . Podemos observar que os testes da razão de verossimilhanças e gradiente apresentam taxas de rejeição muito acima dos níveis nominais considerados, principalmente quando o número de parâmetros de perturbação p aumenta, tornando-os liberais. Comparando estes dois testes, notamos que o teste gradiente tem melhor desempenho que o teste da razão de verossimilhanças. Por exemplo, para $p = 4$ e $\alpha = 5,0\%$ as taxas de rejeição para os testes LR e G são iguais a 9,9% e 9,0%, respectivamente. Por outro lado, os testes LR_m e S, para o mesmo caso, apresentam taxas de rejeição iguais a 4,2% e 3,8%, respectivamente, apresentando taxas de rejeição mais próximas dos níveis nominais, todavia apresentam comportamento conservativo. Vale ressaltar que a modificação de Cox e Reid à verossimilhança perfilada atenua o efeito do número de parâmetros de perturbação, trazendo uma notável melhora para o teste LR_m comparado ao teste LR. Notamos que o aumento no número de parâmetros de perturbação o impacto é bem menos marcante no desempenho dos testes baseados nas estatísticas LR_m , LR_{Boot} e S, mas tratando-se dos testes LR e G o impacto é consideravelmente alto sobre a aproximação por χ^2 para as distribuições das estatísticas baseadas nestes testes. Aqui, o teste da razão de verossimilhança bootstrap foi o que apresentou melhor desempenho, com taxas de rejeição bem mais próximas dos níveis nominais que os demais testes, como por exemplo, para $p = 3$ e para todos os níveis nominais considerados. Para este teste, podemos observar que ele passa a ser liberal quando $p \geq 5$, mas não tão distante dos níveis nominais.

A Figura 2.1 apresenta o gráfico de distorção dos testes, ou seja, a diferença entre a taxa de rejeição empírica e o nível nominal $\alpha = 5,0\%$ para diferentes valores de p . Este gráfico mostra o impacto do número de parâmetros de perturbação nos testes considerados neste capítulo. Vemos, claramente, que à medida que aumentamos o número de parâmetros de perturbação as taxas de rejeição para os testes LR_m e G distorcem do nível nominal 5,0% tornando extremamente liberais e para os testes LR_m , LR_{Boot} e S as taxas de rejeição não sofrem tanto impacto quando aumentamos o número de parâmetros de perturbação, confirmando graficamente o que já havíamos mencionado anteriormente.

Na Tabela 2.3, fixamos o valor de $p = 5$, $q = 2$ e variamos o tamanho da amostra em $n = 25, 35, 45, 100$ para os níveis nominais $\alpha = 10,0\%, 5,0\%, 1,0\%$ e $0,5\%$. Observe que com o aumento do tamanho amostral, as taxas de rejeição dos testes se aproximam dos respectivos níveis nominais, todavia, os testes LR e G continuam liberais. Assim, até mesmo para $n = 100$, os testes LR e G apresentam taxas de rejeição acima dos níveis nominais. Nesta mesma situação e $\alpha = 5,0\%$, as taxas de rejeição dos testes LR, LR_m , LR_{Boot} , S e G foram 8,2%, 4,5%, 5,3%, 4,5%, 7,8%, respectivamente. Dessa maneira, é necessário obter uma correção para as estatísticas LR e G para que, em amostras de tamanho pequeno haja uma diminuição na

distorção das taxas de rejeição empíricas nos testes baseados nestas estatísticas modificadas.

A Tabela 2.4 apresenta resultados da simulação considerando a hipótese alternativa (heteroscedasticidade) para $n = 35$, $p = 5$ e $q = 2$ e diferentes valores de $\delta = \delta_1 = \delta_2$, com δ variando de 0 a 3, 5. Podemos observar que o poder do testes LR e G são maiores que os outros testes, entretando, os testes baseados nas estatísticas razão de verossimilhanças usual e gradiente não podem ser recomendados por serem muito liberais. Considerando $\delta = 3, 5$ e $\alpha = 10, 0\%$ o poder do teste para a LR_m , LR_{Boot} e S foram, respectivamente, 98, 3, 98, 4 e 96, 0. Vemos uma pequena desvantagem da estatística escore com relação as estatísticas LR_m e LR_{Boot} e portanto, o poder dos testes baseados nas estatísticas razão de verossimilhanças perfiladas modificadas e razão de verossimilhanças bootstrap são similares, neste caso.

A Figura 2.2 apresenta o gráfico das discrepâncias relativas entre os quantis amostrais e os quantis assintóticos das estatísticas dos testes contra os quantis assintóticos para todos os níveis nominais $\alpha = 15, 0\%$, $10, 0\%$, $5, 0\%$, $3, 0\%$, $1, 0\%$ e $0, 5\%$. A discrepância relativa é definida como

$$\frac{ST(1 - \alpha) - \chi_q^2(1 - \alpha)}{\chi_q^2(1 - \alpha)},$$

em que denotamos por $ST(1 - \alpha)$ o quantil amostral de ordem $(1 - \alpha)$ do conjunto de valores simulados das estatísticas dos testes (LR, LR_m , S e G, conforme o caso) e o corresponde quantil da distribuição χ_q^2 sendo denotado por $\chi_q^2(1 - \alpha)$. Quanto mais próxima de zero estiver a curva da discrepância, melhor a aproximação da distribuição exata sob a hipótese nula da estatística de teste pela distribuição limite qui-quadrado. Observe que para o caso $\alpha = 5, 0\%$, $n = 35$, $p = 5$ e $q = 2$ a discrepância relativa para as estatísticas LR e G apresentam-se em torno de 43, 3% $((8, 6 - 6, 0)/6, 0 \times 100\%)$ e 36, 7% $((8, 2 - 6, 0)/6, 0 \times 100\%)$; ver Tabela 3.5, respectivamente, confirmando que esses testes são liberais. Em contrapartida, as estatísticas razão de verossimilhanças perfiladas modificadas (LR_m) e escore (S) apresentam discrepâncias relativas em torno de $-5, 0\%$ $((5, 7 - 6, 0)/6, 0 \times 100\%)$, mais próximas de zero que as estatísticas razão de verossimilhanças e gradiente, no entanto reafirmam um comportamento conservativo.

Concluimos destes resultados que os testes baseados nas estatísticas da razão de verossimilhanças usual (LR) e gradiente (G) são muitos liberais em amostras pequenas e moderadas, rejeitando $\mathcal{H}_0$ com frequência muito maior que o esperado. Por outro lado, os testes razão de verossimilhanças perfiladas modificadas (LR_m) e escore (S) apresentam taxas de rejeição empíricas mais próximas dos níveis nominais, mas estes testes são conservativos. Portanto, dos testes analisados neste capítulo o teste razão de verossimilhança bootstrap (LR_{Boot}) é o mais recomendável.

Tabela 2.2: Tamanhos empíricos dos testes LR, LR_m, LR_{Boot}, S e G - Modelo normal não linear para $n = 35$, $q = 2$ e diferentes valores de p .

p	$\alpha(\%)$	LR	LR _m	LR _{Boot}	S	G
2	10,0	13,8	9,6	9,7	8,8	13,1
	5,0	7,7	4,5	4,9	4,3	7,0
	1,0	1,9	0,8	1,0	0,8	1,6
	0,5	1,0	0,4	0,6	0,4	0,8
3	10,0	14,9	8,8	9,8	8,6	14,1
	5,0	8,2	4,2	5,0	4,3	7,6
	1,0	2,3	0,7	1,0	0,8	1,9
	0,5	1,2	0,3	0,6	0,4	0,9
4	10,0	16,4	8,8	9,9	8,3	15,7
	5,0	9,9	4,2	5,3	3,8	9,0
	1,0	2,7	0,8	1,2	0,7	2,3
	0,5	1,7	0,4	0,6	0,3	1,3
5	10,0	20,3	9,0	10,4	8,9	19,2
	5,0	12,7	4,4	5,3	4,2	11,6
	1,0	4,2	0,7	1,2	0,7	3,6
	0,5	2,7	0,3	0,7	0,4	2,0
6	10,0	22,5	8,5	10,3	8,2	21,4
	5,0	14,4	4,1	5,4	4,0	13,2
	1,0	5,2	0,8	1,1	0,8	4,6
	0,5	3,4	0,4	0,6	0,4	2,9
7	10,0	27,5	8,9	10,7	8,4	25,3
	5,0	18,6	4,5	5,5	3,9	16,3
	1,0	7,4	0,8	1,5	0,8	5,8
	0,5	5,4	0,4	0,9	0,4	3,8
8	10,0	34,1	8,1	10,5	7,8	31,9
	5,0	24,5	3,8	5,3	3,7	21,9
	1,0	11,5	0,6	1,2	0,6	9,3
	0,5	8,1	0,3	0,7	0,3	6,5

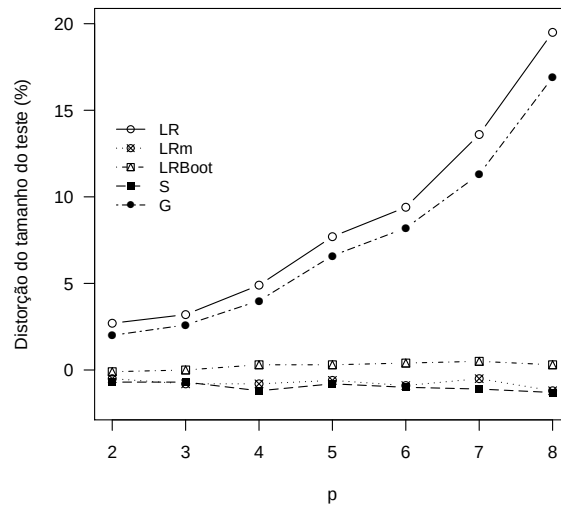
Tabela 2.3: Tamanhos empíricos dos testes LR, LR_m, LR_{Boot}, S e G - Modelo normal não linear para $p = 5$, $q = 2$ e diferentes tamanhos amostrais.

n	$\alpha(\%)$	LR	LR _m	LR _{Boot}	S	G
25	10,0	28,6	8,6	10,1	7,9	25,9
	5,0	19,2	4,0	5,0	4,0	16,8
	1,0	7,9	0,7	1,0	0,8	5,8
	0,5	5,3	0,3	0,6	0,4	3,6
35	10,0	20,3	9,0	10,4	8,9	19,2
	5,0	12,7	4,4	5,3	4,2	11,6
	1,0	4,2	0,7	1,2	0,7	3,6
	0,5	2,7	0,3	0,7	0,4	2,0
45	10,0	16,5	8,8	9,6	8,0	16,0
	5,0	9,8	4,4	4,8	4,1	9,3
	1,0	2,9	0,8	1,1	0,7	2,6
	0,5	1,7	0,4	0,7	0,3	1,4
100	10,0	14,3	9,0	10,2	9,5	13,7
	5,0	8,2	4,5	5,3	4,5	7,8
	1,0	2,2	0,9	1,2	0,9	1,9
	0,5	1,3	0,5	0,7	0,4	1,1

Tabela 2.4: Poder dos testes LR, LR_m, LR_{Boot}, S e G - Modelo normal não linear para $n = 35$, $p = 5$, $q = 2$ e $\alpha = 10,0\%$.

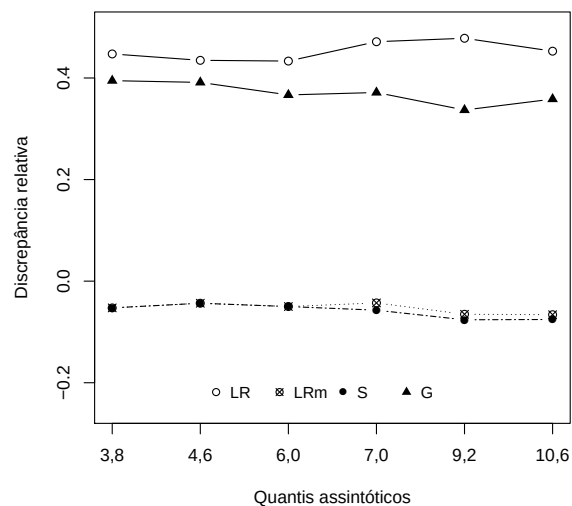
δ	LR	LR _m	LR _{Boot}	S	G
0,0	20,3	9,0	10,4	8,9	19,2
0,5	28,3	13,9	13,5	13,8	26,4
1,0	41,2	27,1	26,2	27,3	40,0
1,5	64,6	51,4	48,4	47,6	63,3
2,0	87,5	77,0	71,2	76,1	87,7
2,5	96,2	91,3	87,6	90,3	96,3
3,0	97,4	94,7	95,0	90,6	97,4
3,5	99,7	98,3	98,4	96,0	99,7

Figura 2.1: Distorção do tamanho dos testes LR, LR_m, LR_{Boot}, S, G - Modelo normal não linear com $n = 35$, $q = 2$ e diferentes valores de p .



Fonte: Elaborada pela autora.

Figura 2.2: Discrepâncias relativas de quantis das estatísticas LR, LR_m, S, G - Modelo normal não linear com $n = 35$, $p = 5$ e $q = 2$.



Fonte: Elaborada pela autora.

2.6 Comentários

Apresentamos primeiramente uma breve introdução sobre os modelos não lineares da família exponencial com dispersão variável. Discutimos alguns aspectos inferências, como testes de heteroscedasticidade usuais tais como razão de verossimilhanças e escore. Também apresentamos testes baseados na estatística da razão de verossimilhanças perfiladas modificada, na recente estatística gradiente e na estatística razão de verossimilhanças via reamostragem bootstrap. Em várias situações práticas observamos que a dispersão não é a mesma para todas as observações, por isso é determinante elaborar testes mais confiáveis e precisos que detectam a presença de heteroscedasticidade nos dados.

Nos resultados numéricos vimos que os testes LR e G apresentam um comportamento liberal, com taxas de rejeição distorcidas acima dos níveis nominais considerados. Vimos que a modificação na verossimilhança perfilada atenuou o efeito do número de parâmetros de perturbação nas inferências, porém este teste produziu um comportamento conservativo, similarmente ao teste S. Notamos a necessidade de técnicas estatísticas que visem corrigir e melhorar a aproximação das distribuições dessas estatísticas para a distribuição χ^2 . Em linhas gerais, dos testes apresentados o mais recomendável é o teste LR_{Boot} . No entanto, este teste requer um tempo de execução (computacionalmente) mais alto que os demais.

Capítulo 3

Aperfeiçoamento de testes em MNLFE

3.1 Introdução

Vimos no Capítulo 2 através de simulação de Monte Carlo que os testes baseados em aproximações assintóticas de primeira ordem não são bons quando o tamanho amostral é pequeno ou mesmo de tamanho moderado, conduzindo a taxas de rejeição distorcidas. Vimos que os testes baseados nas estatísticas da razão de verossimilhanças (LR) e gradiente (G) são bastantes liberais, enquanto que os testes baseados nas estatísticas da razão de verossimilhanças perfiladas modificada (LR_m) e escore (S) são conservativos. Para isso, foram desenvolvidos técnicas que visam corrigir e melhorar a aproximação assintótica das estatísticas de testes. Na literatura são encontradas as correções de Bartlett às estatísticas da razão de verossimilhanças perfiladas modificadas (Cysneiros e Ferrari, 2006), razão de verossimilhanças bootstrap (Rocke, 1989) e correção tipo-Bartlett para a estatística escore (Cavalcanti, 2009), que não será apresentado nessa dissertação, para os MNLFE. Nosso objetivo agora é obter um fator de correção tipo-Bartlett para a estatística gradiente (Terrel, 2002), na classe dos MNLFE, de tal forma que a aproximação por χ_q^2 seja até de ordem n^{-1} . A organização desse capítulo será dada na seguinte forma: definimos na Seção 3.2 cumulantes conjuntos e apresentamos as principais identidades de Bartlett que facilitam o cálculo dos cumulantes. Na Seção 3.3 apresentamos a forma matricial para o fator de correção Bartlett na estatística razão de verossimilhanças. Apresenta-se na Seção 3.4 a correção de Bartlett aplicada a estatística razão de verossimilhanças perfiladas modificadas. Na Seção 3.5 apresentamos a metodologia do cálculo da estatística LR_{Boot} em sua versão corrigida. Como objetivo principal deste capítulo apresentamos na Seção 3.6 a expressão matricial para o fator de correção tipo-Bartlett para a estatística gradiente nos modelos não lineares da família exponencial, sendo baseada no método geral obtido por Vargas, Ferrari e Lemonte (2013). Desenvolvemos na Seção 3.7 estudos de simulação Monte Carlo para avaliar e comparar numericamente os desempenhos dos testes supracitados no Capítulo 2 e neste capítulo, em amostras finitas. Finalmente, realizamos uma aplicação empírica.

3.2 Cumulantes conjuntos e identidades de Bartlett

Primeiramente, apresentamos algumas notações. Seja $\psi = (\delta^\top, \beta^\top, \gamma)^\top$ o vetor de parâmetros que pertence a algum espaço paramétrico, em que ψ^r é o r -ésimo elemento de ψ . Denotamos $U_j = \partial \mathcal{L}^* / \partial \psi^j$, $U_{jr} = \partial^2 \mathcal{L}^* / \partial \psi^j \partial \psi^r$, $U_{jrs} = \partial^3 \mathcal{L}^* / \partial \psi^j \partial \psi^r \partial \psi^s$, etc. Definimos os cumulantes como:

$$\kappa_{jr} = \mathbb{E}(U_{jr}), \quad \kappa_{jrs} = \mathbb{E}(U_{jrs}), \quad \kappa_{j,r} = \mathbb{E}(U_j U_r), \quad \kappa_{j,rs} = \mathbb{E}(U_j U_{rs})$$

e assim por diante. Denotamos as derivadas dos cumulantes por $\kappa_{jr}^{(s)} = \partial \kappa_{jr} / \partial \psi^s$, $\kappa_{jr}^{(st)} = \partial^2 \kappa_{jr} / \partial \psi^s \partial \psi^t$. Adicionalmente, os elementos da matriz informação de Fisher dada em (2.9) possui elementos $-\kappa_{jr}$, sendo $-\kappa^{jr}$ os correspondentes elementos de sua inversa. Para evitar ambiguidades de índices entre os componentes do vetor de parâmetros de interesse e os componentes do vetor de perturbação vamos adotar as seguintes notações: os índices $a, b, c, \dots$ variam em $1, \dots, q$ relacionados aos parâmetros de interesse, os índices $i, j, k, \dots$ variam em $p+1, \dots, p+q$ relacionados aos parâmetros de perturbação β 's e os índices $r, s, t, \dots$ variam em $1, \dots, p+q+1$. Algumas relações podem ser estabelecidas a partir dos cumulantes apresentados acima, essas relações são denominadas identidades de Bartlett. Alguns exemplos dessas relações são:

$$\kappa_{r,s} = -\kappa_{rs},$$

$$\kappa_{rst,u} = \kappa_{rst}^{(u)},$$

$$\kappa_{rst}^{(u)} = \kappa_{rstu} + \kappa_{u,rst},$$

$$\kappa_{r,s,t} = -\kappa_{rst} - \sum_{(3)} \kappa_{r,s},$$

$$\kappa_{r,s,tu} = \kappa_{rstu} - \kappa_{rtu}^{(s)} - \kappa_{stu}^{(r)} + \kappa_{tu}^{(rs)} - \kappa_{rs,tu} \text{ e}$$

$$\kappa_{rstu} + \sum_{(4)} \kappa_{r,stu} + \sum_{(3)} \kappa_{rs,tu} + \sum_{(6)} \kappa_{r,s,tu} + \kappa_{r,s,t,u} = 0, \text{ etc,}$$

em que $\sum_{(k)}$ denota a soma sobre as k permutações dos índices.

As identidades de Bartlett facilitam a obtenção dos cumulantes e podem conduzir a cálculos diretos de alguns cumulantes em algumas situações. Os cumulantes são utilizados principalmente em correções de Bartlett e tipo-Bartlett que serão discutidas posteriormente neste capítulo. Devido a ortogonalidade dos parâmetros nos MNLFE, o cálculo de certos cumulantes tornam-se muito simples, uma vez que vários cumulantes mistos são nulos. No Apêndice A apresentamos alguns cumulantes conjuntos para os modelos não lineares da família exponencial reparametrizados com dispersão variável.

3.3 Estatística da razão de verossimilhanças corrigida

Para melhorar a aproximação da distribuição da estatística LR, (BARTLETT, 1937) propôs um fator de correção $(1 + c/q)^{-1}$ que multiplicado a estatística razão de verossimilhanças gera a estatística definida como razão de verossimilhanças corrigida (LR^*), de tal forma que, a sua média está mais próxima do valor esperado da distribuição de referência χ_q^2 . A estatística LR^* é dada por

$$LR^* = \frac{LR}{1 + \frac{c}{q}},$$

em que q é o número de parâmetros de interesse, ou seja, é o número de restrições impostas pela hipótese nula, n é o tamanho da amostra e a constante c é de ordem n^{-1} e é função dos cumulantes e derivadas dos cumlantes. Sob $\mathcal{H}_0$ o valor esperado da LR, corresponde a $q\{(1 + c/q) + O(n^{-3/2})\}$. Enquanto $P(LR > x_\alpha) = \alpha + O(n^{-1})$, tem-se que $P(LR^* > x_\alpha) = \alpha + O(n^{-2})$, em que x_α é o quantil $(1 - \alpha)$ da distribuição χ_q^2 . Baseando-se no método geral de obtenção da constante c desenvolvida por Lawley (1956), Cordeiro (1993) obteve um fator de correção de Bartlett para a estatística LR para o teste $\mathcal{H}_0$ contra $\mathcal{H}_1$ no modelo heteroscedástico multiplicativo, ou seja, $m_\ell = \exp\{z_\ell^\top \delta\}$. Para $\delta = 0$, obtem-se um modelo homoscedástico. A constante c é dada por

$$c = h(X, R) - \frac{3(k+1)^2 - 1}{6n},$$

sendo

$$\begin{aligned} h(X, R) &= -\frac{1}{2}\text{tr}(A_d^2) + \frac{1}{3}\mathbf{1}^\top A^{(3)}\mathbf{1} + 12\mathbf{1}^\top A_d A A_d \mathbf{1} + \text{tr}(A_d B_d) - \mathbf{1}^\top (B^{(2)} \odot A)\mathbf{1} \\ &+ \frac{1}{2}\mathbf{1}^\top B_d A B_d \mathbf{1} + \mathbf{1} A_d A B_d \mathbf{1}, \end{aligned}$$

em que $A = R(R^\top R)^{-1}R^\top$ com $R = [\mathbf{1}Z]$, $A_d = \text{diag}(a_{11}, \dots, a_{nn})$, $B = X(X'X)^{-1}X^\top$ e $B_d = \text{diag}(b_{11}, \dots, b_{nn})$. Em problemas regulares, a estatística LR^* tem distribuição assintótica χ_q^2 com erro de ordem $O(n^{-2})$.

3.4 Estatística da razão de verossimilhanças perfiladas modificadas corrigida

No Capítulo 2 vimos que o ajuste de Cox e Reid (1987) para a verossimilhança perfilada reduziu o impacto de parâmetros de perturbação no teste baseado na estatística LR conduzindo a taxas de rejeição empíricas mais próximas dos níveis nominais adotados, todavia, o teste LR_m produziu taxas de rejeição abaixo dos níveis nominais. Contudo, a partir dos resultados de DiCiccio e Stern (1994), Cysneiros e Ferrari (2006) desenvolveram um fator de correção de Bartlett para a estatística LR_m definida na Subseção 2.4.4 para a classe dos MNLFE, tal como

a aproximação por χ_q^2 para a estatística corrigida é de ordem n^{-2} . A estatística da razão de verossimilhanças perfiladas modificadas corrigida é dada por

$$\text{LR}_m^* = \frac{\text{LR}_m}{1 + c_m/q},$$

em que, admitindo uma estrutura multiplicativa para a dispersão nos modelos não lineares normal, o fator de correção de Bartlett c_m fica dada por

$$c_m = -\frac{1}{2}\text{tr}(H_d H_d) + \frac{1}{2n}q^2 + \frac{1}{2}\mathbf{1}^\top H_d H H_d \mathbf{1} + \frac{1}{3}\mathbf{1}^\top H^{(3)} \mathbf{1} - \frac{2}{n}q + \frac{1}{n}\mathbf{1}^\top H^{(2)} \mathbf{1},$$

sendo $H = \{h_{\ell s}\} = (Z - \bar{Z})[(Z - \bar{Z})^\top (Z - \bar{Z})]^{-1}(Z - \bar{Z})^\top$, com $(Z - \bar{Z}) = (z_1 - \bar{z}, \dots, z_n - \bar{z})^\top$, $\bar{z} = (\bar{z}_1, \dots, \bar{z}_q)^\top$, $\bar{z}_a = n^{-1} \sum_{\ell=1}^n z_{\ell a}$ para $a = 1, \dots, q$, $H_d = \text{diag}\{h_{11}, \dots, h_{nn}\}$, $H^{(2)} = (h_{\ell s}^2)$ e $H^{(3)} = (h_{\ell s}^3)$.

Observe que o cálculo do fator de correção c_m não depende do número de parâmetros de perturbação, nem de parâmetros desconhecidos. Depende somente do tamanho amostral e da matriz H , cujo é função da matriz Z de variáveis auxiliares usadas para modelar as flutuações na dispersão da resposta. Cysneiros e Ferrari (2006) mostraram por meio de simulação de Monte Carlo que o teste baseado na estatística LR_m^* produz inferências mais precisas que os testes baseados nas estatísticas LR, LR_m e LR^* . Note que a expressão matricial do fator de correção c_m é muito simples visto que requer apenas operações com vetores e matrizes, sendo facilmente computada. No entanto, o fator de correção de Bartlett é aplicado á forma exata da estatística LR_m , que envolve maximização da função de verossimilhança perfilada modificada.

3.5 Estatística da razão de verossimilhanças bootstrap corrigida

Rocke (1989) introduz uma técnica alternativa para calcular o fator de correção de Bartlett utilizando a técnica de reamostragem bootstrap (EFRON, 1979), para melhorar a qualidade do teste baseado na estatística LR e não para se obter um valor crítico da distribuição nula, em outras palavras, a proposta de Rocke (1989) é estimar numericamente o fator de correção de Bartlett para a estatística LR por meio da amostra observada $y = y_1, \dots, y_\ell$. O procedimento para calcular o teste razão de verossimilhanças bootstrap corrigida $\text{LR}_{\text{Boot}}^*$ pode ser descrito como:

- Gerar B pseudos-amostras aleatórias $(y^{*1}, \dots, y^{*B})$ usando o bootstrap paramétrico impondo a hipótese nula.
- Calcular o valor da estatística razão de verossimilhanças para cada pseudo-amostra y^{*b} , com $b = 1, \dots, B$:

$$\text{LR}^{*b} = 2\{\mathcal{L}^*(\hat{\psi}^{*b}; y^{*b}) - \mathcal{L}^*(\tilde{\psi}^{*b}; y^{*b})\},$$

em que $\hat{\psi}^{*b}$ e $\tilde{\psi}^{*b}$ são EMV irrestrito e restrito de ψ , respectivamente da réplica b .

- calcular o valor da estatística LR_{Boot}^* da seguinte maneira:

$$LR_{\text{Boot}}^* = \frac{LR}{\overline{LR}^{*b}} q,$$

em que LR é a estatística de razão de verossimilhanças da amostra original, q representa o número de restrições impostas na hipótese nula e

$$\overline{LR}^{*b} = \frac{1}{B} \sum_{b=1}^B LR^{*b},$$

Sob $\mathcal{H}_0$ a estatística LR_{Boot}^* tem distribuição assintótica χ_q^2 .

3.6 Estatística gradiente corrigida

A partir de uma expansão assintótica da distribuição acumulada, sob hipótese nula, da estatística gradiente e utilizando uma rota bayesiana com base no argumento de encolhimento abordado por Ghosh e Mukerjee (1991), Vargas, Ferrari e Lemonte (2013) propuseram a estatística gradiente corrigida (G^*) utilizando um fator de correção tipo-Bartlett. A estatística gradiente corrigida, sob a hipótese nula, tem distribuição qui-quadrado até um erro de ordem $o(n^{-1})$, enquanto que a estatística gradiente proposta por Terrell (2002) tem erro até de ordem $o(n^{-1/2})$, sob $\mathcal{H}_0$. A partir disso, Vargas, Ferrari e Lemonte (2013) obtiveram expressões matriciais e algébricas para G^* nos MLG com dispersão conhecida e desconhecida. A estatística gradiente corrigida é expressa pela seguinte forma:

$$G^* = G[1 - (c + bG + aG^2)], \quad (3.1)$$

em que

$$a = \frac{A_3}{12q(q+2)(q+4)}, \quad b = \frac{A_2 - 2A_3}{12q(q+2)} \quad \text{e} \quad c = \frac{A_1 - A_2 + A_3}{12q}. \quad (3.2)$$

As expressões algébrica das A_1 , A_2 e A_3 que são funções dos cumulantes e de derivadas dos

cumulantes são dadas por

$$\begin{aligned}
A_1 &= 3 \sum' \kappa_{jrs} \kappa_{klu} [m^{jr} a^{lu} (m^{sk} + 2a^{sk}) + a^{jr} m^{sk} a^{lu} + 2m^{jk} a^{rl} a^{su}] \\
&\quad - \sum' \kappa_{jr}^{(s)} \kappa_{kl}^{(u)} (\kappa^{js} \kappa^{rk} \kappa^{lu} + a^{sj} a^{rk} a^{lu} + \kappa^{sk} \kappa^{lj} \kappa^{ru} + a^{sk} a^{lj} a^{ru}) \\
&\quad - 6 \sum' \kappa_{jrs} \kappa_{kl}^{(u)} [(a^{su} - \kappa^{su}) (\kappa^{jk} \kappa^{lr} - a^{jk} a^{lr}) + m^{jr} (a^{sk} a^{lu} + \kappa^{sk} \kappa^{lu}) \\
&\quad + 2a^{rs} (\kappa^{jk} \kappa^{lu} - a^{jk} a^{lu}) + 2a^{rk} a^{ls} m^{ju}] \\
&\quad + 6 \sum' \kappa_{jrsu} m^{jr} a^{su} - 6 \sum' \kappa_{jrs}^{(u)} [m^{jr} (a^{su} - \kappa^{su}) + 2m^{ju} a^{rs}] \\
&\quad + 12 \sum' \kappa_{jr}^{(su)} (\kappa^{jr} \kappa^{su} - a^{jr} a^{su}), \\
A_2 &= -3 \sum' \kappa_{jrs} \kappa_{klu} [m^{jr} a^{lu} m^{sk} + m^{jr} a^{sk} m^{lu} + 2m^{jk} a^{su} m^{rl} \\
&\quad + \frac{1}{4} (3m^{jr} m^{sk} m^{lu} + 2m^{jk} m^{rl} m^{su})] \\
&\quad + 6 \sum' \kappa_{jrs} \kappa_{kl}^{(u)} [m^{su} (\kappa^{jk} \kappa^{lr} - a^{jk} a^{lr}) + m^{jr} (\kappa^{sk} \kappa^{lu} - a^{sk} a^{lu})] \\
&\quad + 6 \sum' \kappa_{jrs}^{(u)} m^{jr} m^{su} - 3 \sum' \kappa_{jrsu} m^{jr} m^{su}, \\
&\quad \text{e} \\
A_3 &= -3 \sum' \kappa_{jrs} \kappa_{klu} (3m^{jr} m^{sk} m^{lu} + 2m^{jk} m^{rl} m^{su}),
\end{aligned}$$

em que os índices j, r, s, k, l , e u variam sobre as componentes dos vetores δ, β e γ e $\sum'$ denota todas as possíveis combinações de $(p + q + 1)$ parâmetros de $\beta_1, \dots, \beta_p, \delta_1, \dots, \delta_q$ e γ . Sendo a^{jr} e m^{jr} os elementos (j, r) das matrizes

$$A = \begin{pmatrix} 0 & 0 \\ 0 & K^{22} \end{pmatrix} \quad \text{e} \quad M = K^{-1} - A,$$

com $K^{-1} = -((\kappa^{jr}))$ denotando a inversa da matriz de informação de Fisher.

Agora vamos apresentar as formas matriciais para o fator de correção tipo-Bartlett à estatística gradiente em MNLFE considerando o caso de heteroscedasticidade multiplicativa. Daí, definimos as matrizes $Z_\beta = X^* (X^{*\top} W \Phi X^*)^{-1} X^{*\top}$, $Z_\delta = 2(Z - \bar{Z})[(Z - \bar{Z})^\top (Z - \bar{Z})]^{-1} (Z - \bar{Z})^\top$. Como no Capítulo 2, $\Phi = (1/\gamma)Q$, sendo $Q = \text{diag}\{q_1, \dots, q_n\}$, $X^* = \partial \eta / \partial \beta^\top$ e $W = I_n$ para o caso do modelo normal. Denotamos $Z^{(2)} = Z \odot Z$ e $Z^{(3)} = Z \odot Z \odot Z$, sendo $\odot$ o produto direto de matrizes. Utilizamos o sub-índice d para indicar que uma matriz diagonal foi obtida a partir da diagonal principal da matriz original e $\mathbf{1}$ representa um vetor n -dimensional de uns. Portanto, as formas matriciais dos A 's para testar $\mathcal{H}_0 : \delta = \delta_0$ contra a hipótese alternativa $\mathcal{H}_1 : \delta \neq \delta_0$ são dadas da seguinte forma:

$$\begin{aligned}
A_1 &= \frac{3}{2} \mathbf{1}^\top Z_{\delta_d} Z_\delta Z_{\beta_d} W \Phi \mathbf{1} + \frac{2}{n} \mathbf{1}^\top Z_{\delta_d} Z_{\beta_d} W \Phi \mathbf{1} \\
&+ 3 \mathbf{1}^\top \Phi W Z_{\beta_d} Z_\delta Z_{\beta_d} W \Phi \mathbf{1} + \frac{12}{n} \mathbf{1}^\top \Phi W [Z_\beta^{(2)}] W \Phi \mathbf{1} \\
&+ 6 \mathbf{1}^\top \beta_d Z_{\delta_d} W \Phi \mathbf{1}, \\
\\
A_2 &= -\frac{3}{2} \mathbf{1}^\top Z_{\delta_d} Z_\delta Z_{\beta_d} W \Phi \mathbf{1} - \frac{6}{4n} \mathbf{1}^\top [Z_{\delta_d}^2] \mathbf{1} - \frac{3}{n} \mathbf{1}^\top [Z_\delta^{(2)}] \mathbf{1} \\
&- \frac{9}{16} \mathbf{1}^\top Z_{\delta_d} Z_\delta Z_{\delta_d} \mathbf{1} - \frac{3}{8} \mathbf{1}^\top [Z_\delta^{(3)}] \mathbf{1} + \frac{3}{2} \mathbf{1}^\top [Z_{\delta_d}^2] \mathbf{1} \\
\\
&\text{e} \\
A_3 &= \frac{3}{4} \mathbf{1}^\top Z_{\delta_d} Z_\delta Z_{\delta_d} \mathbf{1} + \frac{1}{8} [Z_\delta^{(3)}] \mathbf{1}.
\end{aligned} \tag{3.3}$$

Mais detalhes sobre a obtenção das expressões que compõem os A 's encontram-se no Apêndice B. Observe que os cálculos dessas quantidades envolvem operações simples com vetores e matrizes sendo facilmente computadas. Substituindo A_1 , A_2 e A_3 em (3.2) e aplicando estes resultados na expressão (3.1) obtemos a estatística gradiente corrigida. Sob $\mathcal{H}_0$, a estatística G^* possui distribuição assintótica χ_q^2 .

3.7 Avaliação numérica

O objetivo desta seção é comparar o desempenho dos testes LR, LR_m , LR_{Boot} , S e G apresentados no Capítulo 2 com os testes, LR_m^* , LR_{Boot}^* e G^* apresentados neste capítulo através de simulação de Monte Carlo. Aproveitamos alguns resultados obtidos na Seção 2.5 e comparamos com novos resultados dos testes baseados nas estatísticas LR_m^* , LR_{Boot}^* e G^* proposta nesta dissertação.

Avaliamos estes desempenhos através da proximidade das probabilidades de rejeição da hipótese nula, sendo esta verdadeira (erro tipo I) com os respectivos níveis nominais 10, 0%, 5, 0%, 1, 0% e 0, 5% e o poder dos testes levando em consideração a hipótese alternativa para diferentes valores de δ . Outra vez, aqui vamos testar:

$$\begin{aligned}
\mathcal{H}_0: \delta_1 = \delta_2 = \dots = \delta_q = 0 \text{ (modelo homoscedástico)} \\
\text{versus} \\
\mathcal{H}_1: \delta_1 \neq \delta_2 \neq \dots \neq \delta_q \neq 0 \text{ (modelo heteroscedástico)}.
\end{aligned}$$

Adotamos o mesmo modelo de regressão não linear (2.5) apresentado na Seção 2.5 e os respectivos valores para geração das covariadas e variável resposta, conforme descrito também na

Seção 2.5 para fim de simulação. Usamos diferentes valores de p e q a fim de avaliar o desempenho dos testes quando aumentamos o número de parâmetros de interesse e perturbação. A matriz Z utilizada para modelar as flutuações da variância foi construída utilizando as colunas $2, 3, \dots, p$ da matriz X de covariadas quando $q < p$, para o caso em que $q \geq p$, as colunas extras foram geradas usando amostras aleatórias independentes da distribuição $\mathcal{U}(0, 1)$. Aqui, também o número de réplicas Monte Carlo e bootstrap foram fixadas em 10000 e 600, respectivamente. Para cada tamanho amostral e nível nominal considerado, estimamos, via simulação, as taxas de rejeição, $P(LR \geq \chi^2_{(\alpha, q)})$, $P(LR_m \geq \chi^2_{(\alpha, q)})$, $P(LR_m^* \geq \chi^2_{(\alpha, q)})$, $P(LR_{Boot} \geq \hat{q}_{(1-\alpha)})$, $P(LR_{Boot}^* \geq \chi^2_{(\alpha, q)})$, $P(S \geq \chi^2_{(\alpha, q)})$, $P(G \geq \chi^2_{(\alpha, q)})$ e $P(G^* \geq \chi^2_{(\alpha, q)})$, tal que $\chi^2_{(\alpha, q)}$ é o percentil $(1 - \alpha)$ da distribuição qui-quadrado e $\hat{q}_{(1-\alpha)}$ é estimado de acordo com o procedimento descrito na Seção 2.4.5.

A Tabela 3.1 apresenta os resultados da simulação dos tamanhos dos testes para $n = 35$, número de parâmetros de interesse ($q = 2$) e diferentes valores para p . A finalidade da variação dos valores de p foi para analisar o efeito nos tamanhos dos testes quando o número de parâmetros de perturbação aumenta. Observamos que os testes baseados na estatística gradiente e razão de verossimilhanças apresentam taxas de rejeição acima dos níveis nominais considerados e essas taxas diferenciam mais dos níveis nominais quando aumentamos o número de parâmetros de perturbação. Por exemplo, para $p = 5$ e $\alpha = 5, 0\%$ as taxas de rejeição dos testes LR e G foram 12,7% e 11,6%, respectivamente. No entanto, o fator de correção tipo-Bartlett para a estatística gradiente atenuou significativamente a tendência do teste gradiente em rejeitar com frequência a hipótese nula, exibindo, por exemplo para o mesmo caso, taxa de rejeição igual a 5,3%, satisfatoriamente próximo ao nível nominal. Já o teste S apresenta taxas de rejeição conservativas, ou seja, abaixo dos níveis nominais continuando conservativa à medida que aumentamos p . O teste LR_m tende a corrigir a tendência liberal do teste LR, mas exibe taxas de rejeição abaixo dos níveis nominais para todos valores de p . Observe que a aplicação do fator de correção Bartlett no teste baseado na estatística LR_m^* eleva as taxas para valores mais próximos dos tamanhos nominais, apresentando taxas de rejeição mais estáveis, assim como os testes LR_{Boot} e LR_{Boot}^* .

A Tabela 3.2 apresenta resultados da simulação para $n = 35$, $p = 5$ (fixos) e quando variamos o número de parâmetros de interesse, isto é, $q = 1, \dots, 8$. Observamos que novamente, os testes LR e G são muito liberais, apresentando taxas de rejeição demasiadamente acima dos níveis nominais. Por exemplo, para $q = 5$ e $\alpha = 5, 0\%$, estes testes apresentam taxas de rejeição 24,0% e 20,6%, respectivamente, isto é, quase o quádruplo e o quádruplo do nível nominal selecionado, respectivamente. O fator de correção tipo-Bartlett aplicado à estatística G atenua a liberalidade do teste original. Notamos que, quando aumentamos o número de parâmetros que definem o comportamento de heteroscedasticidade, os tamanhos dos testes LR, G e G^* tendem a se distanciar consideravelmente em todos os níveis nominais adotados. Apesar que o teste G^* permanece liberal, a correção fez com que a taxa de rejeição reduzisse de 20,6% para 8,2%, ou seja, uma redução de 12,4% no mesmo caso citado, em que $p = 5$ e $\alpha = 5, 0\%$. Isso representa uma diminuição significativa mesmo para q elevado. Para a mesma situação a taxa

de rejeição dos testes LR_m , LR_m^* , LR_{Boot} e LR_{Boot}^* são, respectivamente, 4,0%, 4,7%, 5,1% e 5,2%, permanecendo próximas aos tamanhos nominais ao passo que aumentamos o número de parâmetros de interesse. O teste escore permanece conservativo como na Tabela 3.1 quando aumentamos o número de parâmetros de perturbação.

Na Tabela 3.3 variamos o tamanho amostral em $n = 25, 35, 45, 100$ e fixamos em $p = 5$ e $q = 2$. Com o aumento da amostra, as taxas de rejeição dos testes se aproximam do nível nominal, como esperado. Mesmo para $n = 100$ os testes LR e G apresentam taxas de rejeição acima dos níveis considerados. Obviamente $n = 35$ equivale ao caso da Tabela 3.1 quando $p = 5$ e o caso da Tabela 3.2 quando $q = 2$. Com o tamanho amostral igual a 100 e $\alpha = 5,0\%$ temos que as taxas de rejeição dos testes LR, LR_m , LR_{Boot} , G foram 8,2%, 4,5%, 5,3% e 7,8%, respectivamente, e as versões corrigidas das três últimas estatísticas, 4,9%, 5,3% e 4,8%. Já para o teste S a taxa de rejeição foi igual a 4,5%. Por outro lado, todos as três testes corrigidos LR_m^* , LR_{Boot}^* , G^* apresentam taxas de rejeição bem próximas a todos os níveis nominais.

A Tabela 3.4 apresenta as taxas de rejeição adotando como verdadeira a hipótese alternativa, foram fixados $p = 5$, $q = 2$ e tamanho da amostra $n = 35$. Consideramos os níveis nominais $\alpha = 10,0\%$, $5,0\%$, $1,0\%$, $0,5\%$ para diferentes valores de δ , obviamente que para $\delta = 0$ obtemos o tamanho do teste. Não consideramos os testes baseados nas estatísticas razão de verossimilhanças original e gradiente por apresentarem comportamento bastante liberal conforme discutimos no Capítulo 2. Como esperado, à medida que aumentamos o valor de δ , ou seja, afastamos do valor da hipótese nula ($\delta_1 = \delta_2 = 0$) o poder do teste cresce para todos os níveis nominais. Por exemplo, fixando $\alpha = 10,0\%$ temos que para $\delta = 2,0$ os poderes dos testes LR_m , LR_m^* , LR_{Boot} , LR_{Boot}^* , S e G^* são, 77,0%, 78,4%, 71,2%, 71,2%, 76,1% e 79,0%, respectivamente. Enquanto que, para $\delta = 3,5$ os poderes dos testes são, 98,3%, 99,7%, 98,4%, 98,4%, 96,0% e 98,8%, respectivamente. Vemos que os poderes dos testes não apresentam diferenças significativas, porém o teste S apresenta uma pequena desvantagem com relação aos demais testes considerados neste caso.

Na Tabela 3.5 comparamos os quantis amostrais das estatísticas LR, LR_m , LR_m^* , LR_{Boot}^* , S, G e G^* com os da distribuição χ_2^2 . Consideramos $p = 5$ e variamos o tamanho da amostra ($n = 25, 35, 45, 100$). Observe que os quantis amostrais das estatísticas LR e G estão bem acima dos quantis da χ_2^2 , mesmo para amostra de tamanho moderado. Por exemplo, para o quantil de ordem 95% e $n = 45$ temos que as estatísticas LR e G, apresentam quantis 7,8 e 7,6, sendo o quantil 95% da distribuição χ_2^2 igual a 6,0. Contudo, os quantis das estatísticas LR_m^* , LR_{Boot}^* , G^* são mais próximos dos quantis da distribuição χ_2^2 . Para a mesma situação essas estatísticas apresentam quantis amostrais 6,0, 5,9 e 6,0, respectivamente. Já os quantis amostrais de LR_m e S estão abaixo dos quantis da distribuição χ_2^2 , confirmando seus comportamentos conservativos. Note que, como esperado à medida que aumentamos a amostra os quantis amostrais para todas as estatísticas aproximam-se dos quantis da distribuição χ_2^2 . Mesmo assim, para $n = 100$ os quantis amostrais de LR, LR^* e G excedem os quantis da χ_2^2 .

Na Tabela 3.6 apresentamos as médias e as variâncias estimadas de todas estatísticas consideradas neste estudo. Consideramos $p = 5$, $n = 35$ e diferentes valores de q . Observe que as médias e as variâncias de LR e G se distanciam demasiadamente das médias da distribui-

ção χ_q^2 quando q aumenta. Todavia a correção na estatística gradiente melhora a qualidade da aproximação da gradiente exibindo médias e variâncias mais próximas com os da distribuição χ_q^2 . Assim, mesmo para q grande podemos observar uma significativa melhora. Conjuntamente, as médias e variâncias de LR_m^* e LR_{Boot}^* são próximas com os da distribuição χ_q^2 . Dentre as três estatísticas que apresentam médias e variâncias mais próximas com as da distribuição χ_q^2 , a estatística LR_m^* é que obteve menos efeito quando aumentamos q . Mais uma vez temos indícios de que as estatísticas LR_m e S são conservativas apresentando médias e variâncias abaixo das médias e variâncias da χ_q^2 . Seguindo ainda esta tabela, confirmamos que a estatística LR_{Boot} não possui distribuição assintótica χ^2 , ou seja, possui distribuição nula desconhecida, assim estimamos o valor crítico de acordo com o procedimento descrito na Subseção 2.4.5.

A Figura 3.1 apresenta o gráfico da discrepâncias relativas para as estatísticas LR , LR_m , LR_m^* , LR_{Boot}^* , S , G e G^* . Quanto mais próxima de zero a curva da discrepância estiver, melhor a aproximação da distribuição exata sob a hipótese nula da estatística de teste pela distribuição limite χ^2 . Observe que as discrepâncias relativas baseadas nas estatísticas LR_m^* , LR_{Boot}^* e G^* são mais próximas de zero, mostrando uma boa aproximação das distribuição dessas estatísticas para a distribuição qui-quadrado. As estatísticas da razão de verossimilhanças perfiladas modificadas (LR_m) e escore (S) apresentam discrepâncias relativas próximas de zero, porém abaixo de zero, confirmando o comportamento conservativo dos testes baseados nestas estatísticas. Em contrapartida, as estatísticas LR e G apresentam-se curvas afastadas de zero, confirmando que os testes baseados nestas estatísticas são liberais.

Tabela 3.1: Tamanhos empíricos dos testes LR, LR_m, LR_{Boot}, S, G, LR_m^{*}, LR_{Boot}^{*} e G^{*}- Modelo normal não linear para $n = 35$, $q = 2$ e diferentes valores de p .

p	$\alpha(\%)$	LR	LR _m	LR _m [*]	LR _{Boot}	LR _{Boot} [*]	S	G	G [*]
2	10,0	13,8	9,6	10,3	9,7	9,8	8,8	13,1	10,2
	5,0	7,7	4,5	5,1	4,9	4,8	4,3	7,0	4,8
	1,0	1,9	0,8	1,0	1,0	0,9	0,8	1,6	0,9
	0,5	1,0	0,4	0,5	0,6	0,5	0,4	0,8	0,4
3	10,0	14,9	8,8	9,9	9,8	9,6	8,6	14,1	9,5
	5,0	8,2	4,2	4,8	5,0	4,9	4,3	7,6	4,5
	1,0	2,3	0,7	0,9	1,0	0,9	0,8	1,9	0,7
	0,5	1,2	0,3	0,4	0,6	0,4	0,4	0,9	0,3
4	10,0	16,4	8,8	9,6	9,9	10,0	8,3	15,7	9,6
	5,0	9,9	4,2	4,7	5,3	5,1	3,8	9,0	4,7
	1,0	2,7	0,8	0,9	1,2	1,0	0,7	2,3	0,8
	0,5	1,7	0,4	0,5	0,6	0,5	0,3	1,3	0,4
5	10,0	20,3	9,0	10,0	10,4	10,3	8,9	19,2	10,6
	5,0	12,7	4,4	5,0	5,3	5,2	4,2	11,6	5,3
	1,0	4,2	0,7	0,9	1,2	1,0	0,7	3,6	0,8
	0,5	2,7	0,3	0,4	0,7	0,6	0,4	2,0	0,4
6	10,0	22,5	8,5	9,4	10,3	10,3	8,2	21,4	10,9
	5,0	14,4	4,1	4,7	5,4	5,3	4,0	13,2	5,3
	1,0	5,2	0,8	1,0	1,1	0,9	0,8	4,6	0,9
	0,5	3,4	0,4	0,5	0,6	0,4	0,4	2,9	0,4
7	10,0	27,5	8,9	9,8	10,7	10,5	8,4	25,3	12,3
	5,0	18,6	4,5	5,0	5,5	5,3	3,9	16,3	6,1
	1,0	7,4	0,8	0,9	1,5	1,2	0,8	5,8	1,1
	0,5	5,4	0,4	0,5	0,9	0,6	0,4	3,8	0,5
8	10,0	34,1	8,1	9,0	10,5	10,3	7,8	31,9	13,1
	5,0	24,5	3,8	4,3	5,3	5,0	3,7	21,9	6,9
	1,0	11,5	0,6	0,7	1,2	0,9	0,6	9,3	1,4
	0,5	8,1	0,3	0,4	0,7	0,5	0,3	6,5	0,6

Tabela 3.2: Tamanhos empíricos dos testes LR, LR_m, LR_{Boot}, S, G, LR_m^{*}, LR_{Boot}^{*} e G^{*} - Modelo normal não linear para $n = 35$, $p = 5$ e diferentes valores de q .

q	$\alpha(\%)$	LR	LR _m	LR _m [*]	LR _{Boot}	LR _{Boot} [*]	S	G	G [*]
1	10,0	17,0	9,1	9,8	10,2	10,1	8,8	16,5	9,1
	5,0	10,4	4,4	4,8	5,2	5,2	4,3	10,0	4,4
	1,0	3,2	0,8	0,9	1,1	1,1	0,6	2,8	0,8
	0,5	2,0	0,3	0,5	0,7	0,5	0,3	1,6	0,4
2	10,0	20,3	9,0	10,0	10,4	10,3	8,9	19,2	10,6
	5,0	12,7	4,4	5,0	5,3	5,2	4,2	11,6	5,3
	1,0	4,2	0,7	0,9	1,2	1,0	0,7	3,6	0,8
	0,5	2,7	0,3	0,4	0,7	0,6	0,4	2,0	0,4
3	10,0	22,9	9,1	10,0	10,5	10,3	8,5	21,5	12,0
	5,0	14,8	4,2	4,8	5,3	5,2	4,3	13,3	5,9
	1,0	5,3	2,7	0,9	1,1	0,9	0,8	4,0	1,0
	0,5	3,1	0,3	0,4	0,6	0,4	0,5	2,3	0,4
4	10,0	28,5	9,1	10,0	9,7	9,8	8,0	25,9	13,6
	5,0	18,7	4,0	4,7	4,6	4,7	4,1	16,7	6,7
	1,0	7,5	0,6	0,8	1,0	1,1	0,8	5,9	1,1
	0,5	5,0	0,3	0,4	0,6	0,4	0,4	3,9	0,4
5	10,0	34,1	8,5	9,7	9,9	10,0	7,8	30,7	15,8
	5,0	24,0	4,0	4,7	5,1	5,2	4,0	20,6	8,2
	1,0	10,1	0,6	0,8	1,2	1,2	0,9	8,0	1,1
	0,5	6,7	0,2	0,3	0,7	0,7	0,6	5,1	0,3
6	10,0	40,5	9,4	10,4	10,4	10,9	8,2	36,5	20,3
	5,0	30,3	4,7	5,3	5,0	5,8	4,2	26,2	11,3
	1,0	14,9	0,9	1,0	1,0	1,3	1,1	12,0	2,3
	0,5	11,1	0,4	0,5	0,5	0,8	0,6	8,6	0,8
7	10,0	44,8	9,3	10,4	10,3	10,9	8,5	39,9	22,8
	5,0	34,4	4,7	5,1	5,5	6,2	4,9	29,5	12,8
	1,0	17,8	0,8	0,9	1,1	1,7	1,3	14,6	2,5
	0,5	13,4	0,3	0,5	0,7	1,0	0,7	10,7	0,7
8	10,0	52,4	10,0	10,7	9,5	10,2	7,2	45,6	26,7
	5,0	41,6	4,9	5,5	4,7	5,6	4,0	35,2	16,1
	1,0	23,6	0,9	0,9	1,1	1,4	1,0	19,1	2,6
	0,5	18,8	0,4	0,5	0,6	0,8	0,5	14,9	0,0

Tabela 3.3: Tamanhos empíricos dos testes LR, LR_m, LR_{Boot}, S, G, LR_m^{*}, LR_{Boot}^{*} e G^{*} - Modelo normal não linear para $p = 5$, $q = 2$ e diferentes tamanhos amostrais.

n	$\alpha(\%)$	LR	LR _m	LR _m [*]	LR _{Boot}	LR _{Boot} [*]	S	G	G [*]
25	10,0	28,6	8,6	10,1	10,1	10,0	7,9	25,9	11,5
	5,0	19,2	4,0	5,0	5,0	4,8	4,0	16,8	5,4
	1,0	7,9	0,7	0,9	1,0	0,8	0,8	5,8	0,8
	0,5	5,3	0,3	0,4	0,6	0,4	0,4	3,6	0,4
35	10,0	20,3	9,0	10,0	10,4	10,3	8,9	19,2	10,6
	5,0	12,7	4,4	5,0	5,3	5,2	4,2	11,6	5,3
	1,0	4,2	0,7	0,9	1,2	1,0	0,7	3,6	0,8
	0,5	2,7	0,3	0,4	0,7	0,6	0,4	2,0	0,4
45	10,0	16,5	8,8	9,5	9,6	9,4	8,0	16,0	9,8
	5,0	9,8	4,4	4,8	4,8	4,7	4,1	9,3	5,9
	1,0	2,9	0,8	0,8	1,1	0,9	0,7	2,6	0,8
	0,5	1,7	0,4	0,5	0,7	0,5	0,3	1,4	0,3
100	10,0	14,3	9,0	9,6	10,2	10,2	9,5	13,7	9,7
	5,0	8,2	4,5	4,9	5,3	5,3	4,5	7,8	4,8
	1,0	2,2	0,9	1,0	1,2	1,1	0,9	1,9	0,9
	0,5	1,3	0,5	0,6	0,7	0,5	0,4	1,1	0,5

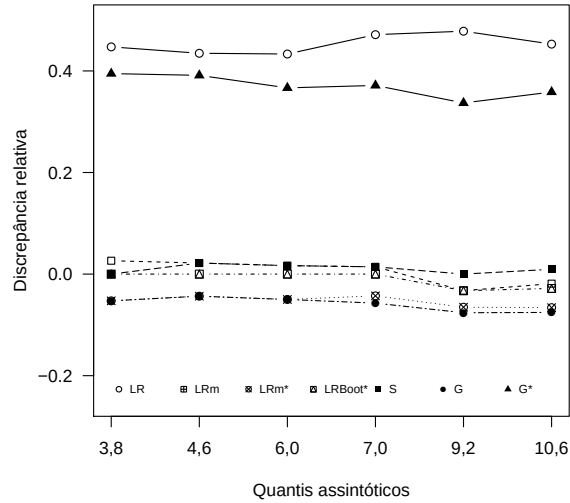
Tabela 3.4: Poder dos testes LR_m , LR_{Boot} , S , LR_m^* , LR_{Boot}^* e G^* - Modelo normal não linear para $n = 35$, $p = 5$, $q = 2$ e $\alpha = 10,0\%$.

δ	LR_m	LR_m^*	LR_{Boot}	LR_{Boot}^*	S	G^*
0,0	9,0	10,0	10,4	10,3	8,9	10,6
0,5	13,9	14,9	13,5	13,3	13,8	17,7
1,0	27,1	28,9	26,2	26,1	27,3	27,5
1,5	51,4	53,3	48,4	48,3	47,6	48,6
2,0	77,0	78,4	71,2	71,2	76,1	79,0
2,5	91,3	92,0	87,6	87,7	90,3	90,1
3,0	94,7	95,0	95,5	95,5	90,6	91,7
3,5	98,3	99,7	98,4	98,4	96,0	98,8

Tabela 3.5: Quantis das estatísticas LR, LR_m, S, G, LR_m^{*}, LR_{Boot}^{*} e G^{*} - Modelo normal não linear para $p = 5$, $q = 2$ e diferentes tamanhos amostrais.

n	quantis(%)	χ^2_2	LR	LR _m	LR _m [*]	LR _{Boot} [*]	S	G	G [*]
25	85,0	3,8	6,9	3,6	3,8	3,8	3,5	6,3	4,1
	90,0	4,6	8,4	4,3	4,6	4,6	4,2	7,5	4,8
	95,0	6,0	10,8	5,7	6,0	5,9	5,5	9,7	6,1
	97,0	7,0	12,5	6,6	7,0	6,8	6,6	11,1	6,6
	99,0	9,2	16,5	8,5	9,1	9,0	8,7	14,5	8,9
	99,5	10,6	19,2	9,8	10,4	10,0	10,0	16,4	9,9
35	85,0	3,8	5,5	3,6	3,8	3,8	3,6	5,3	3,9
	90,0	4,6	6,6	4,4	4,6	4,7	4,4	6,4	4,7
	95,0	6,0	8,6	5,7	6,0	6,1	5,7	8,2	6,1
	97,0	7,0	10,3	6,7	7,0	7,1	6,6	9,6	7,1
	99,0	9,2	13,6	8,6	8,9	9,2	8,5	12,3	8,9
	99,5	10,6	15,4	9,9	10,3	10,7	9,8	14,4	10,4
45	85,0	3,8	5,0	3,7	3,8	3,7	3,4	4,8	3,9
	90,0	4,6	6,0	4,4	4,6	4,5	4,2	5,8	4,6
	95,0	6,0	7,8	5,8	6,0	5,9	5,6	7,6	6,0
	97,0	7,0	9,1	6,8	7,0	6,9	6,6	8,8	7,0
	99,0	9,2	12,6	9,5	11,7	9,1	8,6	11,7	9,2
	99,5	10,6	14,1	10,3	10,7	10,5	9,7	13,4	10,5
100	85,0	3,8	4,1	3,7	3,7	3,8	3,7	4,1	3,8
	90,0	4,6	5,1	4,5	4,6	4,7	4,5	5,0	4,6
	95,0	6,0	6,5	5,8	5,9	6,1	5,8	6,4	5,9
	97,0	7,0	7,6	6,8	6,9	7,1	6,8	7,6	6,9
	99,0	9,2	10,4	9,3	9,4	9,3	8,9	10,2	9,3
	99,5	10,6	11,9	10,6	10,8	10,5	10,5	11,7	10,7

Figura 3.1: Discrepâncias relativas de quantis das estatísticas LR, LR_m, S, G, LR_m^{*}, LR_{Boot}^{*}, G^{*}
- Modelo normal não linear com $n = 35$, $p = 5$ e $q = 2$.



Fonte: Elaborada pela autora.

Tabela 3.6: Média e variância das estatísticas LR, LR_m, LR_{Boot}, S, G, LR_m^{*}, LR_{Boot}^{*} e G^{*} - Modelo normal não linear para $p = 5$ e $n = 35$ e diferentes valores de q .

q	momentos	χ_q^2	LR	LR _m	LR _m [*]	LR _{Boot}	LR _{Boot} [*]	S	G	G [*]
1	média	1,000	1,457	0,943	0,984	1,443	1,022	0,929	1,395	0,940
	variância	2,000	4,831	1,775	1,934	4,194	2,040	1,645	3,755	1,727
2	média	2,000	2,890	1,891	1,970	2,594	2,030	1,868	2,769	2,061
	variância	4,000	8,956	3,583	3,890	6,495	4,143	3,608	7,383	3,956
3	média	3,000	4,360	2,872	2,983	4,138	3,036	2,834	4,183	3,222
	variância	6,000	12,858	5,441	5,872	11,694	5,943	5,543	11,132	6,130
4	média	4,000	6,176	3,839	3,979	6,022	3,980	3,680	5,879	4,467
	variância	8,000	19,355	7,188	7,723	20,303	8,015	7,355	16,850	8,554
5	média	5,000	8,108	4,819	4,974	8,061	5,016	4,572	7,659	5,864
	variância	10,000	25,824	8,789	9,364	28,677	10,496	9,309	22,462	10,975
6	média	6,000	10,555	5,899	6,069	11,058	6,027	5,541	9,910	7,507
	variância	12,000	41,032	11,510	12,184	48,705	13,121	12,015	36,182	15,044
7	média	7,000	12,638	6,916	7,094	12,623	7,003	6,500	11,796	8,926
	variância	14,000	52,210	13,287	13,981	56,384	16,567	14,861	46,152	17,480
8	média	8,000	15,374	7,997	8,148	15,976	7,933	7,401	14,233	10,562
	variância	16,000	67,329	15,539	16,135	76,761	17,481	17,103	62,405	20,246

3.8 Aplicações

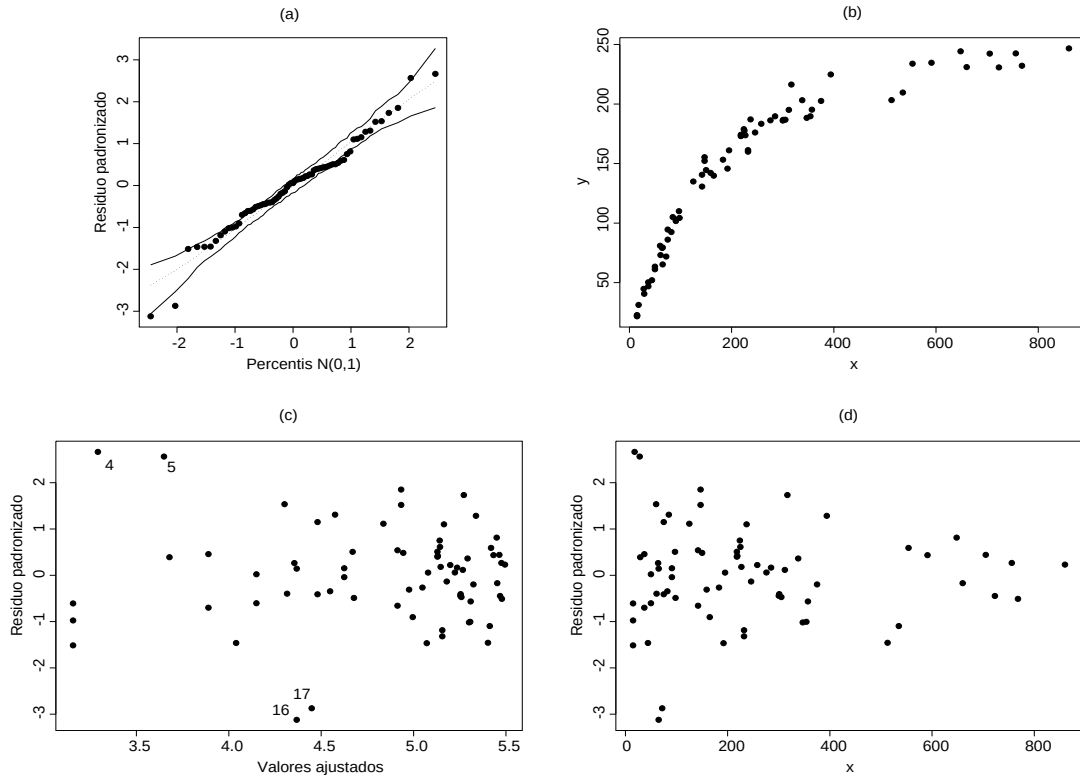
3.8.1 Exemplo 1

Este exemplo utiliza dados analisados por Wei (1998, p.155) e Cysneiros e Ferrari (2006) apresentados no Apêndice D. Os dados analisados representam os pesos em mg (y) das lentes dos olhos dos coelhos europeus (*Oryctolagus Cuniculus*) e à idade média do animal (x) em dias numa amostra composta por 71 observações. O modelo utilizado foi:

$$\log(y_\ell) = \beta_1 - \frac{\beta_2}{x_\ell - \beta_3} + u_\ell, \quad \ell = 1, \dots, 71,$$

em que $\text{Cov}(u_\ell, u_m) = 0$ para $\ell \neq m$. Inicialmente Cysneiros e Ferrari (2006) assumiram que $u_\ell \sim N(0, \sigma^2)$ e estimaram os parâmetros do modelo por máxima verossimilhança, obtendo $\beta_1 = 5,640$, $\beta_2 = 130,58$, $\beta_3 = 37,603$ e $\sigma^2 = 0,004$. Observe que os gráficos da Figura 3.2 mostram que as observações 4, 5, 16 e 17 apresentam resíduos grandes (em valor absoluto) apresentando alguma evidência de heteroscedasticidade.

Figura 3.2: (a) Gráfico dos resíduos padronizados contra percentis normais com envelope simulado; (b) Gráfico de dispersão do peso das lentes dos olhos contra a idade; (c) Gráfico dos resíduos studentizados t_{s_i} contra os valores ajustados e (d) Gráfico dos resíduos studentizados t_{s_i} contra x para o modelo homoscedástico - Coelhos europeus na Austrália.



Fonte: Elaborada pelas autoras Cysneiros e Ferrari (2006).

Então, considerando o modelo heteroscedástico, em que $u_\ell \sim N(0, \sigma^2 \exp\{\delta x\})$, o nosso

interesse consiste em testar $\mathcal{H}_0 : \delta = 0$ (homoscedasticidade) contra $\mathcal{H}_1 : \delta \neq 0$.

Para este teste, temos $LR = 10,250$, $LR_m = 9,556$, $LR_m^* = 9,776$, $G = 3,944$ e $G^* = 3,541$ que conduzem aos respectivos níveis descritivos 0,001, 0,001, 0,002, 0,047 e 0,060. Segue então que os testes baseados nas estatísticas da razão de verossimilhanças (LR), razão de verossimilhanças perfiladas modificada (LR_m), sua versão corrigida (LR_m^*), gradiente (G) e gradiente corrigida (G^*) conduzem à rejeição da hipótese nula ao nível de significância de 10%. Observe que a 5% o teste G rejeita a hipótese nula enquanto que o teste G^* não rejeita a hipótese nula.

3.8.2 Exemplo 2

Neste segundo exemplo foram considerados os dados analisados por Rawlings (1988) e Cysneiros e Ferrari (2006) apresentados no Apêndice E. Howard Grimes do Departamento de Botânica da North Carolina State University realizou um experimento para análise bioquímica de armazenamento intracelular e transporte de cálcio através da membrana plasmática. O experimento consistiu em células suspensas em uma solução de cálcio radioativo por um certo período de tempo x (em minutos) e então a quantidade de cálcio radioativo y (em nmoles/mg) absorvida foi medida numa amostra com 27 observações. O modelo usado foi:

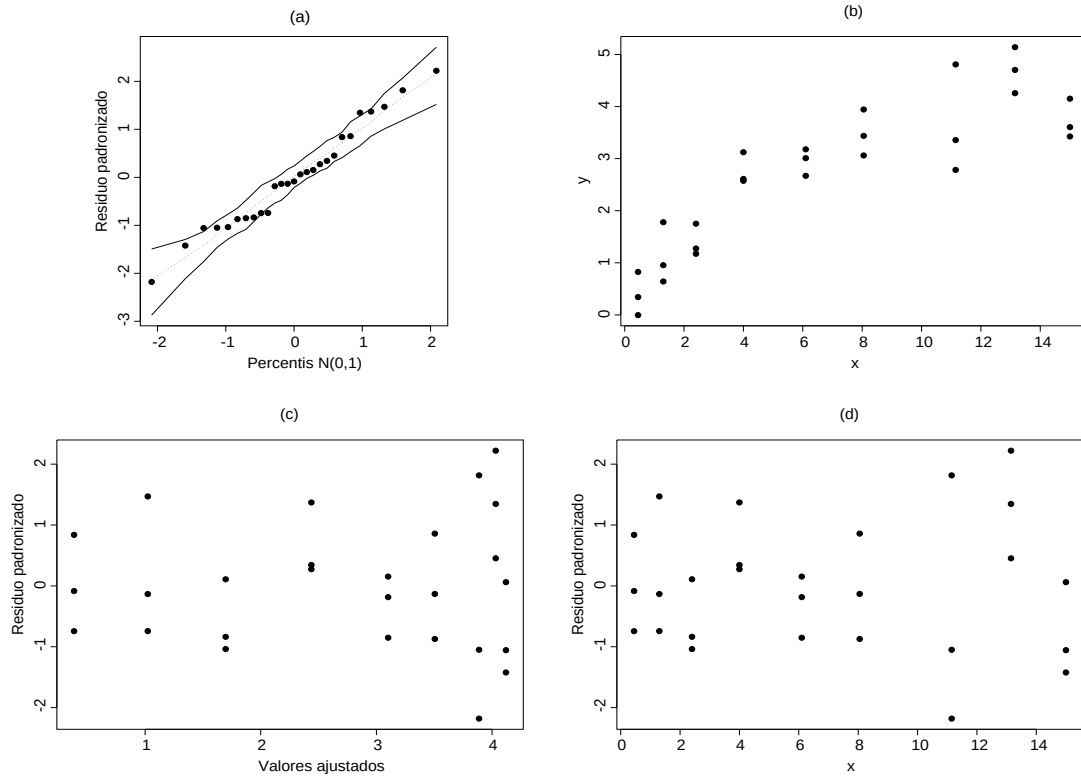
$$y_\ell = \beta_0 \{1 - \exp(-\beta_1 x_\ell)\} + u_\ell, \quad \ell = 1, \dots, 27,$$

em que $\text{Cov}(u_\ell, u_m) = 0$ para $\ell \neq m$. Inicialmente Cysneiros e Ferrari (2006) assumiram que $u_\ell \sim N(0, \sigma^2)$ e estimaram os parâmetros do modelo por máxima verossimilhança, obtendo $\beta_0 = 4,309$, $\beta_1 = 0,208$ e $\sigma^2 = 0,276$. Os gráficos da Figura 3.3 mostram alguma evidência de heteroscedasticidade.

Então, considerando o modelo heteroscedástico, em que $u_\ell \sim N(0, \sigma^2 \exp\{\delta x\})$, o interesse é testar $\mathcal{H}_0 : \delta = 0$ (homoscedasticidade) contra $\mathcal{H}_1 : \delta \neq 0$.

Para este teste, temos $LR = 3,097$, $LR_m = 2,914$, $LR_m^* = 3,056$, $G = 3,055$ e $G^* = 2,540$ que conduzem aos respectivos níveis descritivos 0,078, 0,088, 0,080, 0,081 e 0,111. Segue então que os testes baseados nas estatísticas da razão de verossimilhanças (LR), razão de verossimilhanças perfiladas modificada (LR_m), sua versão corrigida (LR_m^*) e gradiente (G) conduzem a rejeição da hipótese nula ao nível de 10%, por outro lados o teste baseado na estatística gradiente corrigida (G^*) não rejeita $\mathcal{H}_0$. Para o nível de significância de 5% nenhum dos testes rejeitam a hipótese nula.

Figura 3.3: (a) Gráfico dos resíduos padronizados contra percentis normais com envelope simulado; (b) Gráfico de dispersão da quantidade absorvida de cálcio radioativo contra o tempo; (c) Gráfico dos resíduos studentizados t_{s_i} contra os valores ajustados e (d) Gráfico dos resíduos studentizados t_{s_i} contra x para o modelo homoscedástico - Solução de cálcio radioativo.



Fonte: Elaborada pelas autoras Cysneiros e Ferrari (2006).

3.9 Comentários

Trabalhamos neste capítulo com testes de heteroscedasticidade em modelos não lineares da família exponencial com dispersão variável. Em especial utilizamos o modelo de regressão não linear reparmetrizado. Admitimos uma forma funcional que relaciona os parâmetros de dispersão com alguns parâmetros desconhecidos, que não dependem do vetor de parâmetros de regressão e algumas variáveis auxiliares.

Como maior contribuição deste trabalho derivamos um fator de correção tipo-Bartlett para a estatística gradiente. Mostramos a partir dos resultados da simulação Monte Carlo que o teste baseado na estatística gradiente não apresenta uma aproximação muito boa para a distribuição de referência χ^2 . Vale salientar que esse fator de correção tipo-Bartlett é válido para as distribuições normal e normal inversa, pois para a distribuição gamma não é possível obter uma transformação em que implica na ortogonalidade geral dos parâmetros.

Comparamos o desempenho em amostras finitas de nove testes, a saber: o teste da razão de verossimilhanças original (LR), o teste da razão de verossimilhanças perfiladas modificadas LR_m , o teste da razão de verossimilhanças bootstrap LR_{Boot} , o teste escore S, o teste gradiente

G, o teste da razão de verossimilhanças perfiladas modificadas corrigido LR_m^* , o teste da razão de verossimilhanças bootstrap corrigido LR_{Boot}^* e o teste G^* proposto como objeto de estudo dessa dissertação.

Inicialmente o tamanho dos testes levou a concluir que os testes baseados nas estatísticas LR e G são liberais, conduzindo a rejeição da hipótese de homoscedasticidade erroneamente com uma probabilidade maior do que os níveis nominais considerados. Por outro o lado, os testes LR_m e S exibiram taxas de rejeição menores que os níveis nominais dos testes. Os testes LR_m^* , LR_{Boot} , LR_{Boot}^* , G^* apresentaram taxas de rejeição mais próximas dos níveis nominais. Posteriormente, fizemos estudos sobre o impacto dos parâmetros de interesse e perturbação nos testes e vimos que houve um impacto considerável principalmente nos testes baseados nas estatísticas LR e G. Esse impacto também foi observado no teste baseado na estatística G^* porém de forma mais moderada, sobretudo quando aumentamos o número de parâmetros de interesse. No entanto, vemos claramente o quanto a correção tipo-Bartlett melhorou o teste gradiente, tornando-o um teste mais preciso.

Nas simulações envolvendo poder do teste, não consideramos os testes baseados nas estatísticas LR e G por serem muito liberais, sendo os testes LR_m^* , LR_{Boot} , LR_{Boot}^* e G^* não significativamente mais poderoso que o outro.

Por fim, dos quatro testes com melhores desempenhos LR_m^* , LR_{Boot} , LR_{Boot}^* e G^* podemos recomendar o teste G^* , pois apresenta um bom desempenho e pela sua simplicidade e tempo de execução baixo, já que não exige a matriz de informação de Fisher e nem a sua inversa, que em muitas situações práticas pode ser difícil ou mesma não possível de ser calculada. Além disso, o fator de correção tipo-Bartlett envolve operações simples com vetores e matrizes.

Capítulo 4

Considerações finais

A principal contribuição teórica deste trabalho foi a obtenção de um fator de correção tipo-Bartlett para um teste de heteroscedasticidade baseado na estatística gradiente nos modelos não lineares da família exponencial (MNLFE). O fator de correção foi obtido a partir dos resultados apresentados por (VARGAS; FERRARI; LEMONTE, 2013). Consideramos os modelos de regressão não linear normal e normal inversa, pois utilizamos uma transformação que sob essa nova parametrização garantimos a ortogonalidade entre os parâmetros facilitando o cálculo dos cumulantes conjuntos de derivadas do logaritmo da função de verossimilhança e o cálculo das derivadas desses cumulantes conjuntos. Para o modelo gama não é possível obter uma reparametrização que torna os parâmetros, em geral, ortogonais.

Nessa dissertação fizemos um estudo de comparação dos desempenhos dos testes em relação ao tamanho e ao poder. No Capítulo 2 comparamos apenas os testes baseados na estatística razão de verossimilhanças, na razão de verossimilhanças perfiladas modificada proposta por (COX; REID, 1987), na estatística escore introduzida por (RAO, 1973), na estatística razão de verossimilhanças bootstrap e na estatística gradiente proposta recentemente por (TERRELL, 2002). O tamanho dos testes indicou que os testes razão de verossimilhanças e gradiente, são em geral, liberais, conduzindo a taxas de rejeição da hipótese nula com demasiada frequência. Por outro lado, o teste da razão de verossimilhanças perfiladas modificado baseado na LR_m mostrou-se eficiente corrigindo a tendência liberal do teste original. Entretanto, exibiu taxas abaixo dos níveis nominais considerados para o teste. Similarmente, o teste escore apresentou o mesmo comportamento conservativo. Deste capítulo o teste que mostrou melhor desempenho foi o teste baseado na estatística razão de verossimilhanças utilizando reamostragem bootstrap.

No Capítulo 3 discutiu-se alguns testes de heteroscedasticidade para as estatísticas do Capítulo 2 em suas versões corrigidas, exceto para a estatística escore. Desenvolvemos o fator de correção tipo-Bartlett à estatística gradiente nos MNLFE e fizemos uma comparação dos desempenhos dos testes abordados nos Capítulos 2 e 3. Das quais tiramos as seguintes conclusões:

- Os resultados para os tamanhos dos testes indicaram que a correção de Bartlett atenuou fracamente a tendência liberal do teste baseado na estatística razão de verossimilhanças, não houve também uma melhora significativa quando utilizamos a razão de verossimilhanças bootstrap corrigida proposta por (ROCKE, 1989), visto que o teste baseado na

razão de verossimilhanças já apresentava um bom desempenho. Em destaque, o teste gradiente corrigido mostrou-se bastante eficiente corrigindo a tendência liberal do teste original, levando as taxas de rejeição próximas dos níveis nominais.

- Analisou-se estudos de simulação para avaliar o desempenho dos testes quando aumentamos o número de parâmetros de perturbação e interesse. Neste cenário observou-se que os testes baseados nas estatísticas *score*, razão de verossimilhanças perfiladas modificadas, razão de verossimilhanças bootstrap e essas duas últimas em suas versões corrigidas não sofreram impactos significativos tanto para o aumento de parâmetros de perturbação quanto para o aumento de parâmetros de interesse. Em relação aos testes baseados nas estatísticas razão de verossimilhanças original, gradiente e em suas versões corrigidas foram os que apresentaram impacto nas taxas de rejeição, principalmente quando aumentamos o número de parâmetros de interesse. Ainda assim, o teste baseado na estatística gradiente corrigida foi o que teve melhor desempenho quando aumentamos o número de parâmetros de perturbação e interesse comparados aos testes que sofreram impacto significativo.
- Nas simulações envolvendo o poder desconsideramos os testes baseados nas estatísticas razão de verossimilhanças e gradiente, pois estes testes apresentaram taxas de rejeição distorcidas, portanto não podem ser recomendados. Observou-se que os quatros testes: LR_m^* , LR_{Boot} , LR_{Boot}^* e G^* não são significativamente mais poderosos que o outro para os níveis nominais altos.
- Em síntese, os testes que apresentaram melhor desempenho para testar a presença de heteroscedasticidade formam os testes LR_m^* , LR_{Boot} , LR_{Boot}^* e G^* . Vale ressaltar que estes quatro testes apresentam comportamentos semelhantes quando $p \leq 6$ e número de parâmetros de interesse pequeno. Salientamos que o fator de correção tipo-Bartlett corrige eficientemente o comportamento não-conservativo do teste baseado na estatística gradiente. Por isso, pode ser recomendável a sua utilidade em testes de heteroscedasticidade nos MNLFE.

Referências

- BARTLETT, M. S. Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, JSTOR, p. 268–282, 1937.
- BATES, D.; WATTS, D. *Nonlinear Regression Analysis and Its Applications*. second. [S.l.]: New York: Wiley, 1988.
- CAVALCANTI, A. B. *Aperfeiçoamento de métodos estatísticos em modelos de regressão da família exponencial*. Tese (Doutorado) — Universidade de São Paulo, 2009.
- CORDEIRO, G. M. Bartlett corrections and bias correction for two heteroscedastic regression models. *Communications in Statistics-Theory and Methods*, Taylor & Francis, v. 22, n. 1, p. 169–188, 1993.
- CORDEIRO, G. M.; FERRARI, S. L. D. P. A modified score test statistic having chi-squared distribution to order n^{-1} . *Biometrika*, v. 78, n. 3, p. 573–582, 1991.
- CORDEIRO, G. M.; PAULA, G. A. Improved likelihood ratio statistics for exponential family nonlinear models. *Biometrika*, Biometrika Trust, v. 76, n. 1, p. 93–100, 1989.
- COX, D.; HINKLEY, D. *Theoretical Statistics*. [S.l.]: Chapman and Hall, 1974. Distributed in the U.S.A. by Halsted Press, New York.
- COX, D. R.; REID, N. Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 1–39, 1987.
- CRIBARI-NETO, F.; ZARKOS, S. G. R. Yet another econometric programming environment. *Journal of Applied Econometrics*, v. 14, n. 3, p. 319–329, 1999.
- CYSNEIROS, A. H.; FERRARI, S. L. An improved likelihood ratio test for varying dispersion in exponential family nonlinear models. *Statistics & probability letters*, Elsevier, v. 76, n. 3, p. 255–265, 2006.
- DICICCIO, T. J.; STERN, S. E. Frequentist and bayesian bartlett correction of test statistics based on adjusted profile likelihoods. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 397–408, 1994.
- DOORNIK, J. A. An object-oriented matrix programming language ox 6. Timberlake Consultants Ltd, 2009.
- EFRON, B. Bootstrap methods: another look at the jackknife. *The annals of Statistics*, JSTOR, p. 1–26, 1979.

- EFRON, B.; TIBSHIRANI, R. J. *An introduction to the bootstrap*. [S.l.]: CRC press, 1994.
- GHOSH, J.; MUKERJEE, R. Characterization of priors under which bayesian and frequentist bartlett corrections are equivalent in the multiparameter case. *Journal of Multivariate Analysis*, Elsevier, v. 38, n. 2, p. 385–393, 1991.
- LAWLEY, D. A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, JSTOR, p. 295–303, 1956.
- LEMONTE, A. J. Local power of some tests in exponential family nonlinear models. *Journal of Statistical Planning and Inference*, v. 141, n. 5, 2011.
- LEMONTE, A. J. Nonnull asymptotic distributions of the lr, wald, score and gradient statistics in generalized linear models with dispersion covariates. *Statistics*, v. 47, n. 6, p. 1249–1265, 2013.
- LEMONTE, A. J.; FERRARI, S. L. Size and power properties of some tests in the birnbaum–saunders regression model. *Computational Statistics & Data Analysis*, v. 55, n. 2, p. 1109–1117, 2011.
- LEMONTE, A. J.; FERRARI, S. L. Local power and size properties of the lr, wald, score and gradient tests in dispersion models. *Statistical Methodology*, v. 9, n. 5, p. 537–554, 2012.
- LEMONTE, A. J.; FERRARI, S. L. A note on the local power of the lr, wald, score and gradient tests. *Electronic Journal of Statistics*, Institute of Mathematical Statistics, v. 6, p. 421–434, 2012.
- NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society A*, v. 135, p. 370–384, 1972.
- NEYMAN, J.; PEARSON, E. S. On the use and interpretation of certain test criteria for purposes of statistical inference: Part i. *Biometrika*, JSTOR, p. 175–240, 1928.
- NOCEDAL, J.; WRIGHT, S. J. *Numerical optimization*. [S.l.]: Springer New York, 1999. v. 2.
- RAO, C. Linear statistical inference and its applications. *Wiley series in probability and mathematical statistics Show all parts in this series*, New York [etc.]: Wiley, 1973.
- RAWLINGS, J. Applied regression analysis. wadsworth & brooks. *Cole, California*, p. 553, 1988.
- ROCKE, D. M. Bootstrap bartlett adjustment in seemingly unrelated regression. *Journal of the American Statistical Association*, Taylor & Francis, v. 84, n. 406, p. 598–601, 1989.
- SEBER, G.; WILD, C. *Nonlinear Regression*. [S.l.]: New York: Wiley, 1989.
- SIMONOFF, J. S.; TSAI, C.-L. Use of modified profile likelihood for improved tests of constancy of variance in regression. *Applied Statistics*, JSTOR, p. 357–370, 1994.
- TERRELL, G. R. The gradient statistic. *Computing Science and Statistics*, v. 34, p. 206–215, 2002.
- VARGAS, T. M.; FERRARI, S. L.; LEMONTE, A. J. Gradient statistic: Higher-order asymptotics and bartlett-type correction. *Electronic Journal of Statistics*, Institute of Mathematical Statistics, v. 7, p. 43–61, 2013.

WALD, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical society*, JSTOR, v. 54, n. 3, p. 426–482, 1943.

WEI, B.-C. *Exponential family nonlinear models*. [S.l.]: Springer Verlag, 1998. v. 130.

WILKS, S. S. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, JSTOR, v. 9, n. 1, p. 60–62, 1938.

Apêndice A

Cálculo dos cumulantes nos modelos não lineares da família exponencial com dispersão variável

Apresentamos alguns cumulantes conjuntos de derivadas do logaritmo da função de verossimilhança nos MNLFE reparametrizados com dispersão variável que são necessários para correção tipo-Bartlett na estatística gradiente. Calculamos estes cumulantes a partir do logaritmo da função de verossimilhança dado em (2.7). Para a estatística gradiente corrigida pelo fator tipo-Bartlett utiliza-se cumulantes conjuntos até de quarta ordem, por isso adotamos os índices i, j, k, l para as componentes do vetor de parâmetros β e os índices a, b, c, d para as componentes do vetor de parâmetros δ . Utilizando a ortogonalidade entre os parâmetros δ, β, γ e algumas condições de regularidade temos que os cumulantes de segunda ordem são dados por

$$\kappa_{ij} = -\frac{1}{\gamma} \sum_{\ell=1}^n x_{\ell_i}^* q_{\ell} w_{\ell} x_{\ell_j}^*,$$

$$\kappa_{ab} = -\frac{1}{2} \sum_{\ell=1}^n \frac{q_{\ell_{ab}}}{q_{\ell}},$$

$$\kappa_{\gamma\gamma} = -\frac{n}{2\gamma^2},$$

e

$$\kappa_{\beta\delta} = \kappa_{\beta\gamma} = \kappa_{\delta\gamma} = 0,$$

em que $w_{\ell} = (d\mu_{\ell}/d\eta_{\ell})^2 (d\theta_{\ell}/d\mu_{\ell})$, $x_{\ell_i}^* = \partial\eta_{\ell}/\partial\beta^i$ e $q_{\ell_{ab}} = \partial^2 q_{\ell}/\partial\delta^a \partial\delta^b$, para $i = 1, \dots, p$ e $a, b = 1, \dots, q$.

As derivadas primeira dos cumulantes de segunda ordem são dadas por

$$\kappa_{ij}^{(k)} = -\frac{1}{\gamma} \sum_{\ell=1}^n q_{\ell} w_{\ell} (x_{\ell_i}^* x_{\ell_{jk}}^{**} + x_{\ell_j}^* x_{\ell_{ik}}^{**} + x_{\ell_k}^* x_{\ell_{ij}}^{**}) + \frac{1}{\gamma} \sum_{\ell=1}^n 2q_{\ell} \frac{\partial \theta_{\ell}}{\partial \eta_{\ell}} x_{\ell_i}^* x_{\ell_j}^* x_{\ell_k}^*,$$

$$\kappa_{ab}^{(c)} = -\frac{1}{2} \sum_{\ell=1}^n \frac{\partial}{\partial \delta^c} \left(\frac{q_{\ell ab}}{q_{\ell}} \right),$$

$$\kappa_{\gamma\gamma} = \frac{n}{\gamma^3},$$

$$\kappa_{ij}^{(a)} = -\frac{1}{\gamma} \sum_{\ell=1}^n x_{\ell_i}^* q_{\ell a} w_{\ell} x_{\ell_j}^*,$$

$$\kappa_{ij}^{(\gamma)} = \frac{1}{\gamma^2} \sum_{\ell=1}^n x_{\ell_i}^* q_{\ell} w_{\ell} x_{\ell_j}^*,$$

e

$$\kappa_{ab}^{(i)} = \kappa_{ab}^{(\gamma)} = \kappa_{\gamma\gamma}^{(i)} = \kappa_{\gamma\gamma}^{(a)} = \kappa_{ia}^{(j)} = \kappa_{ia}^{(b)} = \kappa_{ia}^{(\gamma)} = \kappa_{i\gamma}^{(j)} = \kappa_{i\gamma}^{(a)} = \kappa_{i\gamma}^{(\gamma)} = \kappa_{a\gamma}^{(i)} = \kappa_{a\gamma}^{(b)} = \kappa_{a\gamma}^{(\gamma)} = 0.$$

em que $x_{\ell_{ij}}^{**} = \partial^2 \eta_{\ell} / \partial \beta^i \partial \beta^j$, para $i, j = 1, \dots, p$.

As derivadas de segunda ordem dos cumulantes de segunda ordem são dadas por

$$\kappa_{ab}^{(cd)} = -\frac{1}{2} \sum_{\ell=1}^n \left\{ \frac{\partial^2}{\partial \delta^c \partial \delta^d} \left(\frac{q_{\ell ab}}{q_{\ell}} \right) \right\},$$

$$\kappa_{\gamma\gamma}^{(\gamma\gamma)} = -\frac{3n}{\gamma^4},$$

$$\kappa_{ij}^{(ka)} = -\frac{1}{\gamma} \sum_{\ell=1}^n x_{\ell_{ik}}^{**} q_{\ell a} w_{\ell} x_{\ell_j}^* + x_{\ell_i}^* q_{\ell a} \frac{\partial w_{\ell}}{\partial \eta_{\ell}} x_{\ell_j}^* x_{\ell_k}^* + x_{\ell_i}^* q_{\ell a} w_{\ell} x_{\ell_{jk}}^{**},$$

$$\kappa_{ij}^{(k\gamma)} = \frac{1}{\gamma^2} \sum_{\ell=1}^n x_{\ell_{ik}}^{**} q_{\ell} w_{\ell} x_{\ell_j}^* + x_{\ell_i}^* q_{\ell} \frac{\partial w_{\ell}}{\partial \eta_{\ell}} x_{\ell_j}^* x_{\ell_k}^* + x_{\ell_i}^* q_{\ell} w_{\ell} x_{\ell_{jk}}^{**},$$

$$\kappa_{ij}^{(ab)} = -\frac{1}{\gamma} \sum_{\ell=1}^n x_{\ell_i}^* q_{\ell ab} w_{\ell} x_{\ell_j}^*,$$

$$\kappa_{ij}^{(\gamma\gamma)} = -\frac{2}{\gamma^3} \sum_{\ell=1}^n x_{\ell_i}^* q_{\ell} w_{\ell} x_{\ell_j}^*,$$

$$\kappa_{ij}^{(a\gamma)} = \frac{1}{\gamma^2} \sum_{\ell=1}^n x_{\ell_i}^* q_{\ell a} w_{\ell} x_{\ell_j}^*,$$

e

$$\begin{aligned}
 \kappa_{ab}^{(c\gamma)} &= \kappa_{ab}^{(\gamma\gamma)} = \kappa_{ab}^{(i\gamma)} = \kappa_{ab}^{(ij)} = \kappa_{ab}^{(ic)} = \kappa_{\gamma\gamma}^{(ij)} = \kappa_{\gamma\gamma}^{(ij)} = \kappa_{\gamma\gamma}^{(i\gamma)} = \kappa_{\gamma\gamma}^{(ia)} = \kappa_{\gamma\gamma}^{(\gamma a)} = \\
 &= \kappa_{\gamma\gamma}^{(ab)} = \kappa_{ia}^{(jk)} = \kappa_{ia}^{(j\gamma)} = \kappa_{ia}^{(jb)} = \kappa_{ia}^{(b\gamma)} = \kappa_{ia}^{(bc)} = \kappa_{ia}^{(\gamma\gamma)} = \kappa_{i\gamma}^{(jk)} = \kappa_{i\gamma}^{(j\gamma)} = \\
 &= \kappa_{i\gamma}^{(ja)} = \kappa_{i\gamma}^{(a\gamma)} = \kappa_{i\gamma}^{(ab)} = \kappa_{i\gamma}^{(\gamma\gamma)} = \kappa_{a\gamma}^{(ij)} = \kappa_{a\gamma}^{(i\gamma)} = \kappa_{a\gamma}^{(ib)} = \kappa_{a\gamma}^{(b\gamma)} = \kappa_{a\gamma}^{(bc)} = \\
 &= \kappa_{a\gamma}^{(\gamma\gamma)} = 0.
 \end{aligned}$$

Os cumulantes de terceira ordem são dados por

$$\kappa_{ijk} = -\frac{1}{\gamma} \sum_{\ell=1}^n \{q_{\ell}(f_{\ell} + 2g_{\ell})x_{\ell_i}^* x_{\ell_j}^* x_{\ell_k}^* + q_{\ell}w_{\ell}(x_{\ell_i}^* x_{\ell_{jk}}^{**} + x_{\ell_j}^* x_{\ell_{ik}}^{**} + x_{\ell_k}^* x_{\ell_{ij}}^{**})\},$$

$$\kappa_{abc} = -\frac{1}{2} \sum_{\ell=1}^n \frac{q_{\ell abc}}{q_{\ell}},$$

$$\kappa_{\gamma\gamma\gamma} = \frac{2n}{\gamma^3},$$

$$\kappa_{ija} = -\frac{1}{\gamma} \sum_{\ell=1}^n x_{\ell_i}^* q_{\ell a} w_{\ell} x_{\ell_j}^*,$$

$$\kappa_{ij\gamma} = \frac{1}{\gamma^2} \sum_{\ell=1}^n x_{\ell_i}^* q_{\ell} w_{\ell} x_{\ell_j}^*,$$

$$\kappa_{ab\gamma} = \frac{1}{2\gamma} \sum_{\ell=1}^n \frac{q_{\ell ab}}{q_{\ell}},$$

$$\kappa_{a\gamma\gamma} = \kappa_{i\gamma\gamma} = \kappa_{abi} = \kappa_{ai\gamma} = 0,$$

em que $q_{\ell abcd} = \partial^3 q_{\ell} / \partial \delta^a \partial \delta^b \partial \delta^c$, para $a, b, c = 1, \dots, q$,

$$f_{\ell} = \frac{d^2 \mu_{\ell}}{d\eta_{\ell}^2} \frac{d\theta_{\ell}}{d\mu_{\ell}} \quad \text{e} \quad g_{\ell} = \frac{d^2 \theta_{\ell}}{d\mu_{\ell}^2} \left(\frac{d\mu_{\ell}}{d\eta_{\ell}} \right)^2.$$

Assim, as derivadas de primeira ordem dos cumulantes de terceira ordem são dados por

$$\kappa_{abc}^{(d)} = -\frac{1}{2} \sum_{\ell=1}^n \frac{\partial}{\partial \delta^d} \left(\frac{q_{\ell abc}}{q_{\ell}} \right),$$

$$\kappa_{\gamma\gamma\gamma}^{(\gamma)} = -\frac{6n}{\gamma^4},$$

$$\kappa_{ijk}^{(a)} = -\frac{1}{\gamma} \sum_{\ell=1}^n \{q_{\ell a}(f_{\ell} + 2g_{\ell})x_{\ell_i}^* x_{\ell_j}^* x_{\ell_k}^* + q_{\ell a}w_{\ell}(x_{\ell_i}^* x_{\ell_{jk}}^{**} + x_{\ell_j}^* x_{\ell_{ik}}^{**} + x_{\ell_k}^* x_{\ell_{ij}}^{**})\},$$

$$\kappa_{ijk}^{(\gamma)} = \frac{1}{\gamma^2} \sum_{\ell=1}^n \{q_{\ell}(f_{\ell} + 2g_{\ell})x_{\ell_i}^* x_{\ell_j}^* x_{\ell_k}^* + q_{\ell}w_{\ell}(x_{\ell_i}^* x_{\ell_{jk}}^{**} + x_{\ell_j}^* x_{\ell_{ik}}^{**} + x_{\ell_k}^* x_{\ell_{ij}}^{**})\},$$

$$\kappa_{ija}^{(k)} = -\frac{1}{\gamma} \sum_{\ell=1}^n \{x_{\ell_j}^* q_{\ell a} w_{\ell} x_{\ell_{ik}}^{**} + x_{\ell_i}^* q_{\ell a} \frac{\partial w_{\ell}}{\partial \eta_{\ell}} x_{\ell_j}^* x_{\ell_k}^* + x_{\ell_i}^* q_{\ell a} w_{\ell} x_{\ell_{jk}}^{**}\},$$

$$\kappa_{ija}^{(b)} = -\frac{1}{\gamma} \sum_{\ell=1}^n x_{\ell_i}^* q_{\ell ab} w_{\ell} x_{\ell_j}^*,$$

$$\kappa_{ija}^{(\gamma)} = \frac{1}{\gamma^2} \sum_{\ell=1}^n x_{\ell_i}^* q_{\ell a} w_{\ell} x_{\ell_j}^*,$$

$$\kappa_{ij\gamma}^{(k)} = \frac{1}{\gamma^2} \sum_{\ell=1}^n \left\{ x_{\ell_j}^* q_{\ell} w_{\ell} x_{\ell_{ik}}^{**} + x_{\ell_i}^* q_{\ell} \frac{\partial w_{\ell}}{\partial \eta_{\ell}} x_{\ell_j}^* x_{\ell_k}^* + x_{\ell_i}^* q_{\ell} w_{\ell} x_{\ell_{jk}}^{**} \right\},$$

$$\kappa_{ij\gamma}^{(a)} = \frac{1}{\gamma^2} \sum_{\ell=1}^n x_{\ell_i}^* q_{\ell a} w_{\ell} x_{\ell_j}^*,$$

$$\kappa_{ij\gamma}^{(\gamma)} = -\frac{2}{\gamma^3} \sum_{\ell=1}^n x_{\ell_i}^* q_{\ell} w_{\ell} x_{\ell_j}^*,$$

$$\kappa_{ab\gamma}^{(c)} = \frac{1}{2\gamma} \sum_{\ell=1}^n \frac{\partial}{\partial \delta^c} \left(\frac{q_{\ell ab}}{q_{\ell}} \right),$$

$$\kappa_{ab\gamma}^{(\gamma)} = -\frac{1}{2\gamma^2} \sum_{\ell=1}^n \frac{q_{\ell ab}}{q_{\ell}},$$

e

$$\begin{aligned} \kappa_{abc}^{(i)} &= \kappa_{abc}^{(\gamma)} = \kappa_{ab\gamma}^{(i)} = \kappa_{\gamma\gamma\gamma}^{(i)} = \kappa_{\gamma\gamma\gamma}^{(\delta)} = \kappa_{abi}^{(j)} = \kappa_{abi}^{(c)} = \kappa_{abi}^{(\gamma)} = \kappa_{a\gamma\gamma}^{(i)} = \kappa_{a\gamma\gamma}^{(b)} \\ &= \kappa_{a\gamma\gamma}^{(\gamma)} = \kappa_{i\gamma\gamma}^{(j)} = \kappa_{i\gamma\gamma}^{(a)} = \kappa_{i\gamma\gamma}^{(\gamma)} = \kappa_{ai\gamma}^{(j)} = \kappa_{ai\gamma}^{(b)} = \kappa_{ai\gamma}^{(\gamma)} = 0. \end{aligned}$$

Finalmente, os cumulantes de quarta ordem são dadas por

$$\begin{aligned} \kappa_{ijkl} &= -\frac{1}{\gamma} \sum_{\ell=1}^n q_{\ell} \left[\left(\frac{\partial^2 w_{\ell}}{\partial \eta_{\ell}^2} \right) + \left(\frac{\partial^2 \mu_{\ell}}{\partial \eta_{\ell}^2} \right) \left(\frac{\partial^2 \theta_{\ell}}{\partial \eta_{\ell}^2} \right) + 2 \left(\frac{\partial \mu_{\ell}}{\partial \eta_{\ell}} \right) \left(\frac{\partial^3 \theta_{\ell}}{\partial \eta_{\ell}^3} \right) \right] x_{\ell_i}^* x_{\ell_j}^* x_{\ell_k}^* x_{\ell_l}^* \\ &+ w_{\ell} (x_{\ell_{ijk}}^{***} x_{\ell_l}^* + x_{\ell_{ijl}}^{***} x_{\ell_k}^* + x_{\ell_{ikl}}^{***} x_{\ell_j}^* + x_{\ell_{jkl}}^{***} x_{\ell_i}^* + x_{\ell_{ij}}^{**} x_{\ell_{kl}}^{**} + x_{\ell_{ik}}^{**} x_{\ell_{jl}}^{**} + x_{\ell_{il}}^{**} x_{\ell_{jk}}^{**}) \\ &+ (x_{\ell_{ij}}^{**} x_{\ell_k}^* x_{\ell_l}^* + x_{\ell_{jk}}^{**} x_{\ell_i}^* x_{\ell_l}^* + x_{\ell_{jl}}^{**} x_{\ell_i}^* x_{\ell_k}^* + x_{\ell_{il}}^{**} x_{\ell_j}^* x_{\ell_k}^* + x_{\ell_{kl}}^{**} x_{\ell_i}^* x_{\ell_j}^* + x_{\ell_{ik}}^{**} x_{\ell_j}^* x_{\ell_l}^*) \\ &\times \left[2 \left(\frac{\partial \mu_{\ell}}{\partial \eta_{\ell}} \right) \left(\frac{\partial^2 \theta_{\ell}}{\partial \eta_{\ell}^2} \right) + \left(\frac{\partial \theta_{\ell}}{\partial \eta_{\ell}} \right) \left(\frac{\partial^2 \mu_{\ell}}{\partial \eta_{\ell}^2} \right) \right], \end{aligned}$$

$$\kappa_{abcd} = -\frac{1}{2} \sum_{\ell=1}^n \frac{q_{\ell abcd}}{q_{\ell}},$$

$$\kappa_{\gamma\gamma\gamma\gamma} = -\frac{9n}{\gamma^4},$$

$$\kappa_{ijk\alpha} = -\frac{1}{\gamma} \sum_{\ell=1}^n \{q_{\ell\alpha}(f_{\ell} + 2g_{\ell})x_{\ell_i}^*x_{\ell_j}^*x_{\ell_k}^* + q_{\ell\alpha}w_{\ell}(x_{\ell_i}^*x_{\ell_jk}^{**} + x_{\ell_j}^*x_{\ell_{ik}}^{**} + x_{\ell_k}^*x_{\ell_{ij}}^{**})\},$$

$$\kappa_{ijk\gamma} = \frac{1}{\gamma^2} \sum_{\ell=1}^n \{q_{\ell}(f_{\ell} + 2g_{\ell})x_{\ell_i}^*x_{\ell_j}^*x_{\ell_k}^* + q_{\ell}w_{\ell}(x_{\ell_i}^*x_{\ell_jk}^{**} + x_{\ell_j}^*x_{\ell_{ik}}^{**} + x_{\ell_k}^*x_{\ell_{ij}}^{**})\},$$

$$\kappa_{abc\gamma} = \frac{1}{2\gamma} \sum_{\ell=1}^n \frac{q_{\ell abc}}{q_{\ell}},$$

$$\kappa_{ijab} = -\frac{1}{\gamma} \sum_{\ell=1}^n x_{\ell_i}^*q_{\ell ab}w_{\ell}x_{\ell_j}^*,$$

$$\kappa_{ij\gamma\gamma} = -\frac{2}{\gamma^3} \sum_{\ell=1}^n x_{\ell_i}^*q_{\ell}w_{\ell}x_{\ell_j}^*,$$

$$\kappa_{ab\gamma\gamma} = -\frac{1}{\gamma^2} \sum_{\ell=1}^n \frac{q_{\ell ab}}{q_{\ell}},$$

$$\kappa_{ija\gamma} = \frac{1}{\gamma^2} \sum_{\ell=1}^n x_{\ell_i}^*q_{\ell a}w_{\ell}x_{\ell_j}^*,$$

$$\kappa_{i\gamma\gamma\gamma} = \kappa_{a\gamma\gamma\gamma} = \kappa_{abci} = \kappa_{abi\gamma} = 0,$$

em que $x_{\ell_{ijk}}^{***} = \partial^3 \eta_{\ell} / \partial \beta^i \partial \beta^j \partial \beta^k$, para $i, j, k = 1, \dots, p$ e $q_{\ell abcd} = \partial^4 q_{\ell} / \partial \delta^a \partial \delta^b \partial \delta^c \partial \delta^d$, para $a, b, c, d = 1, \dots, q$.

Apêndice B

Obtenção das quantidades A_1 , A_2 e A_3 para a correção tipo-Bartlett em modelos não lineares da família exponencial com dispersão variável

As expressões algébricas de A_1 , A_2 e A_3 são calculadas a partir dos cumulantes conjuntos e das derivadas dos cumulantes conjuntos apresentados no Apêndice A. Mostramos aqui, a obtenção da quantidade A_2 na forma matricial e omitimos as demais, pois os cálculos são realizados de forma análoga. Para facilitar os cálculos expandimos a expressão algébrica de A_2 apresentada em 3.3 e escrevemos da seguinte forma:

$$A_2 = A_{21} + A_{22} + A_{23} + A_{24} + A_{25} + A_{26} + A_{27},$$

em que

$$A_{21} = -3 \sum' \kappa_{jrs} \kappa_{klu} m^{jr} a^{lu} m^{sk},$$

$$A_{22} = -3 \sum' \kappa_{jrs} \kappa_{klu} m^{jr} a^{sk} m^{lu},$$

$$A_{23} = -3 \sum' 2 \kappa_{jrs} \kappa_{klu} m^{jk} a^{su} m^{rl},$$

$$A_{24} = -3 \sum' \frac{3}{4} \kappa_{jrs} \kappa_{klu} m^{jr} m^{sk} m^{lu},$$

$$A_{25} = -3 \sum' \frac{2}{4} \kappa_{jrs} \kappa_{klu} m^{jk} m^{rl} m^{su},$$

$$A_{26} = 6 \sum' \kappa_{jrs}^{(u)} m^{jr} m^{su},$$

e

$$A_{27} = -3 \sum' \kappa_{jrsu} m^{jr} m^{su},$$

sendo a^{jr} e m^{jr} os elementos (j, r) das matrizes definidas em 3.6.

Considerando a heteroscedastidade multiplicativa, isto é, $m_\ell = \exp\{z_\ell^\top \boldsymbol{\delta}\}$, pode-se mostrar facilmente que:

$$\kappa_{ab}^{(c)} = 0, \quad \kappa_{ab}^{(cd)} = 0, \quad \kappa_{ij}^{(k)} = 0 \quad \text{e} \quad \kappa_{ij}^{(kl)} = 0.$$

Explorando este fato e a ortogonalidade entre os parâmetros $\boldsymbol{\delta}$, $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$, obtemos

$$A_{21} = -3 \sum' \kappa_{abc} \kappa_{dij} m^{ab} a^{ij} m^{cd},$$

$$A_{22} = -3 \sum' \kappa_{ab\gamma} \kappa_{\gamma cd} m^{ab} a^{\gamma\gamma} m^{cd},$$

$$A_{23} = -3 \sum' 2 \kappa_{cd\gamma} \kappa_{cd\gamma} m^{ac} a^{\gamma\gamma} m^{bd},$$

$$A_{24} = -3 \sum' \frac{3}{4} \kappa_{abc} \kappa_{def} m^{ab} m^{cd} m^{ef},$$

$$A_{25} = -3 \sum' \frac{2}{4} \kappa_{abc} \kappa_{def} m^{ad} m^{be} m^{cf},$$

$$A_{26} = 0,$$

e

$$A_{27} = -3 \sum' \kappa_{abcd} m^{ab} m^{cd}.$$

(B.1)

Substituindo os cumulantes conjuntos e as derivadas dos cumulantes conjuntos obtidos no Apêndice A nas expressões dadas em (B.1) e rearranjando os termos, obtemos

$$A_{21} = -\frac{3}{2} \sum_{r,s=1}^n \sum' (z_r - \bar{z})_a m^{ab} (z_r - \bar{z})_b (z_r - \bar{z})_c m^{cd} (z_s - \bar{z})_d \frac{q_s}{\gamma} w_s x_{s_i}^* a^{ij} x_{s_j}^*,$$

$$A_{22} = -\frac{3}{4} \sum_{r,s=1}^n \sum' \frac{2}{n} (z_r - \bar{z})_a m^{ab} (z_r - \bar{z})_b (z_s - \bar{z})_c m^{cd} (z_s - \bar{z})_d,$$

$$A_{23} = -3 \sum_{r,s=1}^n \sum' \frac{1}{n} (z_r - \bar{z})_a m^{ac} (z_s - \bar{z})_c (z_r - \bar{z})_b m^{bd} (z_s - \bar{z})_d,$$

$$A_{24} = -\frac{9}{4} \sum_{r,s=1}^n \sum' \frac{1}{4} (z_r - \bar{z})_a m^{ab} (z_r - \bar{z})_b (z_r - \bar{z})_c m^{cd} (z_s - \bar{z})_d (z_s - \bar{z})_e m^{ef} (z_s - \bar{z})_f,$$

$$A_{25} = -3 \sum_{r,s=1}^n \sum' \frac{1}{8} (z_r - \bar{z})_a m^{ad} (z_r - \bar{z})_d (z_r - \bar{z})_b m^{be} (z_s - \bar{z})_e (z_r - \bar{z})_c m^{cf} (z_s - \bar{z})_f,$$

e

$$A_{27} = -3 \sum_{r=1}^n \sum' -\frac{1}{2} (z_r - \bar{z})_a m^{ab} (z_r - \bar{z})_b (z_r - \bar{z})_c m^{cd} (z_r - \bar{z})_d.$$

Os termos $\sum x_{r_i}^* a^{ij} x_{r_j}^*$ e $\sum (z_r - \bar{z})_a m^{ab} (z_r - \bar{z})_b$ são elementos das matrizes Z_β e Z_δ , respectivamente, definidas na Seção 3.6. Assim, A_2 se reduz a:

$$\begin{aligned} A_2 = & -\frac{3}{2} \sum_{r,s=1}^n Z_{\delta_{rr}} Z_{\delta_{rs}} \phi_s w_s Z_{\beta_{ss}} - \frac{6}{4} \sum_{r,s=1}^n Z_{\delta_{rr}} Z_{\delta_{ss}} - \frac{3}{n} \sum_{r,s=1}^n Z_{\delta_{rs}} Z_{\delta_{rs}} \\ & - \frac{9}{16} \sum_{r,s=1}^n Z_{\delta_{rr}} Z_{\delta_{rs}} Z_{\delta_{ss}} - \frac{3}{8} \sum_{r,s=1}^n Z_{\delta_{rs}} Z_{\delta_{rs}} Z_{\delta_{rs}} + \frac{3}{2} \sum_{r=1}^n Z_{\delta_{rr}} Z_{\delta_{rr}} \end{aligned}$$

e pode ser escrita em notação matricial conforme apresentamos em (3.3).

Apêndice C

Multiplicação de “array”

Neste apêndice, apresentamos a definição de um array tri-dimensional e uma propriedade de multiplicação de array que nos auxiliará no cálculo de j^* , o bloco da matriz de informação observada para os parâmetros de perturbação $(\beta^\top, \gamma)^\top$ no modelo não-linear heteroscedástico reparametrizado avaliado em $(\delta^\top, \hat{\beta}_{\delta^\top}, \hat{\gamma}_{\delta})^\top$. Um “array” tri-dimensional de dimensão $n \times p \times q$ é denotado por $X = (X_{\ell ij})$, onde os índices ℓ, i e j indicam a face, a linha e a coluna, respectivamente. Um “array” pode ser visto da seguinte forma: $A = (A_\ell)$ e cada A_ℓ é uma matriz $A_\ell = (A_{\ell ij})$ de dimensão $p \times q$ para algum ℓ fixo e A_ℓ é chamada de ℓ -ésima face de A .

A multiplicação de “array” tri-dimensional foi introduzida por Bates e Watts (1980). Desde então, muitos autores têm discutido e utilizado esta multiplicação; ver Bates e Watts (1988), Seber e Wild (1989) e Wei (1998).

Definição 1. Se X é um “array” de dimensão $n \times p \times q$, A e B são matrizes de dimensões $r \times p$ e $q \times s$, respectivamente, então $Y = AXB$ é definido como um “array” de dimensão $n \times r \times s$ com elementos

$$Y_{\ell kt} = \sum_{i=1}^p \sum_{j=1}^q A_{ki} X_{\ell ij} B_{jt}.$$

Definição 2. Se X é um “array” de dimensão $n \times p \times q$, A é uma matriz de dimensão $m \times n$, então $Y = [A][X]$ é chamado de produto colchete de A e X , isto é, um “array” de dimensão $m \times p \times q$ com elemento

$$Y_{\ell ij} = \sum_{t=1}^n A_{\ell t} X_{tij}.$$

Definição 3. (Decomposição de Cholesky) Se A é uma matriz positiva definida, existe uma única matriz triangular superior, U , com elementos diagonais positivos tal que

$$A = U^\top U.$$

A seguinte propriedade pode ser deduzida diretamente das definições acima.

Propriedade 1. *Sejam A, L, M matrizes e X um “array”, então, temos que*

$$[A][LXM] = L[A][X]M.$$

Apêndice D

Coelhos Europeus na Austrália

Tabela D.1: Pesos das lentes dos olhos de coelhos europeus (y), em miligramas e idade (x), em dias, numa amostra de 71 observações.

x	y	x	y	x	y
15	21,66	98	104,30	285	189,66
15	22,75	125	134,90	300	186,09
15	22,30	142	130,68	301	186,70
18	31,25	142	140,58	305	186,80
28	44,79	147	155,30	312	195,10
29	40,55	147	152,20	317	216,41
37	50,25	150	144,50	338	203,23
37	46,88	159	142,15	347	188,38
44	52,03	165	139,81	354	189,70
50	63,47	183	153,22	357	195,31
50	61,13	192	145,72	375	202,63
60	81,00	195	161,10	394	224,82
61	73,09	218	174,18	513	203,30
64	79,09	218	173,03	535	209,70
65	79,51	219	173,54	554	233,90
65	65,31	224	178,86	591	234,70
72	71,90	225	177,68	648	244,30
75	86,10	227	173,73	660	231,00
75	94,60	232	159,98	705	242,40
82	92,50	232	161,29	723	230,77
85	105,00	237	187,07	756	242,57
91	101,70	246	176,13	768	232,12
91	102,90	258	183,40	860	246,70
97	110,00	276	186,26		

Apêndice E

Solução de Cálcio Radioativo

Tabela E.1: Quantidade absorvida de cálcio radioativo (y), em nmoles/mg, e tempo (x), em minutos, numa amostra de 27 observações.

x	y	x	y
0,45	0,34170	6,10	2,67061
0,45	-0,00438	8,05	3,05959
0,45	0,82531	8,05	3,94321
1,30	1,77967	8,05	3,43726
1,30	0,95384	11,15	4,80735
1,30	0,64080	11,15	3,35583
2,40	1,75136	11,15	2,78309
2,40	1,27497	13,15	5,13825
2,40	1,17332	13,15	4,70274
4,00	3,12273	13,15	4,25702
4,00	2,60958	15,00	3,60407
4,00	2,57429	15,00	4,15029
6,10	3,17881	15,00	3,42484
6,10	3,00782		