**Estimação e Testes em Modelos
Lineares Generalizados com
Restrições nos Parâmetros na
Forma de Desigualdades Lineares**

Francisco José de Azevêdo Cysneiros

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

Área de Concentração: **Estatística**

Orientador: **Prof. Dr. Gilberto Alvarenga Paula**

São Paulo - junho - 1997

**Estimação e Testes em Modelos
Lineares Generalizados com
Restrições nos Parâmetros na
Forma de Desigualdades Lineares**

FRANCISCO JOSÉ DE AZEVÊDO CYSNEIROS

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por **Francisco José de Azevêdo Cysneiros** e aprovada pela comissão julgadora.

Aprovado em : 20 de junho de 1997

Comissão julgadora:

- Prof. Dr. Gilberto Alvarenga Paula (Orientador)
- Prof. Dr. José Galvão Leite
- Prof. Dr^a Cícília Wada

IME/USP

IME/USP

IMECC/UNICAMP

À minha mãe e irmãos,
com gratidão,

Ao meu pai Gilberto (in memoriam),
com saudade,

À minha esposa
Audrey
com eterna paixão,

Ao meu filho
Rafael
com admiração,

dedico com carinho e amor.

Agradecimentos

- Ao Professor **Gilberto** pela confiança e excelente orientação dedicada na elaboração deste trabalho.
- À minha esposa, pelo grande apoio a mim concedido, em especial, ao meu filho, **Rafael**, pela compreensão e carinho por ele oferecido.
- Aos meu Pais, **Gilberto** e **Gilvanete**, que me forneceram princípios básicos e fundamentais para minha formação moral e dedicação integral ao meu objetivo.
- Aos professores do Instituto de Matemática e Estatística que ajudaram na minha formação acadêmica.
- Ao Professor **Dr. José Galvão Leite** pela sua contribuição nos resultados do Capítulo 2 desta dissertação.
- Aos meus amigos que me apoiaram e ajudaram permitindo que este passo da minha vida fosse dado.
- Aos colegas do Depto. de Matemática e Estatística da Universidade Federal da Paraíba (campus II) pelo apoio dado durante este curso, em especial, aos amigos e professores da Área de Estatística.
- À **Alba, Chico, Rosana, Cardoso, Gil, Lili, Cláudia Lima e Manoel Senna** no qual sempre me incentivaram nesta caminhada .
- À CAPES pelo apoio financeiro.

Resumo

O objetivo deste trabalho é apresentar de maneira formal, numa primeira etapa, a distribuição nula bem como a equivalência assintótica de alguns testes estatísticos, tais como razão de verossimilhança, Wald e escore, para dois casos gerais de hipóteses restritas na forma de desigualdades lineares. Numa segunda etapa, discutimos a aplicação da teoria em modelos lineares generalizados e apresentamos alguns casos particulares em que simplificações interessantes são obtidas. Algoritmos para a obtenção das estimativas restritas dos parâmetros bem como cinco exemplos ilustrativos são apresentados. Um programa original em S-Plus para a obtenção das estimativas restritas em modelos lineares generalizados é desenvolvido e apresentado num dos apêndices.

Abstract

The aim of the work is to present, in the first part, a formal demonstration of the asymptotic null distribution as well as the asymptotic equivalence among several statistical tests, such as likelihood ratio, Wald and score, for testing hypotheses of linear inequality parameter constraints. Further, we discuss the application of the methodology in generalized linear models. Some particular cases with relevant simplifications are discussed. Algorithms for obtaining the restricted estimates and five illustrative examples are given. An original program in S-Plus is developed for obtaining the restricted estimates in generalized linear models.

Conteúdo

Lista de Figuras	viii
Lista de Tabelas	ix
1 Introdução	1
2 Testes de Hipóteses com Restrições Lineares	6
2.1 Introdução	6
2.2 Conceitos e definições	6
2.3 Algumas propriedades assintóticas	8
2.4 Teste de hipóteses $H_0 : C\theta = 0 \times H_1 : \{C\theta \geq 0\} - H_0$	9
2.5 Equivalência assintótica dos problemas de otimização	16
2.6 Distribuição nula assintótica da estatística ξ^D	18
2.7 Teste de hipóteses $H_0 : C\theta \geq 0 \times H_1 : \mathbb{R}^p - H_0$	29
3 Modelos Lineares Generalizados	36
3.1 Introdução	36
3.2 Hipótese em igualdades lineares	45
3.3 Estimação sob desigualdades lineares	48
3.4 Teste de hipótese em desigualdades lineares	52
3.5 Aplicações	54
4 Alguns Casos Particulares	64
4.1 Introdução	64
4.2 Modelo de análise de variância	64

CONTEÚDO	vii
4.3 Modelo de regressão linear	65
4.4 Ordem simples	66
4.5 Retas paralelas	68
4.6 Aplicações	70
Conclusões	81
A Probabilidades de Nível	82
A.1 Caso de $k = 3$ restrições	82
A.2 Caso de $k = 4$ restrições	82
B Processo Infeccioso Pulmonar	84
C Vírus da Poliomyelitis	87
D Estudo da Relação de Abortos com Casamentos Consangüíneos	89
E Estudo de Câncer Respiratório em Metalúrgicos	90
E.1 Valores observados	90
E.2 Valores esperados	91
F Estudo de Toxicidade de Inseticidas em Insetos da Farinha	92
G Implementação do Algoritmo no S-Plus	93
H PAVA (<i>pool adjacent violator algorithm</i>)	99
Referências	100

Lista de Figuras

2.1	Construções dos cones	21
2.2	Projeções no Cone C_ϕ	22
2.3	Projeções de $\hat{\theta}$	27
2.4	Região crítica do teste de Haussman-Wald	28
2.5	Probabilidades de rejeição sob a hipótese nula	35
4.1	Gráfico de retas separadas	74
4.2	Retas paralelas da proporção de insetos mortos segundo a dose	78

Lista de Tabelas

3.1	Características de algumas distribuições da família exponencial	38
3.2	Modelo logístico para a proporção de abortos naturais	55
3.3	Estatísticas do teste de $H_0 : C\beta = 0 \times H_1 : \{C\beta \geq 0\} - H_0$	56
3.4	Modelo de Poisson para o estudo dos metalúrgicos	58
3.5	Estatísticas do teste para o arsênico moderado	59
3.6	Estatísticas do teste para o arsênico pesado	59
3.7	Estimativas irrestrita e restrita dos parâmetros para o modelo (3.15)	62
3.8	Estatísticas do teste de para a célula HI	63
3.9	Estatísticas do teste para a célula Ff	63
4.1	Estimativas irrestrita e restrita dos parâmetros do modelo (4.6)	73
4.2	Estatística do teste para os dados da Poliomyelitis	74
4.3	Estatística do teste para os dados da Poliomyelitis	75
4.4	Estatísticas do teste de $H_0 : C\beta = 0 \times H_1 : \{C\beta \geq 0\} - H_0$	77
4.5	Estatísticas do teste para o modelo de retas paralelas	79
B.1	Estudo de câncer no processo infeccioso pulmonar	84
C.1	Dados correspondentes ao vírus da Poliomyelitis	87
D.1	Dados referentes ao número de abortos em 6358 casos em Shizuoka City no Japão	89
E.1	Número observado de mortes para metalúrgicos expostos ao arsênico	90
E.2	Número esperado de mortes para metalúrgicos expostos ao arsênico	91
F.1	Toxicidade de inseticidas nos besouros da farinha	92

CAPÍTULO 1

Introdução

Nos anos de 1959 e 1961 deram-se as primeiras publicações sobre testes para hipóteses na forma de desigualdades. Esses trabalhos devidos a Bartholomew, apresentaram resultados importantes para o caso de ordem simples ($\mu_1 \leq \dots \leq \mu_k$) em k populações normais independentes. Entretanto, foram Kudô (1963) e Nüesch (1964,1966) que mostraram que a distribuição nula do teste da razão de verossimilhança para testar hipóteses de igualdades lineares contra desigualdades lineares para os coeficientes de um modelo normal linear é uma mistura de distribuições do tipo qui-quadrado, ponderadas por pesos, diferindo do caso usual, isto é, o caso em que não há restrições sobre os parâmetros do modelo. Quando a variância das observações é desconhecida, a distribuição nula da estatística da razão de verossimilhança é uma mistura de distribuições do tipo beta ou F . Os pesos para o cálculo dessas distribuições no caso de k populações, sob a hipótese nula, não dependem dos parâmetros. Perlman (1969) mostrou que a distribuição nula do teste da razão de verossimilhança para testar hipóteses de desigualdades lineares, em que a hipótese nula é composta por desigualdades é também uma mistura de qui-quadrados e propôs um lema no qual define a situação menos favorável quando os pesos não dependem dos parâmetros. Porém, à medida que o número de restrições aumenta, esses pesos tomam formas complexas. Existem formas fechadas para os pesos até quatro restrições. Bohrer e Chow (1978) escreveram um programa computacional para calcular os pesos até 10 restrições. Esse programa faz o uso de integração numérica utilizando o enfoque dado em Childs (1967). No caso de termos uma estrutura de regressão, os pesos geralmente dependem dos parâmetros do modelo sob a hipótese nula. É importante notar que como os pesos dependem dos parâmetros, a distribuição nula não é mais única como ocorre no caso de k

populações. Em Shapiro (1985) encontra-se um resumo sobre a obtenção desses pesos. Todas as dificuldades em calcular os pesos têm motivado o desenvolvimento de várias linhas de pesquisa nesta área.

Naturalmente, os resultados de k populações foram estendidos para a família exponencial e também para outras distribuições tais como a multinomial. Nesses casos, sob condições gerais de regularidade, o teste da razão de verossimilhança tem distribuição nula assintótica que é uma mistura de qui-quadrados ponderadas com pesos similares aos do caso normal. Lee et al. (1993) fazem uma revisão das diversas aproximações desenvolvidas para os pesos e apresentam limites mais precisos que podem ser usados quando os pesos não tomam formas fechadas. Barlow et al. (1972) e Robertson et al. (1988) lançaram livros sobre este assunto, os quais são excelentes referências para leitura.

O estudo do poder para os testes de hipóteses com restrições é feito através de métodos de Monte Carlo ou de aproximações em torno da hipótese nula, uma vez que os pesos assumem expressões bem mais complexas na hipótese alternativa. Somente em alguns casos particulares essas funções têm forma fechada. Sabe-se que os testes restritos são em geral mais poderosos do que outros testes competitivos (Hillier, 1986).

Nessa última década o estudo de testes para hipóteses com restrições, com enfoque em regressão, teve um grande número de artigos publicados. Kodde e Palm (1986) estenderam os resultados de Kudô (1963) para o caso de hipóteses não-lineares e propuseram o uso de um teste do tipo Wald que, sob certas condições de regularidade, pode ser aplicado para situações em que a função de verossimilhança é desconhecida. Gouriéroux e Monford (1995) mostram a equivalência assintótica da estatística do tipo Wald com algumas estatísticas usuais. Wolak (1987, 1989a) trata o problema de teste de hipóteses com restrição para o modelo de regressão linear com estruturas gerais para a matriz de variância-covariância dos erros. Wolak (1987) propõe uma estatística da razão de verossimilhança modificada quando a matriz de correlação é conhecida, porém as variâncias são desconhecidas, e mostra que a distribuição nula é uma mistura de distribuições do tipo F. Wolak (1989a)

estende os resultados de Gouriéroux et al. (1982) para o caso da matriz de variância-covariância ser desconhecida e depende de um número finito de parâmetros.

Piegorsch (1990) apresenta aplicações desta teoria em modelos lineares generalizados com resposta binária e alguns estudos de simulação comparando o poder do teste da razão de verossimilhança com restrições nos parâmetros com o poder de alguns métodos de comparações múltiplas. Silvapulle (1991,1994) estuda a aplicação desse tipo de teste em modelos de regressão com função de verossimilhança côncava, que englobam os modelos lineares generalizados com ligação canônica e os modelos de regressão de Cox. Wolak (1991) demonstra para uma classe ampla de modelos de regressão, que a distribuição nula menos favorável não é necessariamente atingida quando todas as restrições são satisfeitas na forma de igualdades. Nesse artigo, ele propõe um lema, estendendo o resultado encontrado por Perlman, no qual define um subconjunto da hipótese nula que contém a situação menos favorável. Para determinar a situação menos favorável deve-se percorrer todo esse subconjunto no qual computacionalmente pode ser muito dispendioso. Wolak (1989b) sugere o uso de testes locais que sob condições adicionais de regularidade possam levar à solução do problema. Farhmeir e Klinger (1994) tratam da estimação e teste em modelos lineares generalizados para hipóteses de restrição em desigualdades lineares e apresentam estudos numéricos em que propõem um método para a determinação da situação menos favorável definida no lema de Wolak. Paula e Sen (1995) verificam que os pesos envolvidos na distribuição nula assintótica do teste da razão de verossimilhança, para algumas subclasses de modelos lineares generalizados e para algumas hipóteses com restrição de ordem, não dependem dos parâmetros sob a hipótese nula quando estruturas particulares são assumidas para a matriz modelo. Paula e Rojas (1997) aplicam esta teoria em modelos de regressão com distribuição do valor extremo com parâmetro de dispersão desconhecido.

Na prática, podemos encontrar vários estudos em que é assumido algum tipo de informação a priori, isto é, algum tipo de restrição nos parâmetros do modelo. Existe vários artigos na área de Estatística, Econometria e Farmacologia onde

encontramos este tipo de abordagem. Para ilustrar, apresentamos a seguir dois exemplos que serão discutidos mais detalhadamente no texto.

McDonald e Diamond (1983) propuseram modelos logísticos para explicar a chance de abortos naturais entre pais com algum grau de consangüinidade. Usando conhecimentos de genética humana, eles levantaram a hipótese de que essa chance tende a aumentar com o grau de consangüinidade entre os pais. Em particular, para um conjunto de gestantes em três distritos da Cidade de Shizuoka no Japão, descrito no Apêndice D, um modelo logístico restrito foi proposto para explicar a proporção de abortos naturais. As seguintes variáveis explicativas dicotômicas (sim =1, não =0) foram utilizadas :

- (i) $C2 \equiv$ primos de 2º grau;
- (ii) $C12 \equiv$ primos de $1\frac{1}{2}$ º grau;
- (iii) $C1 \equiv$ primos de 1º grau;
- (iv) $INT \equiv$ vive no distrito intermediário e
- (v) $URB \equiv$ vive no distrito urbano.

O modelo adotado é

$$\log\{\pi/(1-\pi)\} = \beta_1 + \beta_2 C2 + \beta_3 C12 + \beta_4 C1 + \beta_5 INT + \beta_6 URB$$

sujeito a $\beta_4 \geq \beta_3 \geq \beta_2 \geq 0$, onde π é a proporção de abortos naturais.

Será que a proporção de abortos naturais permanece constante em cada distrito, isto é, $\beta_4 = \beta_3 = \beta_2 = 0$? Na Subseção 3.5.1 encontramos tanto a estimação como a solução deste problema de teste de hipóteses com restrições nos parâmetros na forma de desigualdades lineares.

Como segunda ilustração, apresentamos um experimento (Finney, 1978) envolvendo uma preparação padrão do vírus da *poliomyelitis* e quatro preparações teste. Podemos neste caso estar interessados em verificar se a preparação padrão não é menos eficiente do que as preparações teste. Em outras palavras, se não há nenhuma preparação mais potente do que a padrão no sentido de causar a doença no animal num menor tempo médio possível. As preparações foram combinadas em doses os

quais foram inoculadas em cinco ratos machos e cinco fêmeas. Foi observado como resposta, o número de dias decorridos até o aparecimento de algum sintoma da doença, denotado por Y . Esse tipo de experimento induz o seguinte modelo (Paula, 1997)

$$\log \mu_{ijk\ell} = \alpha_i + \delta_i x_{ij},$$

onde x_{ij} denota a dose no nível (i, j) e será assumido que Y segue uma distribuição gama de média μ e parâmetro de dispersão ϕ^{-1} . Como será discutido na Subseção 4.6.1, a comparação da preparação padrão com as demais é equivalente a testarmos $\alpha_1 \leq [\alpha_2, \dots, \alpha_5]$ e $\delta_1 = \dots = \delta_5$. Dois outros exemplos práticos e dois teóricos com hipóteses restritas em desigualdades são apresentados no texto.

CAPÍTULO 2

Testes de Hipóteses com Restrições Lineares

2.1 Introdução

Neste capítulo, discutiremos a distribuição nula assintótica de algumas estatísticas para o problema de testar hipóteses do tipo $H_0 : C\theta = 0 \times H_1 : \{C\theta \geq 0\} - H_0$ e do tipo $H_0 : C\theta \geq 0 \times H_1 : \mathbb{R}^p - H_0$ onde C é uma matriz $(k \times p)$ de posto completo e $\theta = (\theta_1, \dots, \theta_p)^t$ é um vetor de parâmetros p -dimensional. Definimos o vetor $\{C\theta \geq 0\}$, formado por todas as componentes $C_j^t \theta \geq 0$, onde C_j^t é a j -ésima linha da matriz C com $j = 1, \dots, k$. É importante salientar que a distribuição assintótica das estatísticas de teste sob H_0 deixa de ser uma distribuição χ^2 sendo agora uma mistura de χ^2 ponderadas por probabilidades. Trataremos em particular de modelos paramétricos cuja verossimilhança seja côncava. Dentre esses modelos podemos citar os modelos lineares generalizados (Nelder e Wedderburn, 1972, 1976), modelos de regressão de Cox (Cox, 1974) e modelos de regressão com distribuição log-gama generalizada (Lawless, 1980).

2.2 Conceitos e definições

Consideramos uma sequência de variáveis aleatórias $Y_i, i = 1, \dots, n$. Assumimos que as variáveis $Y_i, i = 1, \dots, n$, são independentemente distribuídas com densidade $f(y; \theta), \theta \in \Theta \subset \mathbb{R}^p$. O logaritmo da função de verossimilhança denotado por $L(\theta)$,

$$L(\theta) = L(y; \theta) = \sum_{i=1}^n \log f(y_i; \theta),$$

é assumido ser uma função contínua em θ . Consideramos como função objetivo, $L(\theta)$, que deve satisfazer as condições usuais de regularidade e mais as condições dadas abaixo :

$$(a) \quad \frac{1}{\sqrt{n}} \mathcal{U}(\theta_0) \xrightarrow{D} N_p(0, \mathcal{I}_0) :$$

$$(b) \quad -\frac{1}{n} \mathcal{K}(\theta_0) \xrightarrow{q.c.} \mathcal{J}_0;$$

$$(c) \quad \mathcal{U}(\hat{\theta}) + C^t \tilde{\lambda} = 0;$$

$$(d) \quad C_j^t \tilde{\theta} \geq 0, \quad \tilde{\lambda}_j \geq 0, \quad j = 1, \dots, k;$$

$$(e) \quad \tilde{\lambda}_j \{C_j^t \tilde{\theta}\} = 0, \quad j = 1, \dots, k,$$

onde $\mathcal{I}_0$ e $\mathcal{J}_0$ são matrizes definidas positivas, $\mathcal{U}(\theta) = \frac{\partial}{\partial \theta} L(\theta)$, $\mathcal{K}(\theta) = \frac{\partial^2}{\partial \theta \partial \theta^t} L(\theta)$, $\tilde{\theta}$ é o estimador de máxima verossimilhança obtido maximizando-se $L(\theta)$ sujeito a $C\theta \geq 0$, e $\tilde{\lambda} = (\tilde{\lambda}_1, \dots, \tilde{\lambda}_k)^t$ são os multiplicadores de Kühn-Tücker associados a $\tilde{\theta}$. Denotamos $\hat{\theta}$, o estimador de máxima verossimilhança resultante do problema de maximização de $L(\theta)$ sujeito a $\theta \in \Theta$, denominado estimador irrestrito de θ e θ_0 , como o verdadeiro valor do parâmetros θ . As condições (a) e (b) são usuais no caso de modelos irrestritos, porém as condições (c)-(e) são necessárias para as demonstrações que serão apresentadas nesta seção e são decorrentes das condições de Kühn-Tücker. A condição (d) é conhecida como condição do sinal e (e) como condição de exclusão. Estamos assumindo que sob condições usuais de regularidade $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} N_p(0, \mathcal{J}_0^{-1} \mathcal{I}_0 \mathcal{J}_0^{-1})$. Como é assumido que $\mathcal{I}_0 = \mathcal{J}_0$, logo $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} N_p(0, \mathcal{J}_0^{-1})$. Esse resultado somente é válido para o estimador restrito $\tilde{\theta}$ quando θ_0 pertencer ao interior do conjunto $C\theta \geq 0$, isto é, quando $C\theta_0 > 0$. Se θ_0 está na fronteira do conjunto, $C\theta_0 = 0$, a distribuição assintótica de $\tilde{\theta}$ é muito mais complexa de ser obtida, tendo em geral a forma de um normal multivariada truncada na origem (Wang, 1996). Porém, para demonstrar a equivalência assintótica das várias estatísticas que serão descritas mais adiante precisamos apenas das condições (a)-(e) e das suposições de que $L(\theta)$ é contínua em torno de θ_0 e que os estimadores $\hat{\theta}$, $\tilde{\theta}$ e $\hat{\theta}^0$ são consistentes, onde $\hat{\theta}^0$ é o estimador resultante do problema de maximização de $L(\theta)$ sujeito a $C\theta = 0$. No entanto, como observam

Gouriéroux e Monford (1995, Cap.21) a consistência dos estimadores independe da forma de maximização. Isso quer dizer que se $\hat{\theta}$ for consistente então $\tilde{\theta}$ e $\hat{\theta}^0$ também serão. Assumimos então que $\hat{\theta} \xrightarrow{P} \theta_0$.

Propriedade 2.1 *Tem-se que*

$$\tilde{\lambda} = -(C\Omega_0C^t)^{-1}C\Omega_0\mathcal{U}(\tilde{\theta})$$

onde Ω_0 é uma matriz definida positiva arbitrária de posto k .

Prova. Vimos que

$$\mathcal{U}(\tilde{\theta}) + C^t\tilde{\lambda} = 0.$$

Logo, pré-multiplicando a expressão acima por $C\Omega_0$ obtemos

$$\begin{aligned} C\Omega_0\mathcal{U}(\tilde{\theta}) + C\Omega_0C^t\tilde{\lambda} &= 0, \text{ então} \\ \tilde{\lambda} &= -(C\Omega_0C^t)^{-1}C\Omega_0\mathcal{U}(\tilde{\theta}). \end{aligned} \tag{2.1}$$

□

2.3 Algumas propriedades assintóticas

Como foi mencionado anteriormente a consistência do estimador obtido como a solução do problema de maximização da função objetivo $L(\theta)$ não depende da existência das restrições. Logo, o estimador $\tilde{\theta}$ é consistente para o verdadeiro valor θ_0 desde que o estimador irrestrito $\hat{\theta}$ seja também consistente. Daí segue que

$$\frac{1}{n}\tilde{\lambda} = -(C\Omega_0C^t)^{-1}C\Omega_0\frac{1}{n}\mathcal{U}(\theta_0) \xrightarrow{P} 0,$$

pois a condição assintótica usual de identificabilidade dada em Gouriéroux e Monford (1995, pg.89. Cap.3: pg. 246, Cap 21) em θ_0 implica que

$$\forall \epsilon \geq 0, \lim_{n \rightarrow \infty} Prob\{|\frac{1}{n}\mathcal{U}(\theta_0)| \geq \epsilon\} = 0.$$

Portanto $\frac{1}{n}\tilde{\lambda} \xrightarrow{P} 0$. O estudo da distribuição assintótica de $\frac{\sqrt{n}}{n}\tilde{\lambda} = \frac{1}{\sqrt{n}}\tilde{\lambda}$ é tão complicado quanto o estudo da distribuição assintótica de $\tilde{\theta}$ e não será objeto de estudo neste trabalho

2.4 Teste de hipóteses $H_0 : C\theta = 0 \times H_1 : \{C\theta \geq 0\} - H_0$

Nesta seção, pretendemos descrever algumas estatísticas para testar a hipótese $H_0 : C\theta = 0 \times H_1 : \{C\theta \geq 0\} - H_0$ e mostrar a equivalência assintótica entre elas. Seja $\hat{\lambda}^0$, o vetor de multiplicadores de Lagrange associado às restrições de igualdades $C\theta = 0$ e $\mathcal{U}(\hat{\theta}^0)$ a função escore avaliada em $\hat{\theta}^0$. Similarmente para $\hat{\lambda}^0$, mostra-se que $\frac{1}{n}\hat{\lambda}^0 \xrightarrow{P} 0$. Definimos agora várias estatísticas para testar as hipóteses H_0 e H_1 descritas acima.

2.4.1 Estatística do teste da razão de verossimilhança

A estatística do teste da razão de verossimilhança é definida por

$$\begin{aligned}\xi^R &= 2\{L(\tilde{\theta}) - L(\hat{\theta}^0)\} \\ &= 2\left\{\max_{\theta: C\theta \geq 0} L(\theta) - \max_{\theta: C\theta = 0} L(\theta)\right\}.\end{aligned}\quad (2.2)$$

Sob a hipótese nula $H_0 : C\theta = 0$, os estimadores $\tilde{\theta}$ e $\hat{\theta}^0$ são consistentes para o verdadeiro valor θ_0 . Expandindo em série de Taylor a função $L(\theta)$ em torno de θ_0 , obtemos

$$\begin{aligned}L(\tilde{\theta}) &= L(\theta_0) + \left\{\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}L(\theta_0)\right\}^t \sqrt{n}(\tilde{\theta} - \theta_0) \\ &\quad + \frac{n}{2}(\tilde{\theta} - \theta_0)^t \left\{\frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta^t}L(\theta_0)\right\}(\tilde{\theta} - \theta_0) + R_n,\end{aligned}\quad (2.3)$$

onde $R_n = o_p(\|\tilde{\theta} - \theta_0\|^2)$. Como $\tilde{\theta} - \theta_0 \xrightarrow{P} 0 \Rightarrow \|\tilde{\theta} - \theta_0\|^2 \xrightarrow{P} 0$, logo $\|\tilde{\theta} - \theta_0\|^2 = o_p(1)$. Então podemos escrever $R_n = \frac{\overbrace{R_n}^{o_p(1)}}{\underbrace{\|\tilde{\theta} - \theta_0\|^2}_{o_p(1)}} = o_p(1)$. Temos que, pela condição de regularidade (b), o segundo termo entre $\{ \}$ na expressão (2.3) converge em probabilidade para $-\mathcal{J}_0$. Podemos considerar $\mathcal{J}_0 = \mathcal{I}_0$, então

$$\begin{aligned}\xi^R &= 2\left\{\frac{1}{\sqrt{n}}\mathcal{U}^t(\theta_0)\sqrt{n}(\tilde{\theta} - \theta_0) - \frac{n}{2}(\tilde{\theta} - \theta_0)^t \mathcal{I}_0(\tilde{\theta} - \theta_0) + \frac{n}{2}(\tilde{\theta} - \theta_0)^t \mathcal{I}_0(\tilde{\theta} - \theta_0) \right. \\ &\quad \left. + \frac{n}{2}(\tilde{\theta} - \theta_0)^t \frac{\mathcal{K}(\theta_0)}{n}(\tilde{\theta} - \theta_0) + o_p(1) - \left[\frac{1}{\sqrt{n}}\mathcal{U}^t(\theta_0)\sqrt{n}(\hat{\theta}^0 - \theta_0)\right]\right\}\end{aligned}$$

$$\left. -\frac{n}{2}(\hat{\theta}^0 - \theta_0)' \mathcal{I}_0(\hat{\theta}^0 - \theta_0) + \frac{n}{2}(\hat{\theta}^0 - \theta_0)' \mathcal{I}_0(\hat{\theta}^0 - \theta_0) + \frac{n}{2}(\hat{\theta}^0 - \theta_0)' \frac{\mathcal{K}(\theta_0)}{n}(\hat{\theta}^0 - \theta_0) + o_p(1) \right\}$$

Observe que

$$\begin{aligned} & \frac{n}{2}(\tilde{\theta} - \theta_0)' \mathcal{I}_0(\tilde{\theta} - \theta_0) + \frac{n}{2}(\tilde{\theta} - \theta_0)' \frac{\mathcal{K}(\theta_0)}{n}(\tilde{\theta} - \theta_0) + o_p(1) \\ &= \frac{n}{2}(\tilde{\theta} - \theta_0)' \left(\mathcal{I}_0 + \frac{\mathcal{K}(\theta_0)}{n} \right) (\tilde{\theta} - \theta_0) + o_p(1) \\ &= \frac{\sqrt{n}}{2}(\tilde{\theta} - \theta_0)' \left(\mathcal{I}_0 + \frac{\mathcal{K}(\theta_0)}{n} \right) \sqrt{n}(\tilde{\theta} - \theta_0) + o_p(1) \\ &= \frac{1}{2}o_p(1) + o_p(1) \end{aligned}$$

Similarmente, podemos mostrar que

$$\frac{n}{2}(\hat{\theta}^0 - \theta_0)' \mathcal{I}_0(\hat{\theta}^0 - \theta_0) + \frac{n}{2}(\hat{\theta}^0 - \theta_0)' \frac{\mathcal{K}(\theta_0)}{n}(\hat{\theta}^0 - \theta_0) + o_p(1) = \frac{1}{2}o_p(1) + o_p(1).$$

Portanto,

$$\begin{aligned} \xi^R &= 2 \left\{ \frac{1}{\sqrt{n}} \mathcal{U}^t(\theta_0) \sqrt{n}(\tilde{\theta} - \theta_0) - \frac{n}{2}(\tilde{\theta} - \theta_0)' \mathcal{I}_0(\tilde{\theta} - \theta_0) + o_p(1) \right. \\ &\quad \left. - \left[\frac{1}{\sqrt{n}} \mathcal{U}^t(\theta_0) \sqrt{n}(\hat{\theta}^0 - \theta_0) - \frac{n}{2}(\hat{\theta}^0 - \theta_0)' \mathcal{I}_0(\hat{\theta}^0 - \theta_0) + o_p(1) \right] \right\} \\ &= 2 \left[\frac{1}{\sqrt{n}} \mathcal{U}^t(\theta_0) \sqrt{n}(\tilde{\theta} - \theta_0) - \frac{1}{\sqrt{n}} \mathcal{U}^t(\theta_0) \sqrt{n}(\hat{\theta}^0 - \theta_0) - o_p(1) \right. \\ &\quad \left. - \frac{n}{2}(\tilde{\theta} - \theta_0)' \mathcal{I}_0(\tilde{\theta} - \theta_0) + \frac{n}{2}(\hat{\theta}^0 - \theta_0)' \mathcal{I}_0(\hat{\theta}^0 - \theta_0) + o_p(1) \right] + o_p(1). \quad (2.4) \end{aligned}$$

Implicitamente, assumimos que $L(\theta)$ está bem definida na vizinhança de θ_0 . Expandindo em série de Taylor a função $\mathcal{U}(\theta)$ em torno de θ_0 , obtemos

$$\mathcal{U}(\tilde{\theta}) = \mathcal{U}(\theta_0) + \mathcal{K}(\theta_0)(\tilde{\theta} - \theta_0) + o_p(1),$$

e pelas condições de regularidade iniciais e como

$$\mathcal{U}(\tilde{\theta}) + C^t \tilde{\lambda} = 0.$$

Temos,

$$\begin{aligned} \mathcal{U}(\tilde{\theta}) &= \mathcal{U}(\theta_0) + \mathcal{K}(\theta_0)(\tilde{\theta} - \theta_0) + o_p(1) + C^t \tilde{\lambda} = 0 \\ \frac{1}{\sqrt{n}}\mathcal{U}(\theta_0) + \frac{1}{\sqrt{n}}\mathcal{K}(\theta_0)(\tilde{\theta} - \theta_0) + \frac{1}{\sqrt{n}}C^t \tilde{\lambda} + \frac{1}{\sqrt{n}}o_p(1) &= 0 \\ \frac{1}{\sqrt{n}}\mathcal{U}(\theta_0) - \mathcal{I}_0\sqrt{n}(\tilde{\theta} - \theta_0) + \mathcal{I}_0\sqrt{n}(\tilde{\theta} - \theta_0) \\ + \frac{1}{\sqrt{n}}\mathcal{K}(\theta_0)(\tilde{\theta} - \theta_0) + \frac{1}{\sqrt{n}}C^t \tilde{\lambda} + \frac{1}{\sqrt{n}}o_p(1) &= 0. \end{aligned}$$

Temos que

$$\begin{aligned} \mathcal{I}_0\sqrt{n}(\tilde{\theta} - \theta_0) + \frac{\sqrt{n}}{n}\mathcal{K}(\theta_0)(\tilde{\theta} - \theta_0) + \frac{1}{\sqrt{n}}o_p(1) \\ = \left(\mathcal{I}_0 + \frac{\mathcal{K}(\theta_0)}{n}\right)\sqrt{n}(\tilde{\theta} - \theta_0) + \frac{1}{\sqrt{n}}o_p(1) = o_p(1). \end{aligned}$$

Assim,

$$\frac{1}{\sqrt{n}}\mathcal{U}(\theta_0) - \mathcal{I}_0\sqrt{n}(\tilde{\theta} - \theta_0) + \frac{1}{\sqrt{n}}C^t \tilde{\lambda} + o_p(1) = 0. \quad (2.5)$$

Similarmente para $\hat{\theta}^0$, temos que $\mathcal{U}(\hat{\theta}^0) + C^t \hat{\lambda}^0 = 0$. Logo,

$$\frac{1}{\sqrt{n}}\mathcal{U}(\theta_0) - \mathcal{I}_0\sqrt{n}(\hat{\theta}^0 - \theta_0) + \frac{1}{\sqrt{n}}C^t \hat{\lambda}^0 + o_p(1) = 0. \quad (2.6)$$

Substituindo as expressões (2.5) e (2.6) em (2.4), chegamos ao seguinte :

$$\begin{aligned} \xi^R &= 2 \left\{ \left[\sqrt{n}(\tilde{\theta} - \theta_0)^t \mathcal{I}_0 - \frac{1}{\sqrt{n}} \tilde{\lambda}^t C - o_p(1) \right] \sqrt{n}(\tilde{\theta} - \theta_0) \right. \\ &\quad - \left[\sqrt{n}(\hat{\theta}^0 - \theta_0)^t \mathcal{I}_0 - \frac{1}{\sqrt{n}} (\hat{\lambda}^0)^t C - o_p(1) \right] \sqrt{n}(\hat{\theta}^0 - \theta_0) - o_p(1) \\ &\quad \left. - \frac{n}{2}(\tilde{\theta} - \theta_0)^t \mathcal{I}_0(\tilde{\theta} - \theta_0) + \frac{n}{2}(\hat{\theta}^0 - \theta_0)^t \mathcal{I}_0(\hat{\theta}^0 - \theta_0) + o_p(1) \right\} + o_p(1) \\ &= n(\tilde{\theta} - \theta_0)^t \mathcal{I}_0(\tilde{\theta} - \theta_0) - n(\hat{\theta}^0 - \theta_0)^t \mathcal{I}_0(\hat{\theta}^0 - \theta_0) \\ &\quad - 2 \left\{ \frac{1}{\sqrt{n}} \tilde{\lambda}^t C \sqrt{n}(\tilde{\theta} - \theta_0) + o_p(1) - \left[\frac{1}{\sqrt{n}} (\hat{\lambda}^0)^t C \sqrt{n}(\hat{\theta}^0 - \theta_0) + o_p(1) \right] \right. \\ &\quad \left. - o_p(1) \sqrt{n}(\hat{\theta}^0 - \theta_0) + o_p(1) \sqrt{n}(\tilde{\theta} - \theta_0) \right\} + o_p(1). \end{aligned}$$

Agora, usaremos a condição de exclusão $\{C_j'\tilde{\theta}\}\tilde{\lambda}_j = 0, \quad \forall j$. Note que $C_j'\theta_0 = 0$. Logo temos as igualdades

$$\tilde{\lambda}'C\tilde{\theta} = \frac{\tilde{\lambda}'}{\sqrt{n}}C\sqrt{n}(\tilde{\theta} - \theta_0) + o_p(1) = 0. \quad (2.7)$$

Por outro lado, temos que $C\hat{\theta}^0 = C\theta_0 = 0$. Portanto,

$$(\hat{\lambda}^0)'C\hat{\theta}^0 = \frac{(\hat{\lambda}^0)'}{\sqrt{n}}C\sqrt{n}(\hat{\theta}^0 - \theta_0) + o_p(1) = 0. \quad (2.8)$$

Logo, a estatística do teste ξ^R satisfaz

$$\begin{aligned} \xi^R &= n(\tilde{\theta} - \theta_0)'\mathcal{I}_0(\tilde{\theta} - \theta_0) - n(\hat{\theta}^0 - \theta_0)'\mathcal{I}_0(\hat{\theta}^0 - \theta_0) \\ &\quad - \sqrt{n}(\hat{\theta}^0 - \theta_0)'o_p(1) + \sqrt{n}(\tilde{\theta} - \theta_0)'o_p(1) + o_p(1), \end{aligned} \quad (2.9)$$

e pelo teorema de Slutsky podemos concluir que

$$\xi^R \sim n(\tilde{\theta} - \theta_0)'\mathcal{I}_0(\tilde{\theta} - \theta_0) - n(\hat{\theta}^0 - \theta_0)'\mathcal{I}_0(\hat{\theta}^0 - \theta_0), \quad (2.10)$$

onde $\sim$ significa assintoticamente equivalente.

2.4.2 Estatística do teste tipo Hausman-Wald

A idéia aqui é comparar os estimadores $\tilde{\theta}$ e $\hat{\theta}^0$. A estatística do teste Hausman-Wald é definida por

$$\xi^H = n(\tilde{\theta} - \hat{\theta}^0)'\hat{\mathcal{I}}(\tilde{\theta} - \hat{\theta}^0), \quad (2.11)$$

onde $\hat{\mathcal{I}}$ é estimador consistente de $\mathcal{I}_0$.

Propriedade 2.2 As estatísticas ξ^R e ξ^H são assintoticamente equivalentes sob a hipótese nula, H_0 .

Prova. Substituindo $(\tilde{\theta} - \theta_0)$ por $(\tilde{\theta} - \hat{\theta}^0 + \hat{\theta}^0 - \theta_0)$ na expressão (2.9), obtemos

$$\begin{aligned} \xi^R &= n(\tilde{\theta} - \theta_0)'\mathcal{I}_0(\tilde{\theta} - \theta_0) - n(\hat{\theta}^0 - \theta_0)'\mathcal{I}_0(\hat{\theta}^0 - \theta_0) + o_p[g(\theta_0, \tilde{\theta}, \hat{\theta}^0)] \\ &= n(\tilde{\theta} - \hat{\theta}^0 + \hat{\theta}^0 - \theta_0)'\mathcal{I}_0(\tilde{\theta} - \hat{\theta}^0 + \hat{\theta}^0 - \theta_0) - n(\hat{\theta}^0 - \theta_0)'\mathcal{I}_0(\hat{\theta}^0 - \theta_0) \\ &\quad + o_p[g(\theta_0, \tilde{\theta}, \hat{\theta}^0)] \\ &= n(\tilde{\theta} - \hat{\theta}^0)'\mathcal{I}_0(\tilde{\theta} - \hat{\theta}^0) + 2n(\tilde{\theta} - \hat{\theta}^0)'\mathcal{I}_0(\hat{\theta}^0 - \theta_0) + o_p[g(\theta_0, \tilde{\theta}, \hat{\theta}^0)], \end{aligned} \quad (2.12)$$

onde $o_p[g(\theta_0, \tilde{\theta}, \hat{\theta}^0)] = \sqrt{n}(\tilde{\theta} - \theta_0)' o_p(1) - \sqrt{n}(\hat{\theta}^0 - \theta_0)' o_p(1) + o_p(1)$. Por outro lado, fazendo a diferença (2.6) – (2.5) obtemos

$$\mathcal{I}_0 \sqrt{n}(\tilde{\theta} - \hat{\theta}^0) = \frac{1}{\sqrt{n}} C'(\tilde{\lambda} - \hat{\lambda}^0) + o_p(1) - o_p(1). \quad (2.13)$$

Então,

$$\begin{aligned} \xi^R &= n(\tilde{\theta} - \hat{\theta}^0)' \mathcal{I}_0(\tilde{\theta} - \hat{\theta}^0) + 2 \left\{ \left[\frac{1}{\sqrt{n}}(\tilde{\lambda} - \hat{\lambda}^0)' C + o_p(1) - o_p(1) \right] \sqrt{n}(\hat{\theta}^0 - \theta_0) \right\} \\ &\quad + o_p[g(\theta_0, \tilde{\theta}, \hat{\theta}^0)] \\ &= n(\tilde{\theta} - \hat{\theta}^0)' \mathcal{I}_0(\tilde{\theta} - \hat{\theta}^0) + 2 \frac{1}{\sqrt{n}}(\tilde{\lambda} - \hat{\lambda}^0)' C \sqrt{n}(\hat{\theta}^0 - \theta_0) + o_p(1) \sqrt{n}(\hat{\theta}^0 - \theta_0) \\ &\quad - o_p(1) \sqrt{n}(\hat{\theta}^0 - \theta_0) + o_p[g(\theta_0, \tilde{\theta}, \hat{\theta}^0)] \\ &= n(\tilde{\theta} - \hat{\theta}^0)' \mathcal{I}_0(\tilde{\theta} - \hat{\theta}^0) + 2 \frac{1}{\sqrt{n}}(\tilde{\lambda} - \hat{\lambda}^0)' C \sqrt{n}(\hat{\theta}^0 - \theta_0) + o_p[b(\theta_0, \tilde{\theta}, \hat{\theta}^0)]. \end{aligned}$$

onde $o_p[b(\theta_0, \tilde{\theta}, \hat{\theta}^0)] = o_p(1) \sqrt{n}(\hat{\theta}^0 - \theta_0) - o_p(1) \sqrt{n}(\hat{\theta}^0 - \theta_0) + o_p[g(\theta_0, \tilde{\theta}, \hat{\theta}^0)]$. De (2.7) e (2.8), temos que $\frac{1}{\sqrt{n}}(\tilde{\lambda} - \hat{\lambda}^0)' C \sqrt{n}(\hat{\theta}^0 - \theta_0) \xrightarrow{P} 0$ e pelo teorema de Slutsky $o_p[g(\theta_0, \tilde{\theta}, \hat{\theta}^0)]$ vai em probabilidade para zero. Logo,

$$\xi^R \sim n(\tilde{\theta} - \hat{\theta}^0)' \mathcal{I}_0(\tilde{\theta} - \hat{\theta}^0)$$

e substituindo $\mathcal{I}_0$ por $\hat{\mathcal{I}}$, um estimador consistente, temos que $\xi^R \sim \xi^H$. $\square$

2.4.3 Estatística do teste de multiplicadores de Kühn-Tücker

Seja a estatística do teste de multiplicadores Kühn-Tücker definida por

$$\begin{aligned} \xi^{KT} &= \frac{1}{n}(\tilde{\lambda} - \hat{\lambda}^0)' C \hat{\mathcal{I}}^{-1} C'(\tilde{\lambda} - \hat{\lambda}^0) \\ &= \frac{1}{n} [\mathcal{U}(\tilde{\theta}) - \mathcal{U}(\hat{\theta}^0)]' \hat{\mathcal{I}}^{-1} [\mathcal{U}(\tilde{\theta}) - \mathcal{U}(\hat{\theta}^0)]. \end{aligned} \quad (2.14)$$

Propriedade 2.3 *Sob a hipótese nula H_0 , a estatística ξ^{KT} é assintoticamente equivalente a ξ^R e ξ^H .*

Prova.

$$\begin{aligned} \xi^R &= n(\tilde{\theta} - \hat{\theta}^0)' \mathcal{I}_0 \mathcal{I}_0^{-1} \mathcal{I}_0(\tilde{\theta} - \hat{\theta}^0) + \frac{2}{\sqrt{n}}(\tilde{\lambda} - \hat{\lambda}^0)' C \sqrt{n}(\hat{\theta}^0 - \theta_0) \\ &\quad + o_p[b(\theta_0, \tilde{\theta}, \hat{\theta}^0)] \end{aligned}$$

e substituindo a expressão (2.13) na expressão acima obtemos

$$\begin{aligned}
 \xi^R &= \left[\frac{1}{\sqrt{n}}(\tilde{\lambda} - \hat{\lambda}^0)'C + o_p(1) - o_p(1) \right] \mathcal{I}_0^{-1} \left[\frac{1}{\sqrt{n}}C'(\tilde{\lambda} - \hat{\lambda}^0) + o_p(1) - o_p(1) \right] \\
 &\quad + \frac{2}{\sqrt{n}}(\tilde{\lambda} - \hat{\lambda}^0)'C \sqrt{n}(\hat{\theta}^0 - \theta_0) + o_p[b(\theta_0, \hat{\theta}, \hat{\theta}^0)] \\
 &= \frac{1}{n}(\tilde{\lambda} - \hat{\lambda}^0)'C \mathcal{I}_0^{-1} C'(\tilde{\lambda} - \hat{\lambda}^0) + \frac{1}{\sqrt{n}}(\tilde{\lambda} - \hat{\lambda}^0)'C \mathcal{I}_0^{-1} [o_p(1) - o_p(1)] \\
 &\quad + [o_p(1) - o_p(1)] \mathcal{I}_0^{-1} \frac{1}{\sqrt{n}} C'(\tilde{\lambda} - \hat{\lambda}^0) + [o_p(1) - o_p(1)]' \mathcal{I}_0^{-1} [o_p(1) - o_p(1)] \\
 &\quad + \frac{2}{\sqrt{n}}(\tilde{\lambda} - \hat{\lambda}^0)'C \sqrt{n}(\hat{\theta}^0 - \theta_0) + o_p[b(\theta_0, \hat{\theta}, \hat{\theta}^0)]
 \end{aligned}$$

e pelo teorema de Slutsky podemos dizer que

$$\xi^R \sim \frac{1}{n}(\tilde{\lambda} - \hat{\lambda}^0)'C \mathcal{I}_0^{-1} C'(\tilde{\lambda} - \hat{\lambda}^0). \quad (2.15)$$

Como $\hat{\mathcal{I}}$ é um estimador consistente de $\mathcal{I}_0$ então ξ^{KT} é assintoticamente equivalente a ξ^R e ξ^H .

□

2.4.4 Estatística do teste de Wald

Seja a estatística do teste de Wald definida por

$$\xi^W = n(C\hat{\theta})'(C\hat{\mathcal{I}}^{-1}C')^{-1}C\hat{\theta}. \quad (2.16)$$

Propriedade 2.4 *Sob a hipótese nula, a estatística ξ^{KT} é assintoticamente equivalente a ξ^R, ξ^H e ξ^{KT} .*

Prova. Pela expressão (2.13), temos que

$$(\hat{\theta} - \hat{\theta}^0) = \frac{1}{n} \mathcal{I}_0^{-1} C'(\tilde{\lambda} - \hat{\lambda}^0) + \frac{1}{\sqrt{n}} \mathcal{I}_0^{-1} [o_p(1) - o_p(1)].$$

Então, desde que $C\hat{\theta}^0 = 0$, obtemos

$$C\hat{\theta} = C \frac{1}{\sqrt{n}} \mathcal{I}_0^{-1} C' \frac{1}{\sqrt{n}} (\tilde{\lambda} - \hat{\lambda}^0) + C \frac{1}{\sqrt{n}} \mathcal{I}_0^{-1} [o_p(1) - o_p(1)].$$

Logo,

$$\sqrt{n}C\tilde{\theta} \sim C\mathcal{I}_0^{-1}C^t \frac{1}{\sqrt{n}}(\tilde{\lambda} - \hat{\lambda}^0) \quad (2.17)$$

e substituindo a expressão (2.17) em (2.15) obtemos

$$\begin{aligned} & \sqrt{n}(C\tilde{\theta})^t(C\mathcal{I}_0C^t)^{-1}\sqrt{n}C\tilde{\theta} \\ & \sim \left[C\mathcal{I}_0^{-1}C^t \frac{1}{\sqrt{n}}(\tilde{\lambda} - \hat{\lambda}^0) \right]^t (C\mathcal{I}_0^{-1}C^t)^{-1} \left[C\mathcal{I}_0^{-1}C^t \frac{1}{\sqrt{n}}(\tilde{\lambda} - \hat{\lambda}^0) \right] \\ & = \frac{1}{n}(\tilde{\lambda} - \hat{\lambda}^0)^t C\mathcal{I}_0^{-1}C^t (C\mathcal{I}_0^{-1}C^t)^{-1} C\mathcal{I}_0^{-1}C^t (\tilde{\lambda} - \hat{\lambda}^0) \\ & = \frac{1}{n}(\tilde{\lambda} - \hat{\lambda}^0)^t C\mathcal{I}_0^{-1}C^t (\tilde{\lambda} - \hat{\lambda}^0), \end{aligned}$$

segue-se imediatamente as expressões ξ^{KT} e ξ^W substituindo $\mathcal{I}_0$ por um estimador consistente $\hat{\mathcal{I}}$. $\square$

Em particular se $\mathcal{I}_0 = I$, a matriz identidade, e se a hipótese nula é da forma $H_0 : \theta = 0$ então a estatística do teste de Wald reduz a

$$\xi^W = n \parallel \tilde{\theta} \parallel^2.$$

2.4.5 Estatística do teste escore

A estatística do teste escore é definida por

$$\xi^S = \frac{1}{n}[\mathcal{U}(\hat{\theta}^0) - \mathcal{U}(\tilde{\theta})]^t \hat{\mathcal{I}}^{-1}C^t (C\hat{\mathcal{I}}^{-1}C^t)^{-1} C\hat{\mathcal{I}}^{-1}[\mathcal{U}(\hat{\theta}^0) - \mathcal{U}(\tilde{\theta})]. \quad (2.18)$$

Propriedade 2.5 A estatística ξ^S , sob H_0 , é assintoticamente equivalente a ξ^R, ξ^H, ξ^{KT} e ξ^W .

Prova. Utilizando (2.1) com $\Omega_0 = \mathcal{I}_0^{-1}$, obtemos

$$\frac{1}{\sqrt{n}}\tilde{\lambda} = -(C\mathcal{I}_0^{-1}C^t)^{-1}C\mathcal{I}_0^{-1}\frac{1}{\sqrt{n}}\mathcal{U}(\tilde{\theta}).$$

Obtém-se uma relação similar para $\frac{1}{\sqrt{n}}\hat{\lambda}^0$ e $\frac{1}{\sqrt{n}}\mathcal{U}(\hat{\theta}^0)$. Logo,

$$\xi^{KT} = \frac{1}{\sqrt{n}}(\tilde{\lambda} - \hat{\lambda}^0)^t C\mathcal{I}_0^{-1}C^t \frac{1}{\sqrt{n}}(\tilde{\lambda} - \hat{\lambda}^0) + o_p[v(\theta_0, \tilde{\theta}, \hat{\theta}^0)]$$

$$\begin{aligned}
&= \left[-(C\mathcal{I}_0^{-1}C^t)^{-1}C\mathcal{I}_0^{-1}\frac{1}{\sqrt{n}}\mathcal{U}(\tilde{\theta}) + (C\mathcal{I}_0^{-1}C^t)^{-1}C\mathcal{I}_0^{-1}\frac{1}{\sqrt{n}}\mathcal{U}(\hat{\theta}^0) \right]^t C\mathcal{I}_0^{-1}C^t \\
&\quad \left[-(C\mathcal{I}_0^{-1}C^t)^{-1}C\mathcal{I}_0^{-1}\frac{1}{\sqrt{n}}\mathcal{U}(\tilde{\theta}) + (C\mathcal{I}_0^{-1}C^t)^{-1}C\mathcal{I}_0^{-1}\frac{1}{\sqrt{n}}\mathcal{U}(\hat{\theta}^0) \right] + o_p[v(\theta_0, \tilde{\theta}, \hat{\theta}^0)] \\
&= \frac{1}{\sqrt{n}} [\mathcal{U}(\hat{\theta}^0) - \mathcal{U}(\tilde{\theta})]^t \underbrace{[(C\mathcal{I}_0^{-1}C^t)^{-1}C\mathcal{I}_0^{-1}]^t (C\mathcal{I}_0^{-1}C^t)(C\mathcal{I}_0^{-1}C^t)^{-1}}_{\mathbf{I}} \\
&\quad C\mathcal{I}_0^{-1}\frac{1}{\sqrt{n}} [\mathcal{U}(\hat{\theta}^0) - \mathcal{U}(\tilde{\theta})] + o_p[v(\theta_0, \tilde{\theta}, \hat{\theta}^0)] \\
&\sim \frac{1}{n} [\mathcal{U}(\hat{\theta}^0) - \mathcal{U}(\tilde{\theta})]^t \mathcal{I}_0^{-1}C^t(C\mathcal{I}_0^{-1}C^t)^{-1}C\mathcal{I}_0^{-1} [\mathcal{U}(\hat{\theta}^0) - \mathcal{U}(\tilde{\theta})].
\end{aligned}$$

onde $o_p[v(\theta_0, \tilde{\theta}, \hat{\theta}^0)] = \frac{1}{\sqrt{n}}(\tilde{\lambda} - \hat{\lambda}^0)^t C\mathcal{I}_0^{-1} [o_p(1) - o_p(1)] + [o_p(1) - o_p(1)] \mathcal{I}_0^{-1} \frac{1}{\sqrt{n}} C^t(\tilde{\lambda} - \hat{\lambda}^0) + [o_p(1) - o_p(1)]^t \mathcal{I}_0^{-1} [o_p(1) - o_p(1)] + \frac{2}{\sqrt{n}}(\tilde{\lambda} - \hat{\lambda}^0)^t C \sqrt{n}(\hat{\theta}^0 - \theta_0) + o_p[b(\theta_0, \tilde{\theta}, \hat{\theta}^0)]$

□

2.5 Equivalência assintótica dos problemas de otimização

Sem perda de generalidade, seja a função objetivo $L^*(\theta) = -\frac{n}{2}(\hat{\theta} - \theta)^t \hat{\mathcal{I}}(\hat{\theta} - \theta)$, onde $\hat{\theta}$ é o estimador irrestrito de θ . Podemos definir os seguintes problemas de otimização :

$$\mathcal{P}^* : \begin{cases} \max_{\theta} L^*(\theta) \\ \text{sujeito a } C\theta \geq 0 \end{cases}$$

e

$$\mathcal{P}_0^* : \begin{cases} \max_{\theta} L^*(\theta) \\ \text{sujeito a } C\theta = 0. \end{cases}$$

Propriedade 2.6 As estatísticas $\xi^{*S}, \xi^{*R}, \xi^{*H}, \xi^{*KT}$ e ξ^{*W} , que são baseadas nos problemas de otimização $\mathcal{P}_0^*$ e $\mathcal{P}^*$ são assintoticamente equivalentes às estatísticas $\xi^S, \xi^R, \xi^H, \xi^{KT}$ e ξ^W sob H_0 .

Prova. A prova é semelhante às provas anteriores. □

Considere agora os seguintes problemas de otimização

$$\bar{\mathcal{P}} : \begin{cases} \max_{\theta} -\frac{n}{2}(\hat{\theta} - \theta)^t \hat{\mathcal{I}}(\hat{\theta} - \theta) \\ \text{sujeito a } C\theta_0 + C(\theta - \theta_0) \geq 0 \end{cases}$$

e

$$\bar{\mathcal{P}}_0 : \begin{cases} \max_{\theta} -\frac{n}{2}(\hat{\theta} - \theta)^t \hat{\mathcal{I}}(\hat{\theta} - \theta) \\ \text{sujeito a } C\theta_0 + C(\theta - \theta_0) = 0. \end{cases}$$

Observamos que a estatística para o teste estatístico da razão de verossimilhança sob as hipóteses $H_0 : C\theta = 0 \times H_1 : \{C\theta \geq 0\} - H_0$ é o valor ótimo da função objetivo do problema de otimização

$$\mathcal{P} : \begin{cases} \max_{\theta} -n(\hat{\theta} - \theta)^t \hat{\mathcal{I}}(\hat{\theta} - \theta) + n(\hat{\theta} - \bar{\theta}^0)^t \hat{\mathcal{I}}(\hat{\theta} - \bar{\theta}^0) \\ \text{sujeito a } C\theta_0 + C(\theta - \theta_0) \geq 0, \end{cases}$$

onde $\bar{\theta}^0$ é a solução do problema $\bar{\mathcal{P}}_0$. Isto é, o valor ótimo da função objetivo do problema de otimização $\mathcal{P}$ é dado por

$$2 \left[\max_{\mathcal{P}} L^*(\theta) - \max_{\bar{\mathcal{P}}_0} L^*(\theta) \right].$$

Então, usando a teoria de dualidade para otimização quadrática sob restrições lineares (vide Luenberger, 1969, Cap.8; Avriel, 1976, Cap.7), temos que o valor ótimo da função objetivo do problema $\mathcal{P}$ é também o valor ótimo da função objetivo do problema de otimização abaixo

$$\bar{\mathcal{D}} : \begin{cases} \min_{\lambda} \frac{1}{n}(\lambda - \bar{\lambda}^0)^t C \hat{\mathcal{I}}^{-1} C^t (\lambda - \bar{\lambda}^0) \\ \text{sujeito a } \lambda \geq 0, \end{cases}$$

onde $\bar{\lambda}^0$ é o vetor de multiplicadores de Lagrange no problema de otimização $\bar{\mathcal{P}}_0$. Logo, substituindo a matriz associada à forma quadrática definida na função objetivo por um estimador consistente sob H_0 nada modifica nos resultados assintóticos obtidos.

Propriedade 2.7 *Considere então o problema de otimização*

$$\bar{\mathcal{D}} : \begin{cases} \min_{\lambda} \frac{1}{n}(\lambda - \hat{\lambda}^0)^t C \hat{\mathcal{I}}(\hat{\theta}^0)^{-1} C^t (\lambda - \hat{\lambda}^0) \\ \text{sujeito a } \lambda \geq 0. \end{cases}$$

Seja $\xi^{\mathcal{D}}$ o valor ótimo da função objetivo do problema acima. Essa estatística, que é chamada de estatística dual do problema $\mathcal{P}$, é assintoticamente equivalente a $\xi^S, \xi^R, \xi^H, \xi^{KT} \in \xi^W$ sob H_0 .

2.6 Distribuição nula assintótica da estatística ξ^D

Como foi mostrado anteriormente, as estatísticas do teste são assintoticamente equivalentes a ξ^D sob H_0 . Logo, é suficiente encontrarmos a distribuição nula assintótica de ξ^D .

Note que a variância assintótica de $\frac{1}{\sqrt{n}}\hat{\lambda}^0$ é dada por $Var(\frac{1}{\sqrt{n}}\hat{\lambda}^0) = \Omega_0^{-1} = (C\mathcal{I}_0^{-1}C')^{-1}$ (Gouriéroux e Monfort, Cap. 18). O problema reduz então a encontrar a distribuição assintótica de

$$\xi = \min_{\lambda: \lambda \geq 0} \frac{1}{n} (\lambda - \hat{\lambda}^0)' \Omega_0 (\lambda - \hat{\lambda}^0).$$

Seja $\hat{\mu}^0 = \Omega_0^{1/2} \frac{\hat{\lambda}^0}{\sqrt{n}}$ e $\mu = \Omega_0^{1/2} \frac{\lambda}{\sqrt{n}}$. Então, encontrar a distribuição nula assintótica de ξ é equivalente a encontrar a distribuição assintótica de

$$\xi = \min_{\mu: \Omega_0^{-1/2} \mu \geq 0} \|\mu - \hat{\mu}^0\|^2,$$

onde $\hat{\mu}^0$ é assintoticamente distribuída como uma normal padrão.

Lema 2.1 *Seja X um vetor de variáveis aleatórias normais padrão de dimensão k , isto é, $X \sim N_k(0, I)$. Seja R uma matriz simétrica não-singular de posto k . Então, a distribuição de*

$$\xi = \min_{x: Rx \geq 0} \|X - x\|^2$$

é uma mistura de distribuições do tipo qui-quadrado, isto é

$$\xi \sim \sum_{j=0}^k \omega_j \chi_j^2,$$

onde ω_j , $j = 1, \dots, k$, é uma sequência de pesos satisfazendo

$$\omega_j \geq 0, \text{ e } \sum_{j=0}^k \omega_j = 1,$$

onde χ_0^2 denota o ponto de massa da distribuição na origem.

Prova. Vamos provar inicialmente para o caso de duas restrições. Primeiro vamos

apresentar algumas definições:

1. ϵ_p denota o espaço euclidiano p -dimensional e nós escrevemos $x \geq 0$ ($x > 0$) para indicar que cada componente de x é não-negativa (positiva);
2. Um conjunto $\mathcal{C}$ em ϵ_p é *positivo homogêneo* se $x \in \mathcal{C} \Rightarrow cx \in \mathcal{C}$ para todo real positivo c e assumiremos que esses conjuntos são fechados e convexos;
3. O conjunto $\mathcal{A}$ que contém pelo menos um ponto diferente de zero é dito *unilateral*, se existe um ponto z_* diferente de zero tal que $a^t z_* > 0, \forall a \in \mathcal{A}$ não nulo;
4. Um conjunto $\mathcal{C}$ positivo homogêneo fechado e unilateral é chamado de *cone convexo*.

Sejam as restrições $Rx \geq 0$ para as componentes do vetor Rx , $x = [x_1, x_2]^t$. Desde que R seja uma matriz simétrica então essas condições podem ser escritas como $R_1^t x \geq 0$ e $R_2^t x \geq 0$, onde

$$R_1 = \begin{bmatrix} R_{11} \\ R_{21} \end{bmatrix} \text{ e } R_2 = \begin{bmatrix} R_{12} \\ R_{22} \end{bmatrix}.$$

Então, o conjunto de pontos $C_{(12)} = \{x : R_1^t x \geq 0, R_2^t x \geq 0\}$ constitui um cone convexo em $\mathbb{R}^2$. Para determinarmos tal cone é suficiente encontrarmos dois vetores $\tilde{R}_1$ e $\tilde{R}_2$ tais que

$$\tilde{R}_1^t R_1 > 0, \tilde{R}_2^t R_2 > 0, \tilde{R}_1^t R_2 = 0 \text{ e } \tilde{R}_2^t R_1 = 0.$$

Prova. Considere $\tilde{R} = [\tilde{R}_1, \tilde{R}_2] = R^{-1}$, então o conjunto de pontos $C_{(12)}$ pode ser escrito da forma $\{x : x = \lambda_1 \tilde{R}_1 + \lambda_2 \tilde{R}_2 \text{ com } \lambda_1 \geq 0, \lambda_2 \geq 0\} = C_{(12)}^*$. Neste caso a demonstração segue facilmente.

($\Leftarrow$)

Seja $x_0 \in C_{(12)}^*$, então x_0 pode ser escrito na forma $x_0 = \lambda_1 \tilde{R}_1 + \lambda_2 \tilde{R}_2$ com $\lambda_1 \geq 0, \lambda_2 \geq 0$. Temos então que

$$R_1^t x_0 = R_1^t (\lambda_1 \tilde{R}_1 + \lambda_2 \tilde{R}_2) = \lambda_1 R_1^t \tilde{R}_1 + \lambda_2 R_1^t \tilde{R}_2 = \lambda_1 (\tilde{R}_1^t R_1) + \lambda_2 (\tilde{R}_2^t R_1)^t \Rightarrow R_1^t x_0 \geq 0, \quad (1)$$

$$R_2^t x_0 = R_2^t (\lambda_1 \tilde{R}_1 + \lambda_2 \tilde{R}_2) = \lambda_1 R_2^t \tilde{R}_1 + \lambda_2 R_2^t \tilde{R}_2 = \lambda_1 (\tilde{R}_1^t R_2)^t + \lambda_2 (\tilde{R}_2^t R_2) \Rightarrow R_2^t x_0 \geq 0, \quad (2).$$

De (1) e (2) segue-se $x_0 \in C_{(12)}$.

($\Rightarrow$)

Como $R_1^t x \geq 0$ e $R_2^t x \geq 0$ então $\exists y = [y_1, y_2]^t$ onde $y \geq 0$ tal que

$$\begin{cases} y_1 = R_{11}x_1 + R_{21}x_2 \\ y_2 = R_{12}x_1 + R_{22}x_2. \end{cases}$$

Tomando $\tilde{R}$ que é a inversa de R temos $\tilde{R}y = \tilde{R}R^t x \Rightarrow \tilde{R}y = x$. Isto é,

$$\begin{cases} x_1 = \tilde{R}_{11}y_1 + \tilde{R}_{12}y_2 \\ x_2 = \tilde{R}_{21}y_1 + \tilde{R}_{22}y_2. \end{cases}$$

Portanto $\{x = \lambda_1 \tilde{R}_1 + \lambda_2 \tilde{R}_2 \text{ com } \lambda_1 \geq 0 \text{ e } \lambda_2 \geq 0\} \in C_{(12)}^*$. $\square$

Definindo agora os cones convexos, similarmente ao anterior

$$\begin{aligned} C_{(1)} &= \{x : \tilde{R}_1^t x \geq 0, R_2^t x \leq 0\} = \{x : x = \lambda_1 \tilde{R}_1 + \lambda_2 R_2, \lambda_1 \geq 0, \lambda_2 \leq 0\}; \\ C_{(2)} &= \{x : R_1^t x \leq 0, \tilde{R}_2^t x \geq 0\} = \{x : x = \lambda_1 R_1 + \lambda_2 \tilde{R}_2, \lambda_1 \leq 0, \lambda_2 \geq 0\} \text{ e} \\ C_\phi &= \{x : \tilde{R}_1^t x \leq 0, \tilde{R}_2^t x \leq 0\} = \{x : x = \lambda_1 R_1 + \lambda_2 R_2, \lambda_1 \leq 0, \lambda_2 \leq 0\}. \end{aligned}$$

Seja

$$\xi = \min_{x: Rx \geq 0} \|X - x\|^2 = \|X - Proj_{C_{(12)}} X\|^2,$$

onde $Proj_{C_{(12)}} X$ denota a projeção de X no cone $C_{(12)}$.

Teorema 2.1 *Se $C = L$ é um subespaço linear em $\mathbb{R}^2$ então $C_\phi = L^\perp$ é o complemento ortogonal de L denominado cone dual, com $L^\perp = \{y : x^t y = 0, \forall x \in L\}$.*

Se C é convexo e fechado então $(C_\phi)_\phi = C$, $\forall x, x - Proj_C X = Proj_{C_\phi} X$.

Prova. A demonstração é encontrada em Shapiro (1985b) $\square$

Logo,

$$\xi = \|X - Proj_{C_{(12)}} X\|^2 = \|Proj_{C_\phi} X\|^2.$$

Então, se

$$\begin{cases} x \in C_{(12)}; & \|Proj_{C_\phi} X\|^2 = 0, \text{ pois } Proj_{C_{(12)}} X = X \\ x \in C_{(1)}; & \|Proj_{C_\phi} X\|^2 = \|Proj_{R_2} X\|^2 \\ x \in C_{(2)}; & \|Proj_{C_\phi} X\|^2 = \|Proj_{R_1} X\|^2 \\ x \in C_\phi; & \|Proj_{C_\phi} X\|^2 = \|X\|^2, \text{ pois } \|Proj_{C_{(12)}} X\|^2 = 0. \end{cases}$$

Pelas Figuras 2.1 e 2.2, podemos ver que se $X \in C_{(12)}$, então $Proj X$ em C_ϕ coincide com a projeção no subespaço de dimensão zero. Se $X \in C_{(1)} \cup C_{(2)}$, então $Proj X$ coincide com a projeção no subespaço de dimensão 1 e se $X \in C_\phi$, então $Proj X$ coincide com a projeção no subespaço de dimensão 2.

Figura 2.1 Construições dos cones

Para calcular $Prob\{\xi \in \mathcal{A}\}$, onde $\mathcal{A}$ é um subconjunto arbitrário, temos o seguinte :

$$\begin{aligned}
 Prob\{\xi \in \mathcal{A}\} &= Prob\{\xi \in \mathcal{A} \mid X \in C_{(12)}\} Prob\{X \in C_{(12)}\} \\
 &+ Prob\{\xi \in \mathcal{A} \mid X \in C_{(1)}\} Prob\{X \in C_{(1)}\} \\
 &+ Prob\{\xi \in \mathcal{A} \mid X \in C_{(2)}\} Prob\{X \in C_{(2)}\} \\
 &+ Prob\{\xi \in \mathcal{A} \mid X \in C_{\phi}\} Prob\{X \in C_{\phi}\}.
 \end{aligned}$$

Se $X \in C_{\phi}$, então $\xi = \|X\|^2 = X_1^2 + X_2^2$. Logo, $Prob\{\xi \in \mathcal{A} \mid X \in C_{\phi}\} = Prob\{X_1^2 + X_2^2 \in \mathcal{A} \mid X \in C_{\phi}\} = Prob\{X_1^2 + X_2^2 \in \mathcal{A} \mid X_1 \leq 0, X_2 \leq 0\}$, pois $(X_1, X_2) \in C_{\phi}$ então $\tilde{R}_1' X \leq 0, \tilde{R}_2' X \leq 0 \Rightarrow \tilde{R} X \leq 0 \Rightarrow R \tilde{R} X \leq 0 \Rightarrow X \leq 0$.

Figura 2.2 Projeções no Cone C_ϕ

Como $X \sim N(0,1)$ e fazendo uma transformação em coordenadas polares com $X_1 = d \cos \tau$ e $X_2 = d \sin \tau$ temos que $Prob\{X_1^2 + X_2^2 \in \mathcal{A} \mid X_1 \leq 0, X_2 \leq 0\} = Prob\{d^2 \in \mathcal{A} \mid d \cos \tau \leq 0, d \sin \tau \leq 0\} = Prob\{d^2 \in \mathcal{A} \mid \cos \tau \leq 0, \sin \tau \leq 0\}$. Como d e τ são independentes (Ross, pg.231) obtemos $Prob\{X_1^2 + X_2^2 \in \mathcal{A}\}$, onde $X_1^2 + X_2^2 \sim \chi^2_2$.

Se $X \in C_{(12)} \Rightarrow \xi = 0$. Logo, $Prob\{\xi \in \mathcal{A} \mid X \in C_{(12)}\} = Prob\{0 \in \mathcal{A} \mid X \in C_{(12)}\} = Prob_0\{\mathcal{A}\}$, onde $Prob_0\{\mathcal{A}\}$ é a distribuição do ponto de massa na origem.

Se $X \in C_{(1)} \Rightarrow \xi = \|Proj_{R_2} X\|^2 = X_1^2$. Logo, $Prob\{\xi \in \mathcal{A} \mid X \in C_{(1)}\} = Prob\{X_1^2 \in \mathcal{A} \mid X_1 \geq 0, X_2 \geq 0\}$. Como a distribuição normal é invariante a transformações ortogonais, podemos assumir que cada cone $C_{(1)}$ e $C_{(2)}$, que tem um

ângulo ortogonal ao seu vértice, coincide com o quadrante positivo generalizado. Então, temos que $Prob\{X_1^2 \in \mathcal{A} \mid X_1 \geq 0, X_2 \geq 0\} = Prob\{X_1^2 \in \mathcal{A} \mid X_1 \geq 0\}$. Pelo teorema da probabilidade total podemos escrever $Prob\{X_1^2 \in \mathcal{A}\} = Prob\{X_1^2 \in \mathcal{A} \mid X_1 \geq 0\}Prob\{X_1 \geq 0\} + Prob\{X_1^2 \in \mathcal{A} \mid X_1 \leq 0\}Prob\{X_1 \leq 0\}$. E pela simetria da distribuição normal,

$Prob\{X_1^2 \in \mathcal{A}\} = Prob\{X_1^2 \in \mathcal{A} \mid X_1 \geq 0\}/2 + Prob\{X_1^2 \in \mathcal{A} \mid X_1 \leq 0\}/2$. Logo, $Prob\{X_1^2 \in \mathcal{A} \mid X_1 \geq 0\} = Prob\{X_1^2 \in \mathcal{A}\}$, onde $X_1^2 \sim \chi_1^2$.

Similarmente, se $X \in C_{(2)} \Rightarrow \xi = \|Proj_{R_1} X\|^2 = X_2^2$. Logo, $Prob\{X_2^2 \in \mathcal{A} \mid X_1 \geq 0, X_2 \geq 0\} = Prob\{X_2^2 \in \mathcal{A} \mid X_2 \geq 0\} = Prob\{X_2^2 \in \mathcal{A}\}$, onde $X_2 \sim \chi_1^2$. Portanto,

$$Prob\{\xi \in \mathcal{A}\} = \omega_2 \chi_2^2 + \omega_1 \chi_1^2 + \omega_0 \chi_0^2,$$

onde

$$\begin{aligned} \omega_0 &= Prob\{X \in C_{(12)}\} \\ \omega_1 &= Prob\{X \in C_{(1)}\} + Prob\{X \in C_{(2)}\} \\ \omega_2 &= Prob\{X \in C_\phi\} \end{aligned}$$

e χ_0^2 denota a distribuição degenerada na origem. Para o caso geral, isto é, quando k é arbitrário, a prova acompanha esses passos. É necessário que definamos R_j , $j = 1, \dots, k$, vetores coluna da matriz R e $\tilde{R}_j$, $j = 1, \dots, k$, vetores coluna da matriz R^{-1} . Então, para cada subconjunto $\mathcal{A}$ de $\{1, \dots, k\}$, definimos o cone

$$C_{\mathcal{A}} = \{x = \sum_{j \in \mathcal{A}} \lambda_j \tilde{R}_j + \sum_{j \notin \mathcal{A}} \lambda_j R_j, \text{ com } \lambda_j \geq 0 \text{ se } j \in \mathcal{A}, \text{ e } \lambda_j \leq 0 \text{ se } j \notin \mathcal{A}\}.$$

Então, segue o resultado que

$$\xi \sim \sum_{j=0}^k \omega_j \chi_j^2,$$

onde

$$\omega_j = \sum_{\{\mathcal{A}: \text{cardinal de } \mathcal{A} = k-j\}} Prob\{X \in C_{\mathcal{A}}\}. \quad (2.19)$$

□

Propriedade 2.8 *As estatísticas $\xi^R, \xi^H, \xi^W, \xi^{KT}, \xi^S$ e ξ^D , sob H_0 são todas assintoticamente distribuídas como uma mistura de qui-quadrados, isto é,*

$$\xi^D \sim \sum_{j=0}^k \omega_j \chi_j^2,$$

onde os pesos são dados por (2.19) e a matriz $R = (C\mathcal{I}_0^{-1}C')^{-1/2}$.

Prova. A prova é a utilização imediata do Lema 2.1. $\square$

2.6.1 Pesos

Nas seções anteriores, vimos que a distribuição nula da estatística do teste depende de pesos. Esses pesos são conhecidos como probabilidades de nível pois significam a probabilidade do vetor $C\hat{\theta}$ ter exatamente ℓ componentes maiores que zero, e dependem da matriz R^{-2} que pode depender dos parâmetros. Um caso especial verifica-se quando R^{-2} é uma matriz identidade de ordem k . Nesse caso, os pesos ficam dados por

$$\omega_\ell(k, I_k) = \binom{k}{\ell} 2^{-k}, \ell = 0, \dots, k.$$

No caso normal linear com uma única restrição, $k = 1$, a estatística da razão de verossimilhança que coincide com as demais estatísticas fica dada por

$$\xi^R = \frac{(C\hat{\theta})^2}{\text{Var}(C\hat{\theta})}.$$

Seja

$$C\hat{\theta} = \begin{cases} C\hat{\theta} & \text{se } C\hat{\theta} > 0 \\ 0 & \text{se } C\hat{\theta} \leq 0. \end{cases}$$

Fazendo $\tilde{z} = C\hat{\theta}/\{\text{Var}(C\hat{\theta})\}^{1/2}$, obtém-se

$$\xi^R = \begin{cases} \tilde{z}^2 & \text{se } \tilde{z} > 0 \\ 0 & \text{se } \tilde{z} \leq 0 \end{cases}$$

e como foi mostrado anteriormente sob a hipótese nula. $H_0 : C\theta = 0$, a distribuição nula de ξ^R para $c > 0$ é dada por $\frac{1}{2}\lambda_1^2$ e a região crítica de nível de significância α fica dada por $\{\xi^R > c_\alpha\}$, onde c_α é tal que $\text{Prob}\{\lambda_1^2 > c_\alpha\} = 2\alpha$. Podemos ver

que, se aplicarmos um teste tradicional bicaudal, a região crítica seria definida por $\{\xi^R > c_\alpha^*\} = \alpha$, onde c_α^* é tal que $Prob\{\chi_1^2 > c_\alpha^*\} = \alpha$. Observe que c_α^* é sempre maior do que c_α , logo o teste irrestrito tradicional induz a uma aceitação mais frequente do que o teste restrito.

Suponha agora o caso de $k = 2$ restrições. Como a distribuição de X é invariante sob transformações ortogonais, mostra-se que $Prob\{X \in C_A\} = \alpha_a/2\pi$, onde α_a é o ângulo formado pelo vértice do cone C_A . Temos então que $\omega_1 = 1/2$. Temos ainda que

$$\cos \alpha_{12} = \frac{\tilde{R}_1^t \tilde{R}_2}{\sqrt{\tilde{R}_1^t \tilde{R}_1} \sqrt{\tilde{R}_2^t \tilde{R}_2}}.$$

Desde que R^{-1} é uma matriz simétrica, essa razão é igual a

$$\frac{r_{12}}{\sqrt{r_{11}} \sqrt{r_{22}}},$$

onde r_{ij} é o termo genérico de $R^{-2} = C\mathcal{I}^{-1}(\theta_0)C^t$. Logo, o resultado

$$\cos \alpha_{12} = \frac{C_1^t \mathcal{I}^{-1}(\theta_0) C_2}{\sqrt{C_1^t \mathcal{I}^{-1}(\theta_0) C_1} \sqrt{C_2^t \mathcal{I}^{-1}(\theta_0) C_2}},$$

onde C_i^t é i -ésima linha de C . Assim,

$$\omega_0 = \frac{\alpha_{12}}{2\pi} \text{ e } \omega_2 = \frac{1}{2} - \frac{\alpha_{12}}{2\pi}.$$

Então, ω_0 pode ser calculado como

$$\omega_0 = \frac{1}{2\pi} \cos^{-1}(\rho_{as}(C_1^t \hat{\theta}, C_2^t \hat{\theta})),$$

onde ρ_{as} denota o coeficiente de correlação linear assintótico entre $C_1^t \hat{\theta}$ e $C_2^t \hat{\theta}$. Portanto, para calcular os pesos devemos calcular a probabilidade do quadrante positivo generalizado de uma normal k -variada de média zero e matriz de variância-covariância R^{-2} . Podemos comentar que há forma explícita para até três restrições (Apêndice A). No caso de quatro ou mais restrições são utilizados métodos numéricos de integração. Childs (1967) apresenta uma fórmula reduzida para calcular as probabilidades do quadrante e também uma metodologia para a simplificação

dessa probabilidade em uma única integral no intervalo $[0, 1]$ para o caso de $k = 4$. Sun (1988a) mostra que essa metodologia pode ser estendida para o caso $k \geq 4$, e que a probabilidade do quadrante positivo generalizado é calculada como uma combinação linear de integrais de ordem $([k/2] - 1)$ no intervalo $[0, 1]$. Vemos que para o caso de que $k = 4$ e $k = 5$, somente é preciso calcular uma integral no intervalo $[0, 1]$. Para os casos de $k = 6$ e $k = 7$, calculamos uma integral dupla em $[0, 1] \times [0, 1]$ e assim por diante. Bohrer e Chow (1978) desenvolveram um programa em Fortran para o cálculo das probabilidades de nível até $k = 10$ restrições. Sun (1988b) também desenvolveu um programa para o cálculo dessas probabilidades. A dificuldade no cálculo dessas probabilidades tem motivado o surgimento de pesquisas com o objetivo de encontrar aproximações para os coeficientes de $\omega_\ell(k, R^{-2})$ s. Podemos citar o livro de Robertson et al. (1988, Cap.3) onde encontra-se uma excelente revisão sobre a abordagem e procedimentos para simplificar o cálculo desses pesos. Em alguns casos particulares, tais como hipóteses de quase-ordem ($\mu_i \leq \mu_j$ e $\mu_j \leq \mu_\ell$ então $\mu_i \leq \mu_\ell$) para as médias de k populações normais independentes. Também encontra-se em Kodde e Palm (1986) limites superior e inferior para as probabilidades de nível os quais valem para situações mais gerais,

$$\frac{1}{2} \text{Prob}\{\chi_1^2 \geq c\} \leq \text{Prob}\{\xi^R \geq c\} \leq \frac{1}{2} \text{Prob}\{\chi_{k-1}^2 \geq c\} + \frac{1}{2} \text{Prob}\{\chi_k^2 \geq c\},$$

onde $c > 0$. Kodde e Palm (1986) apresentam também valores críticos para os limites acima para alguns níveis de significância usuais. Com o aumento do número de restrições essa banda fica imprecisa, dificultando a decisão.

2.6.2 Exemplo

Suponha que o nosso problema seja fazer inferências sobre o parâmetro $\theta = [\theta_1, \theta_2]^t$. Assumimos que o problema tenha sido normalizado tal como o estimador de máxima verossimilhança $\hat{\theta}$ de θ satisfaz $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} N(0, \sigma^2 I)$. Esse é o caso do modelo $y = X\theta + u, u \sim N(0, \sigma^2 I)$, onde X é uma matriz $n \times 2$ satisfazendo $X^t X = I$. Nós impomos as seguintes restrições :

$$(i) \begin{cases} \theta_1 & \geq 0 \\ \theta_2 - r\theta_1 & \geq 0, \end{cases}$$

onde r é uma constante.

O interesse é testar $H_0 : \theta = 0 \times H_1 : \{C\theta \geq 0\} - H_0$, onde

$$C = \begin{bmatrix} 1 & 0 \\ -r & 1 \end{bmatrix}.$$

Um estimador, que é assintoticamente equivalente ao estimador restrito por desigualdades $\tilde{\theta}$ pode ser considerado como a projeção ortogonal de $\hat{\theta}$ no conjunto definido pelas inequações (i) (Fig. 2.3).

Figura 2.3 Projeções de $\hat{\theta}$

Então, a região crítica do teste Hausman-Wald fica dada por

$$RC = \left\{ \tilde{\theta} \in \mathbb{R}^2; n \frac{\|\tilde{\theta}\|^2}{\hat{\sigma}^2} \geq c \right\},$$

onde c é o percentil $(1 - \alpha)$ da mistura de qui-quadrados

$$\left(\frac{1}{2} - \frac{\phi}{2\pi} \right) \lambda_0^2 + \frac{1}{2} \lambda_1^2 + \frac{\phi}{2\pi} \lambda_2^2,$$

onde ϕ denota o ângulo das linhas dados por $\theta_1 = 0$, $\theta_2 - r\theta_1 = 0$, $\hat{\sigma}^2$ é um estimador consistente de σ^2 e

$$\alpha = \frac{1}{2} \text{Prob}\{\chi_1^2 > c\} + \frac{\phi}{2\pi} \text{Prob}\{\chi_2^2 > c\}.$$

Figura 2.4 Região crítica do teste de Haussman-Wald

2.7 Teste de hipóteses $H_0 : C\theta \geq 0 \times H_1 : \mathbb{R}^p - H_0$

Quando não existe uma lei que justifique a restrição populacional e sim uma suspeita da existência das restrições é interessante testar hipóteses do tipo $H_0 : C\theta \geq 0 \times H_1 : \mathbb{R}^p - H_0$. Podemos observar que a hipótese nula é composta com restrições nos parâmetros enquanto que a hipótese alternativa é irrestrita. Podemos também, definir algumas estatísticas do teste similarmente ao caso anterior. Por exemplo, a estatística da razão de verossimilhança fica agora dada por

$$\xi^R = 2 \{L(\hat{\theta}) - L(\tilde{\theta})\},$$

já a estatística do tipo Wald toma a forma

$$\xi^W = \min_{\theta: C\theta \geq 0} \{n(\hat{\theta} - \theta)^t \mathcal{I}(\hat{\theta})(\hat{\theta} - \theta)\}.$$

Como a hipótese nula é composta, devemos salientar a seguinte observação :

- A distribuição nula assintótica de ξ^W pode não existir e depende do verdadeiro valor do parâmetro θ_0 que satisfaz $C\theta \geq 0$. Trabalhando sobre um região crítica do tipo $\{\xi^W \geq c\}$, o erro tipo I fica dado por

$$\sup_{\theta_0: C\theta_0 \geq 0} \text{Prob}_{\theta_0} \{\xi^W \geq c\} \leq \alpha.$$

Wolak (1991) propôs um lema no qual apresenta uma metodologia para encontrar a situação nula menos favorável.

2.7.1 Lema de Wolak

Antes de apresentarmos o lema de Wolak, vamos mostrar alguns resultados encontrados por Perlman (1969), que propôs uma maneira de resolver o teste de hipóteses da forma $H_0 : \gamma \geq 0 \times H_1 : \mathbb{R}^k - H_0$ levando em conta que $\hat{\gamma} \sim N_k(\gamma, \Psi_0)$, onde Ψ_0 é uma matriz definida positiva conhecida. Seja

$$W_n^P = \min_{t: t \geq 0} (\hat{\gamma} - t)^t \Psi_0^{-1} (\hat{\gamma} - t). \quad (2.20)$$

Como a hipótese nula é composta, devemos procurar a situação menos favorável

no conjunto determinado sob H_0 para determinar o valor crítico exato de nível α . Perlman (1969) mostrou, para $c > 0$, que

$$\sup_{\gamma \geq 0} \text{Prob}\{W_n^P \geq c \mid \gamma, \Psi_0\} = \text{Prob}\{W_n^P \geq c \mid 0, \Psi_0\},$$

onde $\text{Prob}\{W_n^P \geq c \mid \gamma, \Psi_0\}$ é a probabilidade do evento $[W_n^P \geq c]$ dado que $\hat{\gamma}$ em (2.20) é $N_k(\gamma, \Psi_0)$. É importante salientar que esse resultado somente é válido quando não há dependência funcional entre Ψ_0 e o vetor de parâmetros γ .

Para o caso da dependência funcional entre Ψ_0 e o vetor de parâmetros γ , Wolak (1991) propôs um lema. Para mostrar esse lema, precisamos de algumas suposições adicionais de regularidade que são os Teoremas 4.1.2 e 4.1.3 dados em Amemiya (1985) e mais $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} N_p(0, \mathcal{I}^{-1}(\theta_0)) \quad \forall \theta_0 \in \text{interior } \Theta$ onde $\mathcal{I}(\theta) = \lim_{n \rightarrow \infty} E_{\theta_0} \left[-\frac{1}{n} \frac{\partial^2}{\partial \theta \partial \theta^t} L(\theta) \right]$.

Pelas condições de regularidade iniciais que permitem uma aplicação do teorema da convergência dominada podemos considerar $\mathcal{I}(\theta_0) = \mathcal{J}_0$. Assumindo que $\mathcal{J}_0$ seja estimada consistentemente por $\hat{V}(\theta) = -\frac{1}{n} \frac{\partial^2}{\partial \theta \partial \theta^t} L(\theta)$ e que θ_0 é o verdadeiro valor de θ , então $\hat{V}(\hat{\theta}) \xrightarrow{P} \mathcal{J}_0$, tal como $\hat{V}^{-1}(\hat{\theta})$ é um estimador consistente da matriz de covariância assintótica de $\sqrt{n}(\hat{\theta} - \theta_0)$. Para n suficientemente grande temos o seguinte modelo :

$$\hat{\gamma} = \gamma + \nu, \quad \nu \sim N_k(0, \Delta(\theta)), \quad (2.21)$$

onde $\gamma = \sqrt{n}(C\theta)$ e $\Delta(\theta) = CV^{-1}(\theta)C^t$. Seja agora a estatística

$$W_n = \min_{t: t \geq 0} (\hat{\gamma} - t)^t \hat{\Delta}^{-1}(\hat{\theta})(\hat{\gamma} - t), \quad (2.22)$$

onde $\hat{\Delta}(\hat{\theta}) = C\hat{V}^{-1}(\hat{\theta})C^t$ e $\Delta(\theta_0) = CV^{-1}(\theta_0)C^t$. Para enunciar o lema, precisamos definir alguns conjuntos : seja $S = \{\theta : C\theta \geq 0\}$, $S^i = \{\theta : C\theta > 0\}$ e $S^b = S - S^i$. Seja $A = \{\theta : \theta \in S^b \text{ onde exatamente um restrição em igualdades } C_i^t \theta = 0, i = 1, \dots, k \text{ é satisfeita}\}$ e $B = S^b - A$, onde pelo menos duas desigualdades são satisfeitas na forma de igualdades. Seja o conjunto $E = \{\theta \mid C\theta = 0\}$ que é o conjunto onde todas as desigualdades são satisfeitas em forma de igualdades. E mais, para todo $\theta \in S^b = A \cup B$, seja C_b a submatriz de C com $m \leq k$ linhas tais que $C_i^t \theta = 0$.

Lema 2.2 Para testar a hipótese $H_0 : C\theta \geq 0 \times H_1 : \mathbb{R}^p - H_0$ temos os seguintes resultados :

(i) para todo $\theta_0 \in S^i$,

$$\lim_{n \rightarrow \infty} \text{Prob}\{W_n = 0 \mid \theta_0\} = 1; \quad (2.23)$$

(ii) para algum $c > 0$ e $\theta_0 \in S^b$,

$$\lim_{n \rightarrow \infty} \text{Prob}\{W_n \geq c \mid \theta_0\} = \sum_{j=0}^m \omega_j \text{Prob}\{\chi_{m-j}^2 \geq c\}, \quad (2.24)$$

onde $\omega_j = \omega_j(m, \Delta_b)$, $\Delta_b = C_b \mathcal{J}_0^{-1} C_b^t$;

(iii) para todo $\theta_0 \in B$ e $\theta_0^* \in A$,

$$\lim_{n \rightarrow \infty} \text{Prob}\{W_n \geq c \mid \theta_0\} \geq \lim_{n \rightarrow \infty} \text{Prob}\{W_n \geq c \mid \theta_0^*\}. \quad (2.25)$$

Podemos perceber que a parte (i) deste lema reduz o número de elementos de S^i que podem ser levados em consideração. Essa parte nos diz que $W_n \xrightarrow{P} 0$. A segunda parte caracteriza a distribuição assintótica de W_n para os valores $\theta_0 \in S^b$ e a terceira parte seleciona alguns elementos de S^b como possíveis valores que resultam na situação menos favorável de θ_0 . Este lema especifica que $B \subset S$ deve conter o valor menos favorável de θ_0 sob H_0 , mas, em geral, ele não fornece a solução de θ_0 tal que

$$\sup_{\theta_0 \in S} \lim_{n \rightarrow \infty} \text{Prob}\{W_n \geq c \mid \theta_0\}, \quad (2.26)$$

para um dado $c > 0$. Para resolver o problema (2.26), Wolak propôs um algoritmo que devido às dificuldades computacionais é usado somente para demonstrar que o valor menos favorável existe. Seja $c > 0$ um valor crítico arbitrário. Para um valor fixo de $\bar{\theta}_0 \in B$, o primeiro passo é determinar os elementos $C\theta$ que contêm o vetor $C_b \theta$ descrito na parte (ii) do lema. Aplica-se $C_b \theta$ em $\bar{\theta}_0$ e pela equação (2.24) calcula-se $\lim_{n \rightarrow \infty} \text{Prob}\{W_n \geq c \mid \bar{\theta}_0\}$ utilizando a parte (ii) do lema. Esse processo é repetido para todo $\theta_0 \in B$. O valor de $\theta_0 \in B$ que maximiza as probabilidades

é o valor menos favorável de θ_0 determinando assintoticamente o tamanho exato do teste para o valor crítico de c . Pela dependência funcional desse valor menos favorável de θ_0 em c , denotamos o mesmo por $\theta_0^*(c)$. Então, o valor crítico para tamanho exato α de um teste assintótico irrestrito é o c que resolve

$$\lim_{n \rightarrow \infty} \text{Prob}\{W_n \geq c \mid \theta_0^*(c)\} = \alpha.$$

Como já mencionado, o valor de $\omega_j(k, \Delta)$ na parte (ii) do lema tem forma fechada para $k \leq 4$. Kudô (1963) fornece as expressões para esses pesos para um valor arbitrário de k como sendo a soma de produtos de probabilidades de normais multivariadas. Consequentemente, o principal problema é determinar a matriz $\Delta_b(\theta_0)$. Wolak salienta que existem duas situações em que há unicidade no valor menos favorável de θ_0 para o teste. Quando a matriz $\mathcal{J}_0$ é uma matriz diagonal para todo $\theta_0 \in B$ e quando as restrições de desigualdades tomam formas simples. Nesses casos, o valor menos favorável é o único θ_0 que satisfaz todas as restrições em igualdades.

2.7.2 Exemplo

Considere o modelo linear com observações independentes e igualmente distribuídas de uma distribuição normal bivariada

$$X_i \sim N_2(\mu, \Sigma),$$

onde $\mu = [0, 0]^t$ e $\Sigma = \begin{bmatrix} \sigma^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2 \end{bmatrix}$. Temos que $\theta = [\sigma^2, \tau^2, \rho]^t$ e $\Theta = \{\theta : \theta_1 \geq 0, \theta_2 \geq 0, -1 \leq \theta_3 \leq 1\}$. Os estimadores de máxima verossimilhança são dados por $\hat{\theta} = [\bar{X}_1, \bar{X}_2, r_{12}]^t$, onde $\bar{X}$ representam as médias amostrais e r_{12} a correlação amostral entre X_1 e X_2 (Lehmann, 1983, pg.439-440). Sob condições usuais de regularidade temos que $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} N_3(0, \mathcal{J}_0^{-1})$, onde $\mathcal{J}_0^{-1}$ é dada em Lehmann (1983, pg.441).

A matriz $\mathcal{J}(\theta)$ é dada por

$$(1 - \rho^2)\mathcal{J}(\theta) = \begin{bmatrix} \frac{2 - \rho^2}{4\sigma^4} & \frac{-\rho^2}{4\sigma^2\tau^2} & \frac{-\rho}{2\sigma^2} \\ \frac{-\rho^2}{4\sigma^2\tau^2} & \frac{2 - \rho^2}{4\tau^4} & \frac{-\rho}{2\tau^2} \\ \frac{-\rho}{2\sigma^2} & \frac{-\rho}{2\tau^2} & \frac{1 + \rho^2}{1 - \rho^2} \end{bmatrix}.$$

A matriz de variância - covariância de $\sqrt{n}(\hat{\theta} - \theta_0)$ fica expressa na forma

$$\mathcal{J}(\theta)^{-1} = \begin{bmatrix} 2\sigma^4 & 2\rho^2\sigma^2\tau^2 & \rho(1 - \rho^2)\sigma^2 \\ 2\rho^2\sigma^2\tau^2 & 2\tau^4 & \rho(1 - \rho^2)\tau^2 \\ \rho(1 - \rho^2)\sigma^2 & \rho(1 - \rho^2)\tau^2 & (1 - \rho^2)^2 \end{bmatrix}.$$

Nosso interesse é testar

$$H_0 : \theta \leq \theta^V = [1, 0; 1, 0; 0.95]^t \times H_1 : \mathbb{R}^3 - H_0.$$

onde podemos observar que queremos testar a suspeita de estrutura linear mais fraca. A matriz Δ é dada por

$$\Delta = \begin{bmatrix} 1 & \rho^2 & \rho\sqrt{2} \\ \rho^2 & 1 & \rho\sqrt{2} \\ \rho\sqrt{2} & \rho\sqrt{2} & 1 \end{bmatrix},$$

$$C = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

e $d = [1, 0; 1, 0; 0.95]^t$.

Uma das estatísticas do teste toma a forma

$$W_n = \min_{t: t \leq \theta^V} [n(\hat{\theta}^* - t)\mathcal{J}(\hat{\theta}^*)(\hat{\theta}^* - t)],$$

onde $\theta^* = C\theta - d$. Para $\theta_0 = \theta^V$ a distribuição limite exata de ξ^W é dada por

$$\lim_{n \rightarrow \infty} Prob\{\xi^W \geq c\} = 0,015263 Prob\{\chi_3^2 \geq c\} + 0,168204 Prob\{\chi_2^2 \geq c\} \\ + 0,484737 Prob\{\chi_1^2 \geq c\},$$

onde os pesos são calculados de acordo com a matriz avaliada em $\theta = \theta^V$ e as

fórmulas são dadas no Apêndice A. Para o caso em especial em que $\rho \leq 0.95$ avaliada em $\theta_0 = \theta^B = [1, 0; 1, 0; 0, 0]^t$, a matriz Δ fica dada por

$$\Delta = \begin{bmatrix} 1 & \rho^2 \\ \rho^2 & 1 \end{bmatrix},$$

onde os pesos são calculados de acordo com a matriz avaliada em $\theta = \theta^B$ e as fórmulas são dadas no Apêndice A. Tomando $\theta_0 = \theta^B = [1, 0; 1, 0; 0, 0]^t$ a distribuição assintótica fica expressa por

$$\lim_{n \rightarrow \infty} Prob\{\xi^W \geq c\} = \frac{1}{4} Prob\{\chi_2^2 \geq c\} + \frac{1}{2} Prob\{\chi_1^2 \geq c\},$$

com os pesos calculados de acordo com a matriz Δ . Wolak (1991) mostra que θ^B é a situação menos favorável. Podemos ver através da Figura 2.5 que a distribuição nula assintótica menos favorável para este teste de hipóteses nem sempre é alcançada quando todas as restrições são satisfeitas em igualdades. Denotando $\Gamma(c | \theta) = \lim_{n \rightarrow \infty} Prob\{\xi^W \geq c | \theta\}$, onde na Figura θ^V e θ^B referem-se a $\Gamma(c | \theta^V)$ e $\Gamma(c | \theta^B)$, respectivamente. Vimos na Figura que para todo valor de c , $\Gamma(c | \theta^B) > \Gamma(c | \theta^V)$, apesar do fato de que θ^V é o valor de θ que satisfaz todas as restrições na forma de igualdades. Observamos que neste caso, por causa da dependência funcional da matriz de variância-covariância assintótica de $\sqrt{n}(\hat{\theta} - \theta_0)$ em θ_0 , o valor menos favorável de $\theta_0 \in S$ somente satisfaz 2 das 3 desigualdades na forma de igualdades.

Figura 2.5 Probabilidades de rejeição sob a hipótese nula

CAPÍTULO 3

Modelos Lineares Generalizados

3.1 Introdução

3.1.1 Definição

Suponha $Y_1, \dots, Y_n$ variáveis aleatórias independentes, cada uma com densidade na forma

$$f(y; \theta_i, \phi) = \exp\{\phi[y\theta_i - b(\theta_i)] + c(y, \phi)\}, \quad (3.1)$$

onde $c(\cdot)$ é uma função conhecida, $E(Y_i) = \mu_i = b'(\theta_i)$, $Var(Y_i) = \phi^{-1}V_i$, $V = d\mu/d\theta$ é a função de variância e $\phi^{-1} > 0$ é o parâmetro de dispersão conhecido. A função de variância determina de uma forma biunívoca a classe correspondente de distribuições. Essa propriedade é muito importante, pois permite a comparação de distribuições através de um teste simples para a função de variância. Os modelos lineares generalizados (MLGs) são definidos por (3.1) e pela componente sistemática

$$g(\mu_i) = \eta_i, \quad (3.2)$$

onde $\eta = x^t\beta$ é o preditor linear, $\beta = (\beta_1, \dots, \beta_p)^t$, $p < n$, é um vetor de parâmetros desconhecidos a serem estimados, $x_i = (x_{i1}, \dots, x_{ip})^t$ representa os valores de p variáveis explicativas e $g(\cdot)$ é uma função monótona e diferenciável, denominada função de ligação.

3.1.2 Casos particulares

Podemos citar algumas distribuições pertencentes à família exponencial de distribuições :

Normal

Seja Y uma variável aleatória com distribuição normal com média μ e variância σ^2 , $Y \sim N(\mu, \sigma^2)$. A densidade de Y é da forma

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(y - \mu)^2\right\} = \exp\left\{\frac{1}{\sigma^2}\left(\mu y - \frac{\mu^2}{2}\right) - \frac{1}{2}\left[\log 2\pi\sigma^2 + \frac{y^2}{\sigma^2}\right]\right\},$$

onde $-\infty < \mu, y < \infty$ e $\sigma^2 > 0$. É fácil ver que $\theta = \mu$, $b(\theta) = \theta^2/2$, $\phi = \sigma^2$ e $c(y, \phi) = \frac{1}{2} \log \phi/2\pi - \frac{\phi y^2}{2}$ e a função de variância é dada por $V(\mu) = 1$.

Poisson

Para $Y \sim P(\mu)$, a densidade fica dada por

$$e^{-\mu} \mu^y / y! = \exp\{y \log \mu - \mu - \log y!\},$$

onde $\mu > 0$ e $y = 0, 1, \dots$. Fazendo $\log \mu = \theta$, $b(\theta) = e^\theta$, $\phi = 1$ e $c(y, \phi) = -\log y!$ e a função de variância fica dada por $V(\mu) = \mu$.

Binomial

Seja Y^* a proporção de sucessos em n ensaios de Bernoulli com probabilidade de sucesso μ . Assumiremos que $nY^* \sim B(n, \mu)$. A densidade de Y^* fica definida por

$$\binom{n}{ny^*} \mu^{ny^*} (1 - \mu)^{n - ny^*} = \exp\left\{\log \binom{n}{ny^*} + ny^* \log \left[\frac{\mu}{1 - \mu}\right] + n \log(1 - \mu)\right\},$$

onde $0 < \mu, y^* < 1$. Temos (3.1) fazendo $\phi = n$, $\theta = \log\{\mu/(1 - \mu)\}$, $b(\theta) = \log(1 + e^\theta)$ e $c(y, \phi) = \log \binom{\phi}{\phi y^*}$. A função de variância fica dada por $V(\mu) = \mu(1 - \mu)$.

Outras distribuições são apresentadas na Tabela 3.1

Tabela 3.1 *Características de algumas distribuições da família exponencial*

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

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

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

3.1.3 Estatísticas suficientes e ligações canônicas

O log da função de verossimilhança de um MLG com respostas independentes pode ser expresso na forma

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

Um caso importante dos MLGs é quando o parâmetro natural da família exponencial (θ) coincide com o preditor linear, isto é, $\theta_i = \eta_i = \sum_{j=1}^p x_{ij} \beta_{ij}$. Nesse caso, $L(\beta; y)$ fica definida por

$$L(\beta; y) = \sum_{i=1}^n \phi \left\{ y_i \sum_{j=1}^p x_{ij} \beta_{ij} - b \left(\sum_{j=1}^p x_{ij} \beta_{ij} \right) \right\} + \sum_{i=1}^n c(y_i, \phi).$$

Seja a estatística $S_j^* = \phi \sum_{i=1}^n Y_i x_{ij}$, então $L(\beta; y)$ fica dada por

$$L(\beta; y) = \sum_{j=1}^p s_j^* \beta_j - \phi \sum_{i=1}^n b \left(\sum_{j=1}^p x_{ij} \beta_{ij} \right) + \sum_{i=1}^n c(y_i, \phi).$$

Logo pelo teorema da fatorização a estatística $S^* = (S_1^*, \dots, S_p^*)$ é suficiente minimal para β . As ligações que correspondem a estatísticas suficientes são chamadas de ligações canônicas. Os MLGs com essa característica possuem propriedades importantes tais como a concavidade de $L(\beta; y)$ que garante a unicidade da estimativa de máxima verossimilhança (m.v.) de β , quando essa estimativa existe. Outra consequência é que os resultados assintóticos são derivados mais facilmente. As ligações canônicas para os modelos normal, Poisson, binomial, gama e normal inversa são dadas, respectivamente, por

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

Algumas ligações usuais são:

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

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

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

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

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

Os MLGs podem ser ajustados pelos aplicativos GLIM (Payne, 1986) e S-Plus (Chambers e Hastie, 1992). Mais detalhes sobre como ajustar MLGs no GLIM e no S-Plus são encontrados em Aitkin et al (1990) e Chambers e Hastie (1992), respectivamente.

3.1.4 Função desvio

Sem perda de generalidade, suponha que o log da função de verossimilhança seja agora definido por

$$L(\mu; y) = \sum_{i=1}^n L(\mu; y_i),$$

onde $\mu_i = g^{-1}(\eta_i)$ e $\eta_i = x_i^t \beta$. Para o modelo saturado ($p = n$) a função $L(\mu; y)$ é estimada por

$$L(y; y) = \sum_{i=1}^n L(y; y_i).$$

Temos que a estimativa de m.v. de μ_i fica nesse caso dada por $\mu_i^s = y_i$. Quando $p < n$, denotaremos a estimativa de $L(\mu; y)$ por $L(\hat{\mu}; y)$. Aqui, a estimativa de m.v. será dada por $\hat{\mu}_i = g^{-1}(\hat{\eta}_i)$, onde $\hat{\eta}_i = x_i^t \hat{\beta}$. A qualidade do ajuste do MLG é avaliada através da função desvio

$$D^*(y; \hat{\mu}) = \phi D(y; \hat{\mu}) = 2\{L(y; y) - L(\hat{\mu}; y)\}.$$

Se denotarmos $\hat{\theta}_i = \hat{\theta}_i(\hat{\mu}_i)$ e $\theta_i^s = \theta_i(\mu_i^s)$, as estimativas de m.v. de θ para os modelos com p parâmetros ($p < n$) e saturado ($p = n$), respectivamente, temos que a função $D(y; \hat{\mu})$ fica dada por

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

Apresentamos abaixo a expressão da função desvio para alguns casos particulares.

Normal

Neste caso, temos $\theta_i = \mu_i$ então $\theta_i^s = y_i$ e $\hat{\theta}_i = \mu_i$. A função desvio fica definida por

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

Poisson

Temos aqui $\theta_i = \log \mu_i$, então $\theta_i^s = \log y_i$ e $\hat{\theta}_i = \log \hat{\mu}_i$. Assim,

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

Binomial

No caso binomial temos $\theta_i^s = \log\{y_i/(n_i - y_i)\}$ para $0 < y_i < n_i$ e $\theta_i^s = 0$ em caso contrário. Analogamente, $\hat{\theta}_i = \log\{\hat{\mu}_i/(1 - \hat{\mu}_i)\}$ para $0 < y_i < n_i$ e $\hat{\theta}_i = \log(1 - \hat{\mu}_i)$ para $y_i = n_i$ e $y_i=0$, respectivamente. A função desvio fica dada por

$$D(y; \hat{\mu}) = 2 \sum_{i=1}^n D_i(y; \hat{\mu}) \text{ com}$$

$$D_i(y; \hat{\mu}) = \begin{cases} -n_i \log(1 - \hat{\mu}_i) & , \text{ se } y_i = 0 \\ -n_i \log(\hat{\mu}_i) & , \text{ se } y_i = n_i \\ D^1(y; \hat{\mu}) & , \text{ caso contrário,} \end{cases}$$

onde $D^1(y; \hat{\mu}) = y_i \log(y_i/n_i \hat{\mu}_i) + (n_i - y_i) \log[(1 - y_i/n_i)/(1 - \hat{\mu}_i)]$.

Usualmente compara-se os valores observados da função desvio com os percentis da distribuição qui-quadrado com $n - p$ graus de liberdade. No entanto $D(y; \hat{\mu})$ não segue assintoticamente uma χ_{n-p}^2 . No caso da binomial quando k é fixo e $n_i \rightarrow \infty$ para cada i , $D(y; \hat{\mu})$ segue sob a hipótese de que o modelo é verdadeiro uma χ_{k-p}^2 . Porém, isso não vale quando $n \rightarrow \infty$ e $n_i \mu_i (1 - \mu_i)$ permanece limitado. Para o modelo de Poisson, quando $\mu_i \rightarrow \infty$ para todo i , tem-se que $D(y; \hat{\mu}) \sim \chi_{n-p}^2$. No caso normal, para σ^2 fixo, $D(y; \hat{\mu}) \sim \sigma^2 \chi_{n-p}^2$. Quando $D^*(y; \hat{\mu})$ depende do parâmetro de dispersão ϕ^{-1} , temos o seguinte resultado (Jørgensen, 1987) para a distribuição nula da função desvio :

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

Logo, quando a dispersão é pequena, é razoável comparar os valores observados de $D^*(y; \hat{\mu})$ com os percentis da χ^2_{n-p} . Em particular, para o caso normal linear, temos $D^*(y; \hat{\mu}) \sim \chi^2_{n-p}$ quando $\sigma^2 \rightarrow 0$.

3.1.5 Função de escore, matriz de informação e processo iterativo para os parâmetros

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

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

e

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

onde X é uma matriz $n \times p$ de posto completo cujas linhas serão denotadas por $x_i^t, i = 1, \dots, n$, $W = \text{diag}(w_1, \dots, w_n)$ com

$$w_i = \left(\frac{d\mu_i}{d\eta_i} \right)^2 \frac{1}{V_i},$$

onde $V = \text{diag}(V_1, \dots, V_n)$, $y = [y_1, \dots, y_n]^t$ e $\mu = [\mu_1, \dots, \mu_n]^t$. Para ligações canônicas, as expressões ficam simplificadas $U(\beta) = \phi X^t (y - \mu)$ e $K(\beta) = \phi X^t V X$, respectivamente.

Para obtenção da estimativa de m.v. de β , utilizamos o processo iterativo de Newton-Raphson expandindo a função escore $U(\beta)$ em série de Taylor em torno de um valor inicial $\beta^{(0)}$, tal que

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

onde $U'(\beta)$ é a primeira derivada de $U(\beta)$ com respeito a β . Repetindo-se o procedimento acima, obtém-se o processo iterativo abaixo

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

$m = 0, 1, \dots$. Como a matriz $-U'(\beta)$ pode não ser positiva definida, a aplicação do

método *scoring* de Fisher substituindo a matriz $-U'(\beta)$ pelo correspondente valor esperado, pode ser mais apropriado. Isso resulta no seguinte processo iterativo:

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

$m = 0, 1, \dots$. Pode-se reescrever o processo iterativo acima como um processo iterativo de mínimos quadrados ponderados

$$\beta^{(m+1)} = (X^T W^{(m)} X)^{-1} X^T W^{(m)} z^{(m)}, \quad (3.3)$$

$m = 0, 1, \dots$, onde $z = \eta + W^{-1/2}V^{-1/2}(y - \mu)$. Observe que z faz o papel de uma variável dependente modificada, enquanto que W é uma matriz de pesos que muda a cada passo do procedimento iterativo. Alguns estudos de convergência para o processo (3.3) podem ser encontrados em Wedderburn (1976) que para algumas combinações da distribuição da variável resposta e da função de ligação mostra que o processo converge em um número finito de passos independentemente dos valores iniciais utilizados. É usual iniciar o processo (3.3) com $\eta_i^{(0)} = g(y_i)$ para $i = 1, \dots, n$. Sob condições gerais de regularidade (Sen e Singer, 1993, Cap. 7) mostra-se que $\hat{\beta}$ é um estimador consistente e eficiente de β e que

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

onde

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

sendo $\Sigma(\beta)$ uma matriz positiva definida. Nem sempre é fácil verificar a existência da matriz $\Sigma(\beta)$, às vezes é necessário supor condições de suficiência tais que levem à existência dessa matriz. Como exemplo suponha um MLG com respostas Y_{ij} , $i = 1, \dots, g$ e $j = 1, \dots, n_i$, tais que $E(Y_{ij}) = \mu_{ij}$ e $g(\mu_{ij}) = x_i^t \beta$. Tais condições suficientes são que $n_i/n \rightarrow a_i > 0$ quando $n \rightarrow \infty$ e que $\sum_{i=1}^g x_i x_i^t$ seja de posto completo, onde $n = n_1 + n_2 + \dots + n_g$. Uma importante referência sobre consistência e normalidade assintótica dos estimadores de máxima verossimilhança em MLGs é descrito por Fahrmeir e Kaufmann (1985), que garantem que $K^{-1/2}U(\beta) \xrightarrow{D}$

$N_p(0, I)$ no qual implica em $\frac{1}{\sqrt{n}}U(\beta) \xrightarrow{D} N(0, \Sigma(\beta))$ assegurando a condição (a) do Capítulo 2. Vale ressaltar que a estimação do parâmetro de dispersão ϕ^{-1} , quando é desconhecido, pode ser encontrada em Cordeiro e McCullagh (1991).

3.1.6 Teste de hipóteses

Suponha a partição para o vetor de parâmetros $\beta = (\beta_1^t, \beta_2^t)^t$, onde β_1 e β_2 são vetores q -dimensional e $(p-q)$ -dimensional, respectivamente. Suponha ϕ conhecido. Seja a hipótese de interesse $H_0 : \beta_1 = \beta_1^{(0)} \times H_1 : \beta_1 \neq \beta_1^{(0)}$. Nesse caso, a estatística da razão de verossimilhança fica definida por

$$\xi^R = \phi\{D(y; \mu^{(0)}) - D(y; \hat{\mu})\},$$

onde $\mu^{(0)}$ é a estimativa de m.v. do modelo sob H_0 . A estatística de Wald fica definida por

$$\xi^W = [\hat{\beta}_1 - \beta_1^{(0)}]^t \hat{Var}^{-1}(\hat{\beta}_1) [\hat{\beta}_1 - \beta_1^{(0)}],$$

onde $Var(\hat{\beta}_1) = \phi[X_1^t W^{1/2} \{I - H_2\} W^{1/2} X_1]^{-1}$ com $H_2 = W^{1/2} X_2 (X_2^t W X_2)^{-1} X_2^t W^{1/2}$. O teste de escore é expresso na forma

$$\xi^S = U_1(\hat{\beta}^0)^t \hat{Var}^0(\hat{\beta}_1) U_1(\hat{\beta}^0)$$

onde $U_1(\beta) = \phi X_1^t W^{1/2} V^{-1/2} (y - \mu)$ é a função escore de β_1 , $\hat{\beta}^0$ é a estimativa de m.v. de β sob H_0 , $\hat{Var}^0(\hat{\beta}_1)$ é avaliada em $\hat{\beta}^0$ e X_1 tem dimensão $n \times q$ e provém da partição $X = (X_1, X_2)$, enquanto X_2 tem dimensão $n \times (p-q)$. Sob a hipótese nula, assintoticamente as estatísticas ξ^R , ξ^W , e ξ^S têm distribuição nula χ_q^2 . Quando ϕ é desconhecido, essa aproximação também é válida. Uma alternativa para contornar a estimação de ϕ é usar a estatística F dada abaixo

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

cujas distribuição nula assintótica é uma F com q e $n-p$ graus de liberdade.

3.2 Hipótese em igualdades lineares

Na prática, existem situações em que o interesse é testar hipóteses na forma de igualdades lineares, ou melhor, $H_0 : C\beta = d \times H_1 : C\beta \neq d$, onde C é uma matriz $k \times p$ de posto completo e d é um vetor $k \times 1$ de escalares. Podemos ver que a estimativa de máxima verossimilhança, sob a hipótese alternativa, coincide com a estimativa de máxima verossimilhança $\hat{\beta}$, porém, sob H_0 , é necessário o uso de algum processo iterativo, que pode ser mais complexo. Nyquist (1991) propôs um processo iterativo para encontrar as estimativas de m.v. dos parâmetros restritos por $C\beta = d$ em MLGs. O processo consiste em maximizar o logaritmo da verossimilhança sujeito às restrições $C\beta = d$. Uma das abordagens utilizadas em problemas de otimização é o método da função penalizada (Fiocco e McComick, 1968; Ryan, 1974). Considera-se a função quadrática penalizada por

$$P(\beta, \lambda) = \sum_{i=1}^n \phi\{y_i\theta_i - b(\theta_i)\} + \sum_{i=1}^n c(y_i, \phi) - \frac{1}{2} \sum_{j=1}^k \lambda_j (d_j - C_j^t \beta)^2$$

e procura-se a solução do problema irrestrito $\max_{\beta} P(\beta, \lambda)$ para valores de λ_j , $j = 1, \dots, k$, fixados e positivos. O estimador restrito b^c é definido por

$$b^c = \lim_{\lambda_1, \dots, \lambda_k \rightarrow \infty} b(\lambda),$$

onde $b(\lambda)$ é um estimador irrestrito para cada λ finito e $b(0)$ é igual ao estimador irrestrito de m.v. dos MLGs. Para o cálculo de $b(\lambda)$ é similar a abordagem do problema de estimação irrestrita. Primeiro, diferenciamos $P(\beta, \lambda)$ com respeito aos β_j 's

$$Q_j(\beta, \lambda) = \frac{\partial}{\partial \beta_j} P(\beta, \lambda) = \phi \sum_{i=1}^n \frac{y_i - \mu_i}{v_i^{1/2}} w_i^{1/2} x_{ij} + \sum_{\ell=1}^k C_{\ell j} \lambda_{\ell} (d_{\ell} - C_{\ell}^t \beta), \quad j = 1, \dots, p$$

e a matriz esperada das derivadas de 2ª ordem tem o elemento comum dado por

$$s_{j\ell}(\beta, \lambda) = E\left(-\frac{\partial^2}{\partial \beta_j \partial \beta_{\ell}} P(\beta, \lambda)\right) = \phi \sum_{i=1}^n x_{ij} x_{i\ell} w_i + \sum_{i=1}^k \lambda_i C_{ij} C_{i\ell}, \quad \ell, j = 1, \dots, p,$$

onde C_{ij} é o elemento comum da matriz C . Utilizando o método *scoring* de Fisher obtemos o processo iterativo

$$b(\lambda)^{(m+1)} = b(\lambda)^{(m)} + S^{-1}(b^{(m)}, \lambda) Q(b^{(m)}, \lambda), \quad (3.4)$$

onde $S(b^{(m)}, \lambda)$ é uma matrix $p \times p$ com elementos $s_{jt}(b^{(m)}, \lambda)$ e $Q(b^{(m)}, \lambda)$ é o vetor $p \times 1$ com elementos $Q_j(b^{(m)}, \lambda)$, ambos avaliados nas estimativas preliminares de $b(\lambda)^{(m)}$. Podemos escrever $S(\beta, \lambda) = \phi X^t W X + C^t \Lambda C$ com Λ sendo uma matriz diagonal com λ_j , $j = 1 \dots k$. Sem perda de generalidade vamos supor que ϕ está incluído em W . Então a equação (3.4) pode ser escrita como

$$(X^t W^{(m)} X + C^t \Lambda C) b(\lambda)^{(m+1)} = X^t W^{(m)} z^{(m)} + C^t \Lambda d.$$

Se Λ e $X^t W X$ são aplicações inversíveis do teorema da binomial inversa (Woodbury, 1950) temos

$$\begin{aligned} b(\lambda)^{(m+1)} &= [(X^t W^{(m)} X)^{-1} - (X^t W^{(m)} X)^{-1} C^t \Lambda \{I \\ &\quad + C(X^t W^{(m)} X)^{-1} C^t \Lambda\}^{-1} C(X^t W^{(m)} X)^{-1}] \{X^t W^{(m)} z^{(m)} + C^t \Lambda d\} \\ &= (X^t W^{(m)} X)^{-1} X^t W^{(m)} z^{(m)} + (X^t W^{(m)} X)^{-1} C^t \\ &\quad \times \Lambda \{I + C(X^t W^{(m)} X)^{-1} C^t \Lambda\}^{-1} \{I + C(X^t W^{(m)} X)^{-1} C^t \Lambda\} d \\ &\quad - (X^t W^{(m)} X)^{-1} C^t \Lambda \{I + C(X^t W^{(m)} X)^{-1} C^t \Lambda\}^{-1} C(X^t W^{(m)} X)^{-1} \\ &\quad \times C^t \Lambda d \\ &= (X^t W^{(m)} X)^{-1} X^t W^{(m)} z^{(m)} + (X^t W^{(m)} X)^{-1} C^t \\ &\quad \times \{\Lambda^{-1} + C(X^t W^{(m)} X)^{-1} C^t\}^{-1} \{d - C(X^t W^{(m)} X)^{-1} X^t W^{(m)} z^{(m)}\}. \end{aligned}$$

A $(m+1)$ -ésima aproximação de $b^{c(m+1)}$ da estimativa restrita de b^c é dada por

$$\begin{aligned} b^{c(m+1)} &= \lim_{\lambda_1, \dots, \lambda_k \rightarrow \infty} b(\lambda)^{(m+1)} \\ &= (X^t W^{(m)} X)^{-1} X^t W^{(m)} z^{(m)} + (X^t W^{(m)} X)^{-1} C^t \{C(X^t W^{(m)} X)^{-1} C^t\}^{-1} \\ &\quad \times \{d - C(X^t W^{(m)} X)^{-1} X^t W^{(m)} z^{(m)}\}. \end{aligned}$$

Note que o termo ϕ é cancelado não interferindo no processo de estimação. E mais, $b^{c(m+1)}$ pode ser considerado como uma estimativa de mínimos quadrados ponderados, sendo dado por $\tilde{b}^{(m+1)} = (X^t W^{(m)} X)^{-1} X^t W^{(m)} z^{(m)}$ (com pesos avaliados na estimativa restrita de passos anteriores) mais um termo de correção. O termo adicionado é o produto de dois fatores, uma constante e um sendo a diferença entre

$d \in C\tilde{b}^{(m+1)}$,

$$b^{c(m+1)} = \tilde{b}^{(m+1)} + (X^t W^{(m)} X)^{-1} C^t \{C(X^t W^{(m)} X)^{-1} C^t\}^{-1} (d - C\tilde{b}^{(m+1)}). \quad (3.5)$$

Agora, para valores não nulos em Λ , temos que

$$\begin{aligned} & \lim_{\lambda_1, \dots, \lambda_k \rightarrow \infty} (\phi X^t W X + C^t \Lambda C)^{-1} \\ &= \phi^{-1} (X^t W X)^{-1} [I - C^t \{C(X^t W X)^{-1} C^t\}^{-1} C(X^t W X)^{-1}]. \end{aligned} \quad (3.6)$$

Supondo que as condições de regularidade de Fahrmeir e Kaufmann (1985) são satisfeitas sob as restrições $C\beta = d$, mostra-se que b^c é consistente e assintoticamente normal com matriz de covariância dada pela equação (3.6). Esse algoritmo encontra-se implementado na linguagem S-Plus no Apêndice G. Os testes estatísticos são similares aos do caso irrestrito. O teste da razão de verossimilhança fica aqui expresso por

$$\xi^R = \phi \{D(y; \mu^{(0)}) - D(y; \hat{\mu})\},$$

onde $\mu^{(0)}$ denota a estimativa de m.v. de μ sob $H_0 : C\beta = d$. O teste de escore e Wald ficam, respectivamente, dados por

$$\begin{aligned} \xi^W &= (C\hat{\beta} - d)^t \{\hat{V}ar(C\hat{\beta})\}^{-1} (C\hat{\beta} - d) \\ &= \phi(C\hat{\beta} - d)^t \{C(X^t \hat{W} X)^{-1} C^t\}^{-1} (C\hat{\beta} - d), \\ \xi^S &= [U(\hat{\beta}^0) - U(\hat{\beta})]^t \hat{V}ar^0(\hat{\beta}) [U(\hat{\beta}^0) - U(\hat{\beta})] \\ &= \phi^{-1} U(\hat{\beta}^0)^t (X^t \hat{W}^0 X)^{-1} U(\hat{\beta}^0) \\ &= \phi^{-1} (\hat{z}^0 - \hat{\eta}^0)^t \{\hat{W}^0\}^{1/2} \hat{H}^0 \{\hat{W}^0\}^{1/2} (\hat{z}^0 - \hat{\eta}^0), \end{aligned}$$

onde $H = W^{1/2} X (X^t W X)^{-1} X^t W^{1/2}$ é a matriz de projeção (Cook e Weisberg, 1982; Pregibon, 1981), $\hat{z}^0$, $\hat{\eta}^0$ e $\hat{W}^0$ são avaliados sob a hipótese nula. Sob H_0 e para amostras grandes, as estatísticas ξ^R , ξ^S e ξ^W têm uma distribuição χ^2 com k graus de liberdade. Podemos ver que no caso em que $C\beta = 0$ o modelo pode ser reparametrizado como

$$\eta_i = x_i^t \beta = x_i^t F^{-1} F \beta = x_i^* \beta_1^* + x_i^* \beta_2^*$$

tal que $\beta_2^* = C'\beta$. Isso é obtido fazendo com que as últimas k linhas da matriz F de dimensão $p \times p$ sejam iguais a C e selecionando as $p - k$ primeiras linhas tal que F seja inversível. Então, o teste $H_0 : C\beta = 0$ é equivalente a testar $H_0 : \beta_2^* = 0$.

3.3 Estimação sob desigualdades lineares

Theil e Van de Panne (1960) estudaram o problema de maximização de funções quadráticas sujeito a desigualdades lineares. Se a concavidade da função de verossimilhança for garantida podemos encontrar a estimativa restrita fazendo uma variante do algoritmo da Subseção 3.1.5. Primeiro, verifica-se se há alguma violação do tipo $C_j^t\beta < d_j, j = 1, \dots, k$. Por exemplo, suponha que há apenas uma violação $C_\ell^t\beta < d_\ell$, então a estimativa de m.v., sob $C\beta \geq d$, corresponde a ajustar o MLG sujeito a $C\beta = d$, onde C é o vetor linha de zeros com um na ℓ -ésima posição e $d = d_\ell$. Podendo ser usado tanto o algoritmo (3.3) quanto (3.5). Suponha agora o caso de duas ou mais violações. Deveremos ajustar o MLG sob todas as possíveis combinações de igualdades lineares. Devemos escolher o modelo que fornece o menor valor da função desvio, $D(y; \hat{\mu})$, dentre aqueles que não apresentam violações. Waterman (1977) mostra que na solução do problema de maximização sem restrições, se as componentes $\hat{\gamma}_j < 0$, então na solução do problema de maximização com restrições haverá pelo menos um $\gamma_j = 0, j = 1, \dots, k$, ou melhor, correspondente a eliminar pelo menos uma covariável do modelo. Baseado nisso, podemos então formular $2^k - 1$ problemas de maximização sem restrições com $\gamma_j = 0$ e as demais variáveis livres a variar em todos os subconjuntos não vazios de $\{1, 2, \dots, k\}$. Devemos escolher a estimativa que satisfaz todas as restrições e que maximiza a função objetivo. É importante observar que à medida que o número de violações aumenta, o processo torna-se dispendioso. O problema de encontrar as estimativas de m.v. de MLGs quando alguns ou todos os parâmetros são restritos por desigualdades foi abordado por McDonald e Diamond (1990). Para resolver o problema de maximização foram utilizadas as condições de Kühn-Tucker (K.T.) que são necessárias para a existência do máximo restrito. No contexto do problema de maximização o logaritmo da função de verossimilhança $L(\beta)$ é uma função

côncava sujeito a $\beta_j \geq 0$, $\forall j$. Para o caso de restrições não-negativas as condições de K.T. para o máximo local $\hat{\beta}$ são :

$$\forall j, \text{ ambos } \hat{\beta}_j > 0 \text{ e } \left. \frac{\partial L}{\partial \beta_j} \right|_{\hat{\beta}} = 0$$

$$\text{ou } \hat{\beta}_j = 0 \text{ e } \left. \frac{\partial L}{\partial \beta_j} \right|_{\hat{\beta}} \leq 0.$$

Observe que para os MLGs temos que a derivada parcial do logaritmo da verossimilhança para uma amostra de tamanho n com respeito a β_j é dada por

$$\sum_{i=1}^n \frac{\partial L_i}{\partial \beta_j} = \sum_{i=1}^n \phi(Y_i - \mu_i) \frac{\partial \theta_i}{\partial \eta_i} x_{ij}.$$

As condições de K.T. para os MLGs com ligação canônica (não-canônica) são o produto interno (ponderado) entre a j -ésima coluna da matriz modelo e o vetor dos resíduos ordinários. Esse produto interno é zero quando o j -ésimo parâmetro estimado é positivo e não-positivo quando o j -ésimo parâmetro estimado é zero (com cada caso ponderado por $\partial \theta / \partial \eta$). Para restrições não-negativas, as condições de K.T. são suficientes para um (estritamente) máximo local se o logaritmo da verossimilhança é uma função (estritamente) côncava. Wedderburn (1976) apresenta várias combinações entre função de ligação e de variância em que essa condição é garantida. Podemos ver que, se a estimativa irrestrito de m.v. viola as restrições de não negatividade, ou seja, ocorre uma violação na estimativa de β então pelo menos um $\hat{\beta}_j$ ($j = 1, \dots, p$) deve ser zero, desde que o logaritmo da verossimilhança seja estritamente côncava o que implica na solução restrita. Essa situação corresponde a ajustar os submodelos retirando um ou mais regressores do modelo completo em todas as combinações possíveis e buscar dentre aqueles que não ocorre violação o que fornece o maior valor para a função objetivo. McDonald e Diamond (1983) apresentam vários algoritmos de busca.

Outro algoritmo foi proposto por Fahrmeir e Klinger (1994), no qual o método de estimação de m.v. sob a restrição $C\beta \geq d$ é a solução de

$$\max_{\beta: C\beta \geq d} L(\beta, y). \quad (3.7)$$

Denotamos o estimador de m.v. sob as restrições de igualdades $C\beta = d$ por $\hat{\beta}^0$. Seja a função lagrangiana

$$\mathcal{L}(\beta, y) = L(\beta, y) - \lambda^t(C\beta - d) = L(\beta, y) - \sum_{j=1}^k \lambda_j(C_j^t\beta - d_j), \quad (3.8)$$

com o vetor de multiplicadores de Lagrange $\lambda = (\lambda_1, \dots, \lambda_k)^t \geq 0$, C_j^t a j -ésima linha de C e d_j a j -ésima componente de d . As condições suficientes para que $\tilde{\beta}$ seja máximo local são

- (i) $C_j^t\tilde{\beta} = d_j$ para $j \in I \subseteq \{1, \dots, k\}$, $C_j^t\tilde{\beta} > d_j$ para todo $j \notin I$, isto é, $\tilde{\beta}$ é um ponto admissível.
- (ii) Existe um $\tilde{\lambda} = (\tilde{\lambda}_1, \dots, \tilde{\lambda}_k)^t \geq 0$ com $\partial\mathcal{L}(\tilde{\beta}, y)/\partial\beta = 0$, isto é, $\tilde{\beta}$ é um ponto estacionário.
- (iii) $M^t(\partial^2 L(\tilde{\beta}, y)/\partial\beta\partial\beta^t)M < 0$ para todo $M \neq 0$ e $M \in \{M : C_j^tM = 0, j \in I, \tilde{\lambda}_j \geq 0 \text{ e } C_j^tM > 0, j \in I, \tilde{\lambda}_j = 0\}$.

Observe que McDonald e Diamond (1990) formularam essas condições para verossimilhanças côncavas e restrições não-negativas. Eles pesquisaram todas as combinações possíveis para que $\tilde{\beta}$ satisfizesse (i) – (ii) e ajustaram os correspondentes submodelos com $\beta_j = 0$ para $j \in I \subseteq \{1, \dots, k\}$. Para altas dimensões, esse método pode acarretar um alto custo computacional e uma alternativa é buscar métodos numéricos de otimização que sejam mais rápidos, por exemplo, Gill, Murray e Wright (1981). O algoritmo proposto por Fahrmeir e Klinger (1994) é o SQP - (programação quadrática sequencial) o qual eles garantem é mais vantajoso uma vez que tem boas propriedades de convergência e é de fácil implementação. O método SQP é bastante discutido por Powell (1978) e Schittkowski (1981). Esses algoritmos estão implementados na NAG e IMSL (bibliotecas do Fortran).

O algoritmo iterativo SQP consiste em gerar uma sequência de estimativas $\beta^{(m)}$, $m = 0, 1, \dots$, que solucionam subproblemas quadráticos expandindo-se a função de verossimilhança em série de Taylor até o termo de segunda ordem, e que converge para $\tilde{\beta}$. Seja $\beta^{(m)}$ o valor de β na m -ésima iteração, o $\beta^{(m+1)}$ é calculado

maximizando-se

$$Q(\beta; \beta^{(m)}) = L(\beta^{(m)}, y) + U^t(\beta^{(m)})(\beta - \beta^{(m)}) - \frac{1}{2}(\beta - \beta^{(m)})^t K(\beta^{(m)})(\beta - \beta^{(m)}) \quad (3.9)$$

sob a restrição $C\beta \geq d$, avaliado em $\beta^{(m)}$. Observe que o lado direito de (3.9) é uma aproximação quadrática de $L(\beta, y)$ em torno de $\beta^{(m)}$. A maximização irrestrita de (3.9) equivale a um dos passos do método de *scoring* de Fisher para calcular $\hat{\beta}$. No caso em que o problema quadrático (3.9) for avaliado sob restrições de igualdades, podemos usar algoritmos mais conhecidos como o método de restrições ativas (Fahrmeir e Klinger, 1994), isto é, aquelas avaliadas nas restrições de igualdades, são pesquisadas e solucionam o problema sob restrições de desigualdades. Outros algoritmos são propostos por Wollan e Dykstra (1987) e Hildreth (1957) em que a solução é baseada no problema dual de (3.9). Esse método é de fácil implementação e apresenta poucos problemas computacionais em altas dimensões dos parâmetros e é de baixo custo computacional. Em resumo, o algoritmo SQP para resolver (3.7) é dado por :

- (1) faça $m = 0$ e $\beta^{(0)} = \hat{\beta}$. Se $C\hat{\beta} \geq d$ então pare e $\tilde{\beta} = \hat{\beta}$. Caso contrário, vá para o passo (2);
- (2) calcule o valor máximo do problema quadrático (3.9) sob a restrição $C\beta^{(m)} \geq d$ por um dos métodos relacionados acima e denote a solução por $\beta^{(m+1)}$. Vá para o passo (3);
- (3) se $\|\beta^{(m+1)} - \beta^{(m)}\| \leq \epsilon$ então pare. Caso contrário, faça $m = m + 1$ e vá para o passo (2).

A convergência do método SQP avaliado em $\beta^{(m)}$ para um estimador de m.v. local restrito, $\tilde{\beta}$, é garantida sob algumas condições (Kredler, 1993) :

- (i) a função objetivo é contínua e duas vezes diferenciável ;
- (ii) existência e unicidade local do estimador de m.v. $\hat{\beta}$;
- (iii) a função log-verossimilhança é côncava para algum conjunto $D \subset R^p$ contendo $\tilde{\beta}$, e $D \cap \{\beta : C\beta \geq 0\}$ é não-vazio.

Essas condições são verificadas por uma ampla classe de MLGs, em particular para todos os MGLs com ligação canônica. Desde que $\beta^{(0)} = \hat{\beta}$ em (1), nós temos que $U(\beta^{(0)}) = 0$, tal que resolver (3.9) para $\beta^{(1)}$ no passo (2) é equivalente a resolver

$$\min_{C\beta \geq d} \{(\hat{\beta} - \beta)^t K(\hat{\beta})(\hat{\beta} - \beta)\}.$$

Piegorsch (1990) propôs um algoritmo para encontrar estimativas de m.v. restritas para o caso de restrições não-negativas, que é equivalente ao algoritmo SQP após a primeira iteração.

3.4 Teste de hipótese em desigualdades lineares

Similarmente ao Capítulo 2, vamos agora estudar a distribuição nula assintótica dos testes usuais, ξ^R , ξ^W e ξ^S , no caso em que $H_0 : C\beta = d \times H_1 : C\beta \geq d$, com pelos menos uma desigualdade estrita em H_1 . Como mencionado na Seção 3.1.5 as condições de Gouriéroux são verificadas para os MLGs com função de verossimilhança côncava e sabendo que $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{D} N_p(0, \Sigma^{-1}(\beta))$, onde $\Sigma = \Sigma(\beta)$ é uma matriz definida positiva. Logo, temos que $\sqrt{n}(C\hat{\beta} - C\beta) \xrightarrow{D} N_k(0, \Delta(\beta))$, onde $\Delta(\beta) = C\Sigma^{-1}C^t$. Para n suficientemente grande, temos o seguinte modelo :

$$\hat{\gamma} = \gamma + \nu, \quad \nu \sim N_k(0, \Delta), \quad (3.10)$$

onde $\gamma = \sqrt{n}(C\beta - d)$ e $\Delta = \Delta(\beta)$ é suposto avaliado no verdadeiro valor do parâmetro. Como mostrado no Capítulo 2, testar as hipóteses $H_0 : \gamma = 0 \times H_1 : \gamma \geq 0$, com pelo menos um componente de γ positivo em H_1 , é assintoticamente equivalente a testar $H_0 : C\beta = d \times H_1 : C\beta \geq d$, com pelos uma desigualdade estrita em H_1 no MLG sob estudo. A estatística de Wald para o modelo (3.10) fica expressa na forma

$$\begin{aligned} \xi^W &= \hat{\gamma} Var^{-1}(\hat{\gamma}) \hat{\gamma} \\ &= \phi(C\hat{\beta} - d)^t \{C(X^t W X)^{-1} C^t\}^{-1} (C\hat{\beta} - d). \end{aligned} \quad (3.11)$$

Quando W é conhecida temos pelo Lema 2.1 que a distribuição nula assintótica de ξ^W é dada por

$$\lim_{n \rightarrow \infty} \text{Prob}\{\xi^W \geq c\} = \sum_{\ell=0}^k \omega_{\ell} \text{Prob}\{\chi_{\ell}^2 \geq c\}, \quad (3.12)$$

onde $c \geq 0$ e ω_{ℓ} são definidos como na Seção 2.6 e dependem de Δ . Comumente, temos W não conhecida e dependente de β , logo se substituirmos W por uma estimativa consistente, a distribuição nula (3.12) não se altera. Note que a distribuição nula assintótica de ξ^W não é mais única pois há uma dependência funcional de Δ em β . A estatística da razão de verossimilhança para testar $H_0 : \gamma = 0 \times H_1 : \gamma \geq 0$, com pelo menos um componente de γ positivo em H_1 , no caso de MLG, fica dada por

$$\xi^R = \phi\{D(y; \mu^{(0)}) - D(y; \tilde{\mu})\},$$

onde $D(y; \mu^{(0)})$ é o desvio sob a hipótese nula e $D(y; \tilde{\mu})$ é o desvio sob a alternativa.

A estatística escore é expressa na forma

$$\xi^S = \phi^{-1}[U(\hat{\beta}^0) - U(\tilde{\beta})]^t (X^t \hat{W}^0 X)^{-1} [U(\hat{\beta}^0) - U(\tilde{\beta})]$$

e a Wald

$$\xi^W = \phi(C\tilde{\beta} - d)^t \{C(X^t \tilde{W} X)^{-1} C^t\}^{-1} (C\tilde{\beta} - d).$$

No caso de $H_0 : C\beta \geq d \times H_1 : \mathbb{R}^p - H_0$, a matriz $\Delta(\beta) = C\Sigma^{-1}C^t$ depende de β nas duas hipóteses. Logo, aplicamos o lema de Wolak mencionado na Subseção 2.7.1. A estatística da razão de verossimilhança, escore e Wald para testar $H_0 \times H_1$ ficam definidas, respectivamente, por

$$\xi^R = \phi\{D(y; \tilde{\mu}) - D(y; \hat{\mu})\},$$

$$\xi^S = \phi^{-1}\{U(\tilde{\beta})\}^t (X^t \tilde{W} X)^{-1} \{U(\tilde{\beta})\} \text{ e}$$

$$\xi^W = \phi(C\hat{\beta} - C\tilde{\beta})^t \{C(X^t \tilde{W} X)^{-1} C^t\}^{-1} (C\hat{\beta} - C\tilde{\beta}),$$

onde $\tilde{W}$ é avaliado em $\tilde{\beta}$. Fahrmeir e Klinger (1994) mostraram através de simulações que a situação menos favorável é em geral alcançada quando $m = k$. Porém,

existe a possibilidade de que a situação menos favorável aconteça para valores de $m < k$. Wolak (1989b) mostra que, se o parâmetro verdadeiro pertencer à fronteira do espaço paramétrico sob H_0 , a distribuição é sempre alcançada para $m = k$. Um outro resultado importante, verificado por Fahrmeir e Klinger, é que quando o tamanho da amostra $n \geq 50$, a diferença entre a probabilidade assintótica de rejeição para a estatística ξ^R , para um c fixo, e a probabilidade empírica de rejeição, obtida através de simulações, é geralmente desprezível.

Com o intuito de aplicar o lema de Wolak, considere um MLG com $\eta = \beta_1 + \beta_2 x_2 + \beta_3 x_3$. Defina $H_0 : \beta \geq [1, 1, 1]^t \times H_1 : \mathbb{R}^3 - H_0$. Conforme a notação da Subseção 2.7.1 temos que $E = \{[1, 1, 1]^t\}$; o conjunto onde está a situação menos favorável é dado por $B = \{[1, 1, 1]^t\} \cup B_1 \cup B_2 \cup B_3$, onde

$$\begin{aligned} B_1 &= \{\beta_1 > 1, \beta_2 = 1, \beta_3 = 1\}, \\ B_2 &= \{\beta_1 = 1, \beta_2 > 1, \beta_3 = 1\} \text{ e} \\ B_3 &= \{\beta_1 = 1, \beta_2 = 1, \beta_3 > 1\}. \end{aligned}$$

A distribuição nula menos favorável deve ocorrer para $m = 2$ ou $m = 3$. Se ocorrer para $m = 2$, deve-se percorrer os três subconjuntos B_1 , B_2 e B_3 e verificar qual deles contém o conjunto menos favorável.

3.5 Aplicações

3.5.1 Casamentos consanguíneos

Como foi apresentado no Capítulo 1, McDonald e Diamond (1983) propuseram modelos logísticos para explicar a chance de abortos naturais entre pais com algum grau de consaguinidade. Um modelo logístico restrito foi proposto para explicar a proporção de abortos naturais. Temos as seguintes variáveis explicativas dicotômicas (sim =1, não =0) :

- (i) $C2 \equiv$ primos de 2º grau;
- (ii) $C12 \equiv$ primos de 1º grau;
- (iii) $C1 \equiv$ primos de 1º grau;

(iv) $INT \equiv$ vive no distrito intermediário e

(v) $URB \equiv$ vive no distrito urbano.

O modelo adotado é

$$\log\{\pi/(1-\pi)\} = \beta_1 + \beta_2 C2 + \beta_3 C12 + \beta_4 C1 + \beta_5 INT + \beta_6 URB \quad (3.13)$$

sujeito a $C\beta \geq 0$, onde π é a proporção de abortos naturais, e

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \end{bmatrix}.$$

Observando as estimativas irrestritas dos parâmetros e seus desvios padrões assintóticos na Tabela 3.2, podemos notar que há uma violação nas estimativas de m.v. irrestritas em $\hat{\beta}_3 > \hat{\beta}_4$. Então, de Theil e Van de Panne (1960) e McDonald e Diamond (1990), segue que a estimativa restrita de m.v. deverá ser obtida após o ajuste do modelo (3.13) restrito por $C\beta = 0$, onde $C = [0 \ 0 \ -1 \ 1 \ 0 \ 0]$. O

Tabela 3.2 *Modelo logístico para a proporção de abortos naturais*

Efeito	Estimativas	
	Irrestrita	Restrita
Constante	-3,6466 (0,1690)	-3,6512 (0,1689)
C2	0,1525 (0,2731)	0,1529 (0,2731)
C12	0,5978 (0,2689)	0,4543 (0,1676)
C1	0,4019 (0,1874)	0,4543 (0,1676)
INT	-0,0099 (0,1825)	-0,0044 (0,1823)
URB	-0,3869 (0,2713)	-0,3777 (0,2708)
Desvio	9,041 (6 g.l.)	9,473 (6 g.l.)

desvio correspondente ao ajuste do modelo final restrito foi de $D(y; \hat{\mu}) = 9.4734$ (6 g.l.), indicando um ajuste adequado. Pelas estimativas dos parâmetros nota-se que a chance de aborto natural parece crescer com o grau de consangüinidade.

Análises de diagnóstico que explicam a violação encontrada são discutidas em Paula (1993). Com o intuito de testar se nessa população a chance de aborto natural

cresce com o grau de consangüinidade. formulamos as hipóteses :

$$H_0 : \beta_2 = \beta_3 = \beta_4 = 0$$

$$H_1 : \beta_4 \geq \beta_3 \geq \beta_2 \geq 0. \text{ com pelo menos uma desigualdade estrita em } H_1.$$

O valor observado da estatística da razão de verossimilhança foi de $\xi^R = 7,036$.

Após calcular $\hat{\Delta}(\hat{\beta}^0) = C\hat{Var}^0(\hat{\beta})C'$ e os respectivos pesos, chega-se ao seguinte nível descritivo:

$$\begin{aligned} P &= \sum_{\ell=1}^3 \omega_{\ell} Pr\{\xi^R \geq 7,036\} \\ &= 0,4840 \times Pr\{\xi^R \geq 7,036\} + 0,1649 \times Pr\{\xi^R \geq 7,036\} \\ &\quad + 0,0160 \times Pr\{\xi_R \geq 7,036\} < 0,01. \end{aligned}$$

Tabela 3.3 *Estatísticas do teste de $H_0 : C\beta = 0 \times H_1 : \{C\beta \geq 0\} - H_0$*

Estatística	Valor observado	Nível descritivo
ξ^R	7,036	0,0099
ξ^W	7,344	0,0085
ξ^S	7,451	0,0080

Portanto, podemos concluir que a chance de aborto natural cresce com o grau de consangüinidade entre os pais ao nível de significância de 1%. Isso é reforçado pelos resultados obtidos para as demais estatísticas decritas na Tabela 3.3.

3.5.2 Estudo de corte de trabalhadores expostos ao arsênico

Na análise de mortes por câncer respiratório dentre trabalhadores expostos ao arsênico, Breslow et al. (1983) propuseram o uso do seguinte modelo de regressão de Poisson :

$$\log E(O_i) = \log(E_i) + x_i^t \beta, \quad (3.14)$$

onde O_i e E_i são, respectivamente, a observação e o valor esperado do número de mortes de câncer respiratório para o i -ésimo subcorte, $i = 1, \dots, 40$. O vetor de parâmetros denota a naturalidade (U.S. ou estrangeiro), cinco níveis de exposição ao arsênico moderado e quatro níveis de exposição ao arsênico pesado. Os dados encontram-se no Apêndice E. Supomos que O_i são variáveis aleatórias independentes com distribuição de Poisson com valores médios $E(O_i)$. Temos portanto um modelo linear generalizado com erro Poisson, função de ligação $\log$ e *offset* $\log(E_i)$.

McDonald e Diamond (1990) argumentam que as taxas de mortes para cada exposição deve formar uma sequência não-decrescente. Isso implica que o modelo (3.14) deve ser ajustado sujeito às restrições do tipo $C\beta \geq 0$, onde

$$C = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}.$$

Pela Tabela 3.4 vimos que existem três violações com as estimativas irrestritas de m.v., $\hat{\beta}_3 < 0$, $\hat{\beta}_4 > \hat{\beta}_5$ e $\hat{\beta}_7 > \hat{\beta}_8$. Aplicando o método de estimação restrita definido por McDonald e Diamond (1983), ajustamos o modelo (3.14) com *offset* $\log(E_i)$ sujeito a $C\beta = 0$, onde

$$C = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \end{bmatrix}.$$

O objetivo aqui é testar a homogeneidade das taxas de morte contra a alternativa

Tabela 3.4 *Modelo de Poisson para o estudo dos metalúrgicos*

Efeito	Estimativas	
	Irrestrita	Restrita
Constante	0,5301 (0,1429)	0,4912 (0,1340)
Nascido-estrangeiro	0,7392 (0,1756)	0,7326 (0,1743)
Anos de exposição ao arsênico moderado		
< 1	-0,2638 (0,2926)	0,0000 (0,0000)
1-4	0,4930 (0,2629)	0,4215 (0,2172)
5-14	0,2133 (0,3394)	0,4215 (0,2172)
15 +	0,8900 (0,2434)	0,9380 (0,2395)
Anos de exposição ao arsênico pesado		
< 1	0,4592 (0,2950)	0,3563 (0,2551)
1-4	0,1843 (0,4582)	0,3563 (0,2551)
5 +	1,1515 (0,3173)	1,1445 (0,3169)
Desvio	30,359 (31 g.l.)	32,033 (31 g.l.)

de que a mesma cresce com o nível de arsênico (moderado ou pesado). Podemos testar os dois casos em separado. Assim, teremos as seguintes hipóteses :

$$H_0 : \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$$

$$H_1 : C\beta \geq 0 \text{ com pelo menos uma desigualdade estrita}$$

e

$$H_0 : \beta_7 = \beta_8 = \beta_9 = 0$$

$$H_1 : C\beta \geq 0 \text{ com pelo menos uma desigualdade estrita,}$$

onde

$$C = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

e

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix},$$

respectivamente.

O valor da estatística ξ^R para o arsênico moderado foi de 14,62 e após calcularmos $\hat{\Delta}(\hat{\beta}^0) = CVar^0(\hat{\beta})C^t$ e os respectivos pesos, chegamos ao seguinte nível descritivo :

$$\begin{aligned} P &= \sum_{i=1}^3 \omega_i Prob\{\xi^R \geq 14,62\} \\ &= 0,4388 \times Prob\{\xi^R \geq 14,62\} + 0,3021 \times Prob\{\xi^R \geq 14,62\} \\ &\quad + 0,0612 \times Prob\{\xi^R \geq 14,62\} < 0,0005. \end{aligned}$$

Tabela 3.5 *Estatísticas do teste para o arsênico moderado*

Estatística	Valor observado	Nível descritivo
ξ^R	14,615	0.0004
ξ^W	16,427	0.0001
ξ^S	17,323	0.0001

O valor da estatística ξ^R para o arsênico pesado foi de 10,96 e após calcularmos $\hat{\Delta}(\hat{\beta}^0) = C\hat{V}ar^0(\hat{\beta})C^t$ e os respectivos pesos, chegamos ao seguinte nível descritivo:

$$\begin{aligned} P &= \sum_{i=1}^4 \omega_i Prob\{\xi^R \geq 10,96\} \\ &= 0,4687 \times Prob\{\xi^R \geq 10,96\} + 0,2219 \times Prob\{\xi^R \geq 10,96\} \\ &\quad + 0,0313 \times Prob\{\xi^R \geq 10,96\} < 0,002. \end{aligned}$$

Tabela 3.6 *Estatísticas do teste para o arsênico pesado*

Estatística	Valor observado	Nível descritivo
ξ^R	10,956	0,0017
ξ^W	14,039	0.0003
ξ^S	15,471	0.0001

As Tabelas 3.5 e 3.6 apresentam um resumo para as três estatísticas correspondentes aos casos de arsênico moderado e pesado. Os níveis descritivos para as

estatísticas ξ^W e ξ^S confirmam o resultado verificado para ξ^R de que a taxa de mortes cresce com o tempo de exposição ao arsênico (moderado e pesado).

3.5.3 Estudo de processo infeccioso pulmonar

No estudo de caso-controle realizado no Setor de Anatomia e Patologia do Hospital Heliópolis, em São Paulo, no período de 1970 a 1982 (Paula e Tudor, 1986), um total de 175 pacientes com processo infeccioso pulmonar foi classificado segundo as seguintes variáveis:

- (i) Resp, tipo de tumor (1: maligno, 0: benigno);
- (ii) Idade, idade em anos;
- (iii) Sexo (1: masculino, 0: feminino);
- (iv) Hl, intensidade da célula histiócitos-linfócitos (1: ausente, 2: discreta, 3: moderada, 4: intensa) e
- (v) Ff, intensidade da célula fibrose-frouxa (1: ausente, 2: discreta, 3: moderada, 4: intensa).

As informações referentes às variáveis Resp, Hl e Ff foram obtidas após biópsia realizada na região pleural de cada paciente ou por autópsia no caso de óbito. Esses dados são descritos no Apêndice B.

Considere o modelo logístico-linear apenas com os efeitos principais

$$\Pr\{\text{Resp} = 1 \mid \eta\} = \{1 + \exp(-\eta)\}^{-1}, \quad (3.15)$$

onde $\eta = \beta_1 + \beta_2 \text{Idade} + \beta_3 \text{Sexo} + \sum_{i=1}^4 \beta_{4i} \text{Hl}_i + \sum_{i=1}^4 \beta_{5i} \text{Ff}_i$, com Hl_i e Ff_i sendo variáveis binárias correspondentes aos níveis de Hl e Ff, respectivamente. É assumido que $\beta_{41} = \beta_{51} = 0$. Vale salientar que devido ao fato da amostragem ter sido retrospectiva, o uso do modelo acima para fazer previsões somente é válido corrigindo-se a estimativa da constante, β_1 (vide, por exemplo, McCullagh e Nelder, 1989, pg.113).

Observações médicas indicam que é muito razoável supor que a chance de tumor maligno no nível i de Hl ou Ff é pelo menos igual à chance no nível $i + 1$. Logo,

pode ser de interesse testar as seguintes hipóteses:

$$H_0 : \beta_{42} = \beta_{43} = \beta_{44} = 0$$

$$H_1 : \beta_{44} \leq \beta_{43} \leq \beta_{42} \leq 0$$

e

$$H_0 : \beta_{52} = \beta_{53} = \beta_{54} = 0$$

$$H_1 : \beta_{54} \leq \beta_{53} \leq \beta_{52} \leq 0,$$

com pelo menos uma desigualdade estrita em H_1 . Ou melhor, verificar se existe evidências de que há pelo menos dois grupos de intensidade em cada caso com chances diferentes de tumor maligno. Podemos testar as hipóteses descritas acima em separado. Logo, a matriz C fica dada por

$$C = \begin{bmatrix} 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

e

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \end{bmatrix},$$

respectivamente. Como há três restrições em H_1 , as expressões para cálculo dos pesos $\omega_t(3, \hat{\Delta})$'s, onde $\hat{\Delta}(\hat{\beta}^0) = C(X^T \hat{V}^0 X)^{-1} C^T$, tornam-se mais complexas que para o caso de duas restrições descrito na Subseção 2.6.1 (vide Apêndice A). A Tabela 3.7 apresenta as estimativas de máxima verossimilhança (desvio padrão assintótico entre parênteses) dos parâmetros do modelo (3.15). Podemos notar pelas estimativas irrestritas, que a violação ocorre entre as estimativas correspondentes aos níveis Ff(3) e Ff(4). Utilizando o algoritmo descrito na Seção 3.3, obtemos o modelo restrito onde o desvio correspondente foi de $D(y; \hat{\mu}) = 158,98$ (166 g.l.), indicando um ajuste adequado. Observando as estimativas dos parâmetros nota-se que a chance de tumor maligno (com relação a benigno) parece crescer com o aumento da idade e que as mulheres são mais pré-dispostas a desenvolverem um processo infeccioso maligno do que os homens. Paula (1995) verifica através de

métodos de diagnóstico que dois pacientes com um perfil atípico são altamente influentes na violação observada entre as estimativas dos níveis Ff(3) e Ff(4). Com relação aos níveis dos fatores Hl e Ff, o valor da estatística da razão de verossimilhança, para testar H_0 contra H_1 , foi de $\xi^R = 15,10$ para o caso Hl e $\xi^R = 3,57$ para o caso Ff.

Tabela 3.7 *Estimativas irrestrita e restrita dos parâmetros para o modelo (3.15)*

Efeito	Estimativa	
	Irrestrita	Restrita
Constante	-1.850 (1,060)	-1,845 (1,060)
Idade	0.065 (0,013)	0,065 (0,013)
Sexo	0.784 (0,469)	0,778 (0,469)
Hl(2)	-0.869 (0,946)	-0,891 (0,947)
Hl(3)	-2.249 (0,970)	-2,210 (0,970)
Hl(4)	-3.294 (1,458)	-3,345 (1,484)
Ff(2)	-0.687 (0,502)	-0,690 (0,502)
Ff(3)	-1.025 (0,526)	-0,874 (0,506)
Ff(4)	0.431 (1,123)	-0,874 (0,506)
Desvio	157,40 (166 g.l.)	158,98 (166 g.l.)

Após computar-se $\hat{\Delta}(\hat{\beta}^0) = C(X^T \hat{V}^0 X)^{-1} C^T$ e os respectivos pesos, chega-se aos seguintes níveis descritivos:

$$\begin{aligned}
 P &= \sum_{\ell=1}^3 \omega_{\ell} \text{Prob}\{\xi^R \geq 15,10\} \\
 &= 0,409 \times \text{Prob}\{\xi^R \geq 15,10\} + 0,353 \times \text{Prob}\{\xi^R \geq 15,10\} \\
 &\quad + 0,091 \times \text{Prob}\{\xi^R \geq 15,10\} < 0,0004 \quad \text{e} \\
 P &= \sum_{\ell=1}^3 \omega_{\ell} \text{Pr}\{\xi^R \geq 3,57\} \\
 &= 0,435 \times \text{Prob}\{\xi^R \geq 3,57\} + 0,301 \times \text{Prob}\{\xi^R \geq 3,57\} \\
 &\quad + 0,064 \times \text{Prob}\{\xi^R \geq 3,57\} < 0,0962,
 \end{aligned}$$

respectivamente. Logo, pode-se afirmar que há pelo menos dois tipos de intensidade para a célula Hl com chances diferentes de tumor maligno. Vimos que essas chances

Tabela 3.8 *Estatísticas do teste de para a célula Hl*

Estatística	Valor observado	Nível descritivo
ξ^R	15,105	0,0004
ξ^W	13,426	0,0009
ξ^S	14,903	0,0004

Tabela 3.9 *Estatísticas do teste para a célula Ff*

Estatística	Valor observado	Nível descritivo
ξ^R	3,573	0,0961
ξ^W	3,574	0,0962
ξ^S	3,633	0,0933

estão em ordem não-decrescente. Para o tipo de célula Ff, não há evidências fortes de diferenças entre as chances de tumor maligno segundo o grau de intensidade. As Tabelas 3.8 e 3.9 apresentam os valores das três estatísticas ξ^R , ξ^W e ξ^S e os correspondentes níveis descritivos para os dois casos sob estudo. Nota-se uma proximidade de valores das estatísticas e uma mesma conclusão sendo induzida.

CAPÍTULO 4

Alguns Casos Particulares

4.1 Introdução

Nesta seção vamos discutir alguns casos particulares nos quais a estatística do teste e também os pesos tomam formas mais simples no sentido que os mesmos não dependam das estimativas dos parâmetros. Não existindo a dependência funcional dos pesos com relação às estimativas dos parâmetros, no caso de testarmos a hipótese $H_0 : C\beta \geq d \times H_1 : \mathbb{R}^p - H_0$, não precisaremos ir em busca da situação menos favorável pois ela é atingida quando todas as restrições estiverem na forma de igualdades.

4.2 Modelo de análise de variância

Suponha o modelo de análise de variância balanceado com um fator e dois grupos

$$g(\mu_{ij}) = \alpha + \beta_i,$$

onde $i = 1, 2, j = 1, \dots, m, \beta_1 = 0$ e $\beta_2 = \beta$. Considere as hipóteses : $H_0 : \beta = 0 \times H_1 : \beta \geq 0$. Temos que $X = [X_1, X_2]$, onde X_2 é um vetor $2m \times 1$ de 1's enquanto X_1 é um vetor $2m \times 1$ com 0's nas m primeiras posições e 1's nas m restantes. Logo,

$$X^t \hat{W} X = m \tilde{w}_2 \begin{bmatrix} \tilde{w}_1/\tilde{w}_2 & 1 \\ 1 & 1 \end{bmatrix}$$

e conseqüentemente a estatística de Wald fica dada por

$$\xi^W = \frac{\phi m \tilde{w}_1 \tilde{w}_2}{(\tilde{w}_1 + \tilde{w}_2)} \tilde{\beta}^2,$$

onde $\tilde{\beta}$ denota a estimativa de m.v. de β sob o modelo restrito. Similarmente, para a estatística de escore obtemos

$$\xi^S = \frac{\phi m}{2W(\bar{y})} \left[\sum_{j=1}^2 \left(\sqrt{\tilde{w}_j/\tilde{v}_j} (\bar{y}_j - \hat{\mu}_j) + \sqrt{W(\bar{y})/V(\bar{y})} (\bar{y} - \bar{y}_j) \right)^2 \right],$$

onde $\bar{y}_1$ e $\bar{y}_2$ são as médias amostrais correspondentes aos dois grupos e $V(\bar{y})$ e $W(\bar{y})$ são a função de variância e a função peso sob a hipótese nula, respectivamente

$$\xi^R = \phi\{D(y; \mu^{(0)}) - D(y; \tilde{\mu})\},$$

onde $D(y; \mu^{(0)})$ é o desvio sob a hipótese nula e $D(y; \tilde{\mu})$ é o desvio sob a alternativa. É importante observar que a distribuição nula assintótica de ξ^R é dada por

$$\lim_{n \rightarrow \infty} \text{Prob}\{\xi^R \geq c\} = \frac{1}{2} \text{Prob}\{\chi_1^2 \geq c\}, \quad c > 0. \quad (4.1)$$

4.3 Modelo de regressão linear

Suponha o modelo

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

considere as hipóteses $H_0 : \beta = 0 \times H_1 : \beta > 0$. Logo, temos que

$$X^t \hat{W} X = m \hat{w}_2 \begin{bmatrix} \sum_{i=1}^n \hat{w}_i & \sum_{i=1}^n \hat{w}_i x_i \\ \sum_{i=1}^n \hat{w}_i x_i & \sum_{i=1}^n \hat{w}_i x_i^2 \end{bmatrix}$$

e conseqüentemente a estatística de Wald fica dada por

$$\xi^W = \phi \tilde{\beta}^2 \left[\sum_{i=1}^n \hat{w}_i x_i^2 - \frac{\left(\sum_{i=1}^n \hat{w}_i x_i \right)^2}{\sum_{i=1}^n \hat{w}_i} \right].$$

A estatística de escore toma forma

$$\xi^S = \phi^{-1} [U(\hat{\beta}^0) - U(\tilde{\beta})]^t (X^t \hat{W}^0 X)^{-1} [U(\hat{\beta}^0) - U(\tilde{\beta})]$$

e estatística da razão de verossimilhança fica dada por

$$\xi^R = \phi\{D(y; \mu^{(0)}) - D(y; \tilde{\mu})\},$$

onde $D(y; \mu^{(0)})$ é o desvio sob a hipótese nula e $D(y; \tilde{\mu})$ é o desvio sob a alternativa.

A distribuição nula assintótica é também dada por (4.1).

4.4 Ordem simples

Suponha agora Y_{ij} variáveis aleatórias mutuamente independentes tais que $E(Y_{ij}) = \mu_i$ e $Var(Y_{ij}) = \phi^{-1}V_i$, $i = 1, \dots, k$ e $j = 1, \dots, n_i$. Desejamos testar $H_0 : \mu_1 = \dots = \mu_k$ contra $H_1 : \mu_1 \leq \dots \leq \mu_k$, com pelo menos uma desigualdade estrita em H_1 . Mostra-se que a estatística da razão de verossimilhança fica agora dada por

$$\xi^R = \phi\{D(y; \mu^{(0)}) - D(y; \tilde{\mu})\},$$

onde $\mu^{(0)} = (\bar{y}, \dots, \bar{y})^t$, $\bar{y} = \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}/n$, $n = n_1 + \dots + n_k$ e $\tilde{\mu} = (\tilde{\mu}_1, \dots, \tilde{\mu}_k)^t$ é a regressão isotônica de $\bar{y}_i$ com pesos n_i , $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$, $i = 1, \dots, k$. Isso quer dizer que o algoritmo PAVA (Apêndice H), utilizado para a obtenção das estimativas restritas sob H_1 , estendido para a família exponencial (Robertson, Wright e Dykstra, 1988, Cap. 4) pode ser aplicado para a obtenção de $\tilde{\mu}_1, \dots, \tilde{\mu}_k$. Após algumas manipulações algébricas, mostra-se que as estatísticas de escore e de Wald para testar H_0 contra H_1 ficam dadas por

$$\xi^S = \frac{\phi}{V(\bar{y})} \sum_{i=1}^k n_i (\tilde{\mu}_i - \bar{y})^2$$

e

$$\xi^W = \tilde{\mu}^t C^t \{C \hat{V}ar(\hat{\mu}) C^t\}^{-1} C \tilde{\mu},$$

respectivamente, onde $V(\bar{y})$ denota que a função de variância está sendo avaliada em $\bar{y}$, que é a estimativa de μ sob H_0 , $\hat{V}ar(\hat{\mu}) = \phi^{-1}\{\hat{V}_1^{-1}n_1^{-1}, \dots, \hat{V}_k^{-1}n_k^{-1}\}$ e C é uma matriz de ordem $(k-1) \times k$ dada por

$$C = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 0 & \dots & 0 & 0 \\ & & & \ddots & & \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}.$$

Mostra-se que as três estatísticas são assintoticamente equivalentes com distribuição nula assintótica

$$\lim_{n \rightarrow \infty} Prob\{\xi^R \geq c\} = \sum_{\ell=0}^{k-1} \omega_\ell Prob\{\chi_\ell^2 \geq c\}, \quad (4.2)$$

onde $c \geq 0$ e é assumido que $n_i/n \rightarrow a_i > 0$ quando $n \rightarrow \infty$. Portanto, tem-se assintoticamente uma distribuição qui-quadrado barra $\bar{\chi}_k^2$ (vide Barlow et al., 1972; Robertson et al., 1988) com os coeficientes de correlação linear associados à matriz $\Delta = \phi^{-1}C(X^tWX)^{-1}C^t$ dados por ($\rho_{ii} = 1$)

$$\rho_{i(i+1)} = \rho_{(i+1)i} = - \left[\frac{a_i a_{i+2}}{(a_i + a_{i+1})(a_i + a_{i+2})} \right]^{1/2} \quad (4.3)$$

$i = 1, \dots, k-2$ e $\rho_{ij} = 0$ para $|i-j| \geq 2$. Podemos ver que não há dependência funcional das probabilidades $\omega_\ell(k-1, \Delta)$'s e o vetor de parâmetros sob H_0 . Porém, continua dispendioso computacionalmente o cálculo dos pesos para $k \geq 5$. É importante salientar que para estruturas balanceadas, $n_1 = \dots = n_k$, as probabilidades $\omega_\ell(k-1, \Delta)$'s tomam formas mais simples e de obtenção recursiva (Robertson et al., 1988)

$$\begin{aligned} \omega_s(1, k) &= \frac{1}{k!}, \\ \omega_s(k, k) &= \frac{1}{k!}, \end{aligned}$$

e $\omega_s(\ell, k) = \frac{1}{k} \omega_s(\ell-1, k-1) + \frac{k-1}{k} \omega_s(\ell, k-1)$, $\ell = 2, \dots, k-1$, onde $\omega_s(\ell, k) = \omega_{\ell-1}(k-1, \Delta)$. Siskind (1976) conjecturou que a distribuição nula da estatística $\bar{\chi}_k^2$, sob a hipótese de pesos desiguais, tende a aproximar razoavelmente para o caso de pesos iguais quando os tamanhos amostrais não são muito diferentes. Robertson e Wright (1983) têm confirmado essa conjectura para hipóteses em ordem monótona. Eles mostraram que a aproximação é adequada se a razão $n_{\max}/n_{\min}$ não exceder 3,5.

Outra hipótese de interesse é testar $H_0 : \mu_1 \leq \dots \leq \mu_k$ contra $H_1 : \mathbb{R}^k - H_0$. A estatística da razão de verossimilhança fica aqui dada por

$$\xi^R = \phi\{D(y; \hat{\mu}) - D(y; \hat{\mu})\},$$

onde $\hat{\mu} = (\bar{y}_1, \dots, \bar{y}_k)^t$, cuja distribuição nula menos favorável, assumindo que $n_i/n \rightarrow a_i > 0$ quando $n \rightarrow \infty$, é alcançada quando $\mu_1 = \dots = \mu_k$, sendo dada por

$$\lim_{n \rightarrow \infty} Prob\{\xi^R \geq c\} = \sum_{\ell=0}^{k-2} \omega_\ell Prob\{\chi_{k-\ell-1}^2 \geq c\}, \quad (4.4)$$

onde $c \geq 0$ e os coeficientes de correlação linear associados à matriz Δ são também dadas por (4.3). As estatísticas de escore e de Wald, que são assintoticamente equivalentes à estatística da razão de verossimilhança, tomam, respectivamente, as formas

$$\xi^S = \phi \sum_{i=1}^k n_i (\bar{y}_i - \tilde{\mu}_i)^2 / \hat{V}_i$$

e

$$\xi^W = (\hat{\mu} - \tilde{\mu})^t C^t \{C \hat{Var}(\hat{\mu}) C^t\}^{-1} C (\hat{\mu} - \tilde{\mu}),$$

onde $\hat{Var}(\hat{\mu})$ é aqui avaliado em $\hat{\mu}$. Em Robertson et al. (1988) pode-se encontrar a equivalência assintótica das três estatísticas ξ^R , ξ^W e ξ^S com a qui-quadrado barra, verificada nesta seção para o caso de ordem simples. Dachs e Paula (1988) discutem aproximações para a distribuição qui-quadrado barra para o caso de ordem simples em estudos de seguimento com resposta de Poisson. Peers (1995) apresenta testes alternativos para hipóteses restritas, os quais são equivalentes assintoticamente à estatística ξ^R e assumem formas mais simples do tipo qui-quadrado barra.

4.5 Retas paralelas

Modelos de dose-resposta de retas paralelas têm sido largamente aplicados na área de Farmacologia. Esses modelos são usados para comparar a eficiência de drogas do mesmo tipo (ação similar - vide Finney, 1971, 1978; Collet, 1994; Morgan, 1992). O objetivo principal nesses estudos é comparar as potências entre as drogas, definindo uma determinada droga como nível base ou droga padrão. Podemos generalizar esses modelos supondo que $Y_{ij\ell}$ seja o efeito produzido pela j -ésima dose correspondente à i -ésima droga no ℓ -ésimo indivíduo, $i = 1, \dots, k$, $j = 1, \dots, d_i^*$ e $\ell = 1, \dots, n_{ij}$, tem distribuição pertencente à família exponencial de distribuições com média μ_{ij} definida tal que

$$g(\mu_{ij}) = \alpha_i + \beta \log x_{ij}, \quad (4.5)$$

e que as variáveis $Y_{ij\ell}$'s são mutuamente independentes. Podemos ver que se tomarmos a primeira droga como padrão, a potência δ_i da i -ésima droga com relação à

primeira é definida por

$$\log \delta_i = (\alpha_i - \alpha_1) / \beta,$$

$i = 1, \dots, k$. Essa suposição leva à seguinte relação:

$$g(\mu_{ij}) = \alpha_1 + \beta \log \delta_i x_{ij},$$

isto é, x unidades da droga i têm o mesmo efeito que $\delta_i x$ unidades da primeira droga.

Em alguns casos práticos podemos estar interessados em verificar se as potências das drogas estão restritas segundo alguma ordem, tal como se há um aumento do grau de severidade das drogas à medida que varia os níveis de um fator particular. Pode haver interesse em testar a hipótese de homogeneidade das potências contra essa ordem particular. Ou ainda, assumindo que novas drogas sejam tão eficientes quanto o controle, e o interesse é verificar se pelo menos uma das drogas é melhor do que a controle. Essas situações, dentre outras, podem ser tratadas com a teoria apresentada na Seção 3.4 aplicada ao modelo (4.5). Contudo, é possível algumas simplificações interessantes, que em geral não ocorrem na presença de co-variáveis, se o experimento é conduzido de uma forma balanceada. Suponha, então, as seguintes condições experimentais:

- (i) o mesmo número de doses para todas as drogas, $d_i^* = d^*$;
- (ii) o mesmo número de réplicas para as doses de cada droga, $n_{i1} = \dots = n_{id^*} = n_i$;
- (iii) a mesma dose para o nível j de todas as drogas, $x_{1j} = \dots = x_{kj}$, $j = 1, \dots, d^*$.

Supondo essa estrutura balanceada, que é comum na prática (vide, por exemplo, Morgan, 1992), desejamos testar $H_0 : 1 = \delta_2 = \dots = \delta_k$ contra $H_1 : 1 \leq \delta_2 \leq \dots \leq \delta_k$, com pelo menos uma desigualdade estrita em H_1 . Ou seja, vamos testar a homogeneidade dos interceptos contra a hipótese de ordem monótona para os mesmos. Paula e Sen (1995) mostram que a distribuição nula, nesse caso, é uma qui-quadrado barra dada por (4.2) com coeficientes de correlação linear associados à matriz Δ dados por (4.3). Porém, se não há nenhum indício a priori para as potências, podemos testar $H_0 : 1 \leq \delta_2 \leq \dots \leq \delta_k$ contra $H_1 - H_0$, onde $H_1 : \delta_i \in$

$\mathbb{R}^+$, $\forall i$, no qual mostra-se que a distribuição nula menos favorável é também uma qui-quadrado barra dada agora por (4.4) com coeficientes de correlação linear dados por (4.3). Paula e Sen mostram que essa equivalência assintótica com a distribuição qui-quadrado barra ocorre para uma classe bastante geral de restrições que inclui, em particular, a classe de quase-ordem. Paula (1996) estuda a robustez dessa aproximação quando as suposições (ii) e (iii) são violadas e verifica que a aproximação para a distribuição qui-quadrado barra continua valendo para variações moderadas dessas suposições. Na análise de dados de sobrevivência, Paula e Rojas (1997) estendem esses resultados para os modelos de regressão com distribuição do valor extremo.

4.6 Aplicações

4.6.1 Vírus da Poliomyelitis

Como ilustração considere os dados descritos no Apêndice C sobre um experimento envolvendo uma preparação padrão ($n^{\circ} 1$) do vírus *poliomyelitis* e quatro preparações teste (Finney, 1978, p.441). Cinco ratos machos e cinco fêmeas foram inoculados com cada uma das cinco doses, 2, 3, 4, 5 e 6 (dadas em diluição de $\log_{10}$) da preparação padrão; o mesmo número de animais foi inoculado com cada uma das duas doses de cada preparação teste. A resposta considerada foi o número de dias decorridos até o animal apresentar algum sintoma da doença. Cinco animais que estavam aparentemente bem após dezesseis dias tiveram assinalado um valor arbitrário, 20 dias. Finney discute alguns possíveis métodos para analisar esse conjunto de dados e particularmente sugere o uso de transformações do tipo $Y = \log(N - 0,5)$ e $Y = 1/(N - 0,5)$, onde N denota o número de dias decorridos até a ocorrência da doença.

A idéia básica de modelos lineares generalizados é evitar o uso de transformações para a variável resposta, procurando sempre alguma distribuição na família exponencial que possa representar bem os dados. Nesse sentido, será adotado como resposta $Y = N - 0,5$, essa correção é recomendada uma vez que os dias são contados integralmente mesmo que o animal fique doente no início do dia, e será assumido que

Y segue uma distribuição gama de média μ e parâmetro de dispersão ϕ^{-1} . Denote então por $Y_{ijk\ell}$ o tempo decorrido até o aparecimento da doença para o k -ésimo animal do ℓ -ésimo sexo que recebeu a j -ésima dose da i -ésima preparação, $i = 1, \dots, 5$; $j = 1, \dots, d_i^*$; $k = 1, \dots, 10$ e $\ell = 1, 2$ com $d_1^* = 5$ e $d_2^* = d_3^* = d_4^* = d_5^* = 2$. A ligação utilizada (Paula, 1997) é a log que induz ao seguinte modelo:

$$\log \mu_{ijk\ell} = \alpha_i + \delta_i x_{ij}, \quad (4.6)$$

onde x_{ij} denota a dose no nível (i, j) . Deve-se notar aqui que em geral as doses são dadas em concentração das substâncias, diferente desse exemplo em que as mesmas são dadas em diluição das preparações. Inicialmente este experimento nos leva a pensar que a preparação padrão é pelo menos tão eficiente quanto cada preparação teste (suposição de quase-ordem). Finney sugere o uso do modelo de regressão de retas paralelas. Uma forma de testar a adequacidade desse modelo e a suposição de quase-ordem é formulando a seguinte hipótese :

$$H_0 : \alpha_1 \leq [\alpha_2, \dots, \alpha_5], \delta_1 = \dots = \delta_5$$

contra $H_1 : \{\alpha \in \mathbb{R}^5, \delta \in \mathbb{R}^5\} - H_0$. Podemos reescrever a hipótese H_0 na forma $H_0 : C_1 \alpha \geq 0, C_2 \delta = 0$, onde

$$C_1 = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

$$C_2 = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

e

$$C = \begin{bmatrix} C_1 & 0 \\ 0 & C_2 \end{bmatrix}.$$

É importante notar que sob o modelo gama com ligação log temos que $\hat{\beta} \sim N_p(0, \phi^{-1}(X^t X)^{-1})$, onde a matriz de variância-covariância de $\hat{\beta}$ não depende dos parâmetros. Portanto

$$\Sigma^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix}$$

não depende de β . Logo, um fato importante é que a distribuição limite de ξ^R , para o teste proposto acima, é unicamente determinada e dada aproximadamente por

$$Prob\{\xi^R \geq c\} = \sum_{\ell=1}^4 \omega(4, 4 - \ell, \hat{\Delta}) Prob\{\chi_{\ell}^2 \geq c\}, \quad c \geq 0,$$

onde $\hat{\Delta} = \frac{1}{\phi} \{C_1 K^{11} C_1^t - C_1 K^{12} C_2^t (C_2 K^{22} C_2^t)^{-1} C_2 K^{21} C_1^t\}$ e as matrizes K^{11} , K^{12} , K^{21} e K^{22} são obtidas pela partição da inversa da matriz de informação de Fisher a menos do termo ϕ^{-1} ,

$$K^{-1} = (X^t X)^{-1} = \begin{bmatrix} K^{11} & K^{12} \\ K^{21} & K^{22} \end{bmatrix}.$$

Na Tabela 4.1 encontram-se as estimativas de m.v. dos parâmetros do modelo e a estimativa dos momentos de ϕ (McCullagh e Nelder, pg. 296, 1989), que é consistente.

Nesse caso a estatística do teste da razão de verossimilhança, supondo ϕ desconhecido, toma a forma

$$\xi^R = \hat{\phi} t(\hat{\mu}, y) - \tilde{\phi} t(\tilde{\mu}, y) + 2n\{d(\hat{\phi}) - d(\tilde{\phi})\},$$

onde $t(\mu, y) = -2n - D(\mu, y)$, $d(\phi) = \phi \log \phi - \log \Gamma(\phi)$, $\Gamma(\cdot)$ denota a função gama e $\hat{\phi}$ e $\tilde{\phi}$ são as estimativas de momentos irrestrita e restrita para ϕ , respectivamente.

A estatística Wald e escore são dadas, respectivamente, por

$$\xi^W = \hat{\phi} [C\hat{\beta} - C\tilde{\beta}]^t \{C(X^t X)^{-1} C^t\}^{-1} [C\hat{\beta} - C\tilde{\beta}]$$

e

$$\begin{aligned} \xi^S &= \tilde{\phi} [y - \tilde{\mu}]^t \tilde{V}^{-1/2} X \{X^t X\}^{-1} X^t \tilde{V}^{-1/2} [y - \tilde{\mu}] \\ &= \tilde{r}_p^t H \tilde{r}_p, \end{aligned}$$

Tabela 4.1 *Estimativas irrestrita e restrita dos parâmetros do modelo (4.6)*

Efeito	Estimativas	
	Irrestrita	Restrita
α_1	0,3537 (0,1528)	0,2519 (0,0371)
α_2	0,5176 (0,2546)	0,5451 (0,0420)
α_3	0,2433 (0,2546)	0,2519 (0,0371)
α_4	0,5248 (0,2546)	0,6254 (0,0420)
α_5	0,4524 (0,2546)	0,2519 (0,0371)
δ_1	0,3450 (0,0360)	0,3584 (0,0096)
δ_2	0,3676 (0,0805)	0,3584 (0,0096)
δ_3	0,3561 (0,0805)	0,3584 (0,0096)
δ_4	0,3918 (0,0805)	0,3584 (0,0096)
δ_5	0,2516 (0,0805)	0,3584 (0,0096)
Desvio	12,0229 (120 g.l.)	12,6677 (124 g.l.)
ϕ	7,6571	7,3177

onde $r_p = \phi^{1/2} V^{-1/2}(y - \mu)$ é o resíduo de Pearson, $H = X\{X^t X\}^{-1} X^t$, $V = \text{diag}\{\mu_1^2, \dots, \mu_n^2\}$. Temos então que $\xi^R = 6,663$ e calculando os pesos sob $\Delta(\tilde{\beta})$ obtemos

$$\begin{aligned}
 P &= 0 + 0,372 \times \text{Prob}\{\chi_1^2 \geq 6,663\} + 0,336 \times \text{Prob}\{\chi_2^2 \geq 6,663\} \\
 &\quad + 0,128 \times \text{Prob}\{\chi_3^2 \geq 6,663\} + 0,013 \times \text{Prob}\{\chi_4^2 \geq 6,663\} \\
 &\cong 0,03.
 \end{aligned}$$

Observando a Tabela 4.2 através dos níveis descritivos dos valores observados para as estatísticas ξ^R , ξ^S e ξ^W podemos verificar uma fraca evidência em favor da hipótese nula. Na Figura 4.1, podemos notar que a suposição de retas paralelas é em geral bem razoável. Contudo, a reta da preparação teste 4 tem um comportamento diferente das demais preparações teste que pode ter ocasionado os baixos níveis descritivos dos testes acima. Podemos tentar relaxar a restrição de que a preparação padrão não é menos eficiente que a preparação 4 e manter a hipótese

Tabela 4.2 *Estatística do teste para os dados da Poliomyelitis*

Estatística	Valor observado	nível descritivo
ξ^R	6,663	0.029
ξ^W	5,196	0.057
ξ^S	4,273	0.089

de paralelismo. A hipótese nula fica então dada por :

Figura 4.1 *Gráfico de retas separadas*

$$H_0 : \alpha_1 \leq [\alpha_2, \alpha_3, \alpha_4], \delta_1 = \dots = \delta_5$$

contra $H_1 : \{\alpha \in \mathbb{R}^5, \delta \in \mathbb{R}^5\} - H_0$. Podemos reescrever a hipótese H_0 da seguinte

forma $H_0 : C_1\alpha \geq 0, C_2\delta = 0$, onde

$$C_1 = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \end{bmatrix},$$

$$C_2 = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

e

$$C = \begin{bmatrix} C_1 & 0 \\ 0 & C_2 \end{bmatrix}.$$

Neste caso, temos a distribuição nula assintótica da ξ^R dada por

$$Prob\{\xi^R \geq c\} = \sum_{\ell=1}^3 \omega(3, 3 - \ell, \hat{\Delta}) Prob\{\chi_{\ell}^2 \geq c\}, \quad c \geq 0.$$

Calculando as estimativas sob H_0 , obtemos $\hat{\alpha}_1 = \hat{\alpha}_3 = 0,3165(0,038)$, $\hat{\alpha}_2 = 0,5728(0,041)$, $\hat{\alpha}_4 = 0,6533(0,041)$, $\hat{\alpha}_5 = 0,1641(0,041)$, $\hat{\delta} = 0,3492(0,009)$, $\hat{\phi} = 7,5256$ com desvio de 12,333 com 124 graus de liberdade e mais a estatística $\xi^R = 3,061$ e os pesos sob $\Delta(\hat{\beta})$ obtemos

$$\begin{aligned} P &= 0 + 0,436 \times Prob\{\chi_1^2 \geq 3,061\} + 0,293 \times Prob\{\chi_2^2 \geq 3,061\} \\ &\quad + 0,064 \times Prob\{\chi_3^2 \geq 3,061\} \\ &< 0,13. \end{aligned}$$

Tabela 4.3 *Estatística do teste para os dados da Poliomyelitis*

Estatística	Valor observado	nível descritivo
ξ^R	3,061	0,123
ξ^W	2,391	0,173
ξ^S	2,322	0,179

Observando a Tabela 4.3 através dos níveis descritivos dos valores observados para as estatísticas ξ^R , ξ^S e ξ^W , aceitamos a hipótese de que a preparação padrão é mais eficiente que as preparações teste 2,3 e 4 ao nível de significância de 10%.

Podemos também analisar este problema numa outra direção. Pela Tabela 4.2, a um nível de significância de 5% as estatísticas ξ^W e ξ^S não rejeitam H_0 , o que indica que parece ser razoável, segundo esses valores, assumir que as preparações teste não são mais potentes que a preparação padrão no sentido de levar o animal à doença num menor tempo médio possível. Isso induz às restrições $\alpha_1 \leq [\alpha_2, \dots, \alpha_5]$. Pode então ser de interesse testar se há pelo menos uma preparação com um tempo médio, até a ocorrência da doença, maior do que o tempo médio induzido pela preparação padrão.

As hipóteses seriam agora formuladas na forma

$$H_0 : \alpha_1 = \dots = \alpha_5$$

$$H_1 : \alpha_1 \leq [\alpha_2, \dots, \alpha_5],$$

com pelo menos uma desigualdade estrita em H_1 . As estimativas sob a hipótese nula são $\hat{\alpha}^0 = 0,4512$ (0,1053), $\hat{\delta}^0 = 0,333$ (0,029) e $\hat{\phi}^0 = 5,568$. As estatísticas ξ^R , ξ^W e ξ^S ficam expressas, respectivamente, por

$$\xi^R = \tilde{\phi}t(\hat{\mu}, y) - \hat{\phi}^0 t(\tilde{\mu}, y) + 2n\{d(\tilde{\phi}) - d(\hat{\phi}^0)\},$$

$$\xi^W = \tilde{\phi}(C\tilde{\beta})^t \{C(X^t X)^{-1}C^t\}^{-1}(C\tilde{\beta})$$

e

$$\xi^S = \hat{\phi}^0 [\tilde{r}_p - \hat{r}_p^0]^t H [\tilde{r}_p - \hat{r}_p^0],$$

onde $\hat{\phi}^0$ é o estimador de momentos de ϕ sob H_0 , $\hat{r}_p^0$ denota o resíduo de Pearson

avaliado em $\hat{\mu}^0$ e

$$C = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Obteve-se para a estatística ξ^R o valor de 31,837. Para o cálculo do nível descritivo deve-se antes computar os pesos $\omega_\ell(4, \hat{\Delta})$, $\ell = 0, \dots, 4$, onde $\hat{\Delta}(\hat{\beta}^0) = \frac{1}{\hat{\phi}^0}(X^t X)^{-1}$. Utilizando a fórmula encontrada no Apêndice A para calcular os pesos, obtemos

$$\begin{aligned} P &= 0 + 0,128 \times Prob\{\chi_1^2 \geq 31,837\} + 0,336 \times Prob\{\chi_2^2 \geq 31,837\} \\ &\quad + 0,372 \times Prob\{\chi_3^2 \geq 31,837\} + 0,151 \times Prob\{\chi_4^2 \geq 31,837\} \\ &< 0,0001. \end{aligned}$$

A Tabela 4.4 apresenta os valores das três estatísticas ξ^R , ξ^W e ξ^S e os correspondentes níveis descritivos indicando fortemente pela rejeição da hipótese nula. Conclui-se portanto que a preparação padrão não é menos eficiente que as preparações teste e é mais eficiente que pelo menos uma delas.

Tabela 4.4 *Estatísticas do teste de $H_0 : C\beta = 0 \times H_1 : \{C\beta \geq 0\} - H_0$*

Estatística	Valor observado	Nível descritivo
ξ^R	31,837	0,000
ξ^W	22,073	0,000
ξ^S	19,415	0,000

4.6.2 Mortes de besouros por inseticida

A tabela dada no Apêndice F é resultado de um experimento (Collet, 1994) em que três inseticidas são aplicados num determinado tipo de inseto. Os três inseticidas são *DDT* a uma proporção de 2,0% w/v, γ - *BHC* a uma proporção de 1,5% w/v e uma mistura dos dois. Os insetos em lotes de cinquenta foram expostos as doses dos três inseticidas medidas em mg/10 cm², em doses diferentes, onde foi verificado

o número de insetos mortos após um período de 6 dias para cada dose aplicada. Usaremos como variável explicativa o logaritmo da dose aplicada no modelo de retas paralelas em regressão logística, que pela Figura 4.1 parecer ser razoável. É assumido que Y , o número de besouros mortos para cada dose aplicada, segue uma distribuição binomial com probabilidade de sucesso π . Denote então por Y_{ij} o número de besouros mortos após o período de 6 dias para o i -ésimo inseticida referente à j -ésima dose aplicada, $i = 1, \dots, 3$ e $j = 1 \dots 6$. A ligação sugerida pelos dados (Collet, 1994) é a *logit* que induz ao seguinte modelo inicial:

$$\text{logit}(\pi_{ij}) = \alpha_i + \delta x_{ij}, \quad (4.7)$$

Figura 4.2 *Retas paralelas da proporção de insetos mortos segundo a dose*

onde x_{ij} denota a dose no nível (j) referente ao i -ésimo inseticida e α_i denota o efeito do i -ésimo inseticida. Os modelos ajustados são :

$$\begin{aligned} DDT : \text{logit}(\hat{\pi}) &= -4,555 + 2,696 \log(\text{dose}) \\ &\quad (0,361) \quad (0,215) \\ \gamma - BHC : \text{logit}(\hat{\pi}) &= -3,842 + 2,696 \log(\text{dose}) \\ &\quad (0,332) \quad (0,215) \\ DDT + \gamma - BHC : \text{logit}(\hat{\pi}) &= -1,425 + 2,696 \log(\text{dose}) \\ &\quad (0,285) \quad (0,215) \end{aligned}$$

com o desvio de 21,282 com 14 graus de liberdade. Parece ser razoável assumir que o inseticida composto pela mixtura é pelo menos tão eficiente quanto os outros dois inseticidas. Isso induz às restrições $\alpha_3 \geq [\alpha_1, \alpha_2]$. Então, podemos pensar em testar se a mixtura é mais potente que pelo menos um inseticida (sinergismo, vide Paula, 1997). As hipóteses seriam formuladas na forma

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3$$

$$H_1 : \alpha_3 \geq [\alpha_1, \alpha_2],$$

com pelos menos uma desigualdade estrita em H_1 . Como não há violação com as estimativas irrestritas, as estimativas restritas são iguais às irrestritas. A matriz C fica expressa na forma

$$C = \begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \end{bmatrix}.$$

Utilizando a fórmula dada no Apêndice A para calcular os pesos, obtemos

$$\begin{aligned} P &= 0 + 0,5 \times Pr\{\chi_1^2 \geq 225,55\} + 0,333 \times Pr\{\chi_2^2 \geq 225,55\} \\ &= 0,000, \end{aligned}$$

indicando fortemente pela rejeição da hipótese nula.

Tabela 4.5 *Estatísticas do teste para o modelo de retas paralelas*

Estatística	Valor observado	Nível descritivo
ξ^R	225,55	0,000
ξ^W	160,08	0,000
ξ^S	200,74	0,000

A Tabela 4.4 apresenta também os valores das estatísticas ξ^W e ξ^S com os respectivos níveis descritivos, reforçando a hipótese de sinergismo.

Conclusões

Em resumo, nesta dissertação mostramos a equivalência assintótica, sob H_0 , entre as estatísticas do teste da razão de verossimilhança Wald, Hausman-Wald, Kühn-Tücker e escore para testar $H_0 : C\theta = 0 \times H_1 : \{C\theta \geq 0\} - H_0$. Comentamos sobre métodos de estimação para verossimilhanças côncavas que é garantida em modelos lineares generalizados para função de ligação canônica e outras (Wedderburn, 1976). Através de exemplos, ilustramos a metodologia descrita. Acreditamos que o interesse de procurar testes mais potentes para inferência estatística levou os pesquisadores a desenvolverem seus trabalhos nesta área iniciada por Bartholomew. Várias linhas de pesquisa podem ser ainda tratadas, tais como:

- (i) encontrar expressões mais simples para as estatísticas do teste de hipóteses;
- (ii) encontrar casos particulares em que os pesos não dependam dos parâmetros;
- (iii) encontrar aproximações para os pesos;
- (iv) desenvolver algoritmos para o cálculo dos pesos para mais de 10 restrições;
- (v) estudo do poder das estatísticas através de simulações entre outros.

Concluindo, esta dissertação é um esforço inicial para apresentar alguns tópicos nesta área de pesquisa e divulgar sua utilidade.

APÊNDICE A

Probabilidades de Nível

A.1 Caso de $k = 3$ restrições

Para o caso de $k = 3$ restrições os pesos ficam dados por (vide, por exemplo, Wolak, 1987)

$$\omega_0(3, \Delta) = \frac{1}{2} - \omega_2(3, \Delta);$$

$$\omega_1(3, \Delta) = \frac{1}{2} - \omega_3(3, \Delta);$$

$$\omega_2(3, \Delta) = \frac{1}{4}\pi^{-1}\{3\pi - \cos^{-1}(\rho_{12.3}) - \cos^{-1}(\rho_{13.2}) - \cos^{-1}(\rho_{23.1})\} \text{ e}$$

$$\omega_3(3, \Delta) = \frac{1}{4}\pi^{-1}\{2\pi - \cos^{-1}(\rho_{12}) - \cos^{-1}(\rho_{13}) - \cos^{-1}(\rho_{23})\},$$

onde ρ_{ij} denota o elemento (i, j) da matriz de correlações lineares associadas à matriz Δ , enquanto $\rho_{ij.t}$'s correspondem aos coeficientes de correlação linear parcial, os quais são definidos por

$$\rho_{ij.t} = \frac{\rho_{ij} - \rho_{it}\rho_{jt}}{\sqrt{(1 - \rho_{it}^2)(1 - \rho_{jt}^2)}}.$$

A.2 Caso de $k = 4$ restrições

No caso de $k = 4$ restrições temos as expressões abaixo para os pesos (vide, por exemplo, Wolak 1987)

$$\omega_0(4, \Delta) = \frac{1}{2} - \omega_4(4, \Delta) - \omega_2(4, \Delta);$$

$$\begin{aligned}
\omega_1(4, \Delta) &= -\frac{1}{2} + \frac{1}{8\pi} \sum_{i>j; i, j \neq k} \cos^{-1}(\rho_{ij,k}); \\
\omega_2(4, \Delta) &= \frac{1}{4\pi^2} \sum_{i>j, k>\ell; \ell \neq i, j} \cos^{-1}(\rho_{ij}) \{ \pi - \cos^{-1}(\rho_{k\ell, ij}) \}; \\
\omega_3(4, \Delta) &= \frac{1}{2} - \omega_1(4, \Delta) \text{ e} \\
\omega_4(4, \Delta) &= \frac{1}{16} + \frac{1}{8\pi} \{ \sin^{-1}(\rho_{12}) + \sin^{-1}(\rho_{13}) + \sin^{-1}(\rho_{14}) \\
&\quad + \sin^{-1}(\rho_{23}) + \sin^{-1}(\rho_{24}) + \sin^{-1}(\rho_{34}) \} + \frac{1}{4\pi^2} \eta,
\end{aligned}$$

onde η (vide, Childs, 1967; Sun, 1988a) é dado por

$$\eta = \sum_{k=2}^4 \frac{\rho_{1k}}{\sqrt{1 - \rho_{1k}^2}} I_2(R_2^{1,k}),$$

com $I_2(R_2^{1,2})$, $I_2(R_2^{1,3})$ e $I_2(R_2^{1,4})$ sendo integrais no intervalo (0,1) dadas abaixo

$$\begin{aligned}
I_2(R_2^{1,2}) &= \int_0^1 \sin^{-1} \left(\frac{r_{34}^{12}(t)}{\sqrt{r_{33}^{12}(t)r_{44}^{12}(t)}} \right) dt; \\
I_2(R_2^{1,3}) &= \int_0^1 \sin^{-1} \left(\frac{r_{24}^{13}(t)}{\sqrt{r_{22}^{13}(t)r_{44}^{13}(t)}} \right) dt \text{ e} \\
I_2(R_2^{1,4}) &= \int_0^1 \sin^{-1} \left(\frac{r_{23}^{14}(t)}{\sqrt{r_{22}^{14}(t)r_{33}^{14}(t)}} \right) dt,
\end{aligned}$$

onde

$$\begin{aligned}
r_{ij}^{1k}(t) &= \{ \rho_{ij} - \rho_{ki}\rho_{kj} - t^2(\rho_{1k}^2\rho_{ij} + \rho_{1i}\rho_{1j} \\
&\quad - \rho_{1k}\rho_{1i}\rho_{kj} - \rho_{1k}\rho_{1j}\rho_{ki}) \},
\end{aligned}$$

$i, j, k = 1, 2, 3$. A correlação parcial $\rho_{k\ell, ij}$ é definida abaixo

$$\rho_{k\ell, ij} = \frac{\rho_{k\ell, i} - \rho_{kj, i}\rho_{\ell j, i}}{\sqrt{(1 - \rho_{kj, i}^2)(1 - \rho_{\ell j, i}^2)}}.$$

APÊNDICE B

Processo Infeccioso Pulmonar

Os dados abaixo estão na ordem:
Id, Tumor, Idade, Sexo, HL e FF

1	0	26	1	3	1	2	0	21	1	3	1
3	0	45	1	3	3	4	0	19	2	4	3
5	0	16	2	4	3	6	0	72	2	4	3
7	0	53	1	3	1	8	0	33	1	4	2
9	0	39	1	3	2	10	0	41	1	3	2
11	0	26	2	3	3	12	0	27	2	3	3
13	0	46	1	3	1	14	0	27	1	3	3
15	0	65	1	3	2	16	0	27	1	4	1
17	0	32	2	4	1	18	0	22	2	3	2
19	0	23	1	4	2	20	0	42	2	4	2
21	0	82	2	3	2	22	0	23	2	3	1
23	0	55	1	3	2	24	0	43	1	3	1
25	0	49	1	3	1	26	0	20	2	2	1
27	0	23	1	4	2	28	0	28	1	3	2
29	0	34	1	3	1	30	0	18	1	3	1
31	0	22	1	3	2	32	0	50	2	3	3
33	0	64	1	2	1	34	0	29	2	4	2
35	0	24	2	3	3	36	0	50	1	2	3

Tabela B.1 *Estudo de câncer no processo infeccioso pulmonar*

37	0	38	1	2	3	38	0	20	2	3	3
39	0	44	1	3	3	40	0	59	1	3	3
41	0	43	1	3	3	42	0	27	1	4	3
43	0	20	1	2	1	44	0	24	1	3	2
45	0	46	1	3	4	46	0	40	1	2	3
47	0	21	2	3	3	48	0	21	1	3	2
49	0	42	1	3	3	50	0	23	1	3	3
51	0	38	2	3	4	52	0	53	1	3	3
53	0	53	1	3	2	54	0	21	1	1	1
55	0	57	1	3	1	56	0	63	2	3	1
57	0	21	1	3	2	58	0	45	1	3	2
59	0	77	1	2	1	60	0	58	1	3	2
61	0	28	1	3	2	62	0	83	1	3	1
63	0	22	1	2	1	64	0	36	1	2	1
65	0	43	1	3	2	66	0	22	1	3	3
67	0	30	1	3	3	68	0	46	1	2	1
69	0	78	2	3	3	70	0	23	1	3	3
71	0	56	1	2	1	72	0	56	1	3	3
73	0	44	1	2	1	74	0	64	1	2	3
75	0	18	2	3	3	76	0	23	1	2	2
77	0	62	1	1	1	78	0	53	1	2	2
79	0	23	1	3	1	80	0	23	1	3	1
81	0	49	1	3	2	82	0	21	1	3	3
83	0	17	1	3	3	84	0	41	2	3	3
85	0	45	1	3	3	86	0	51	1	2	3
87	0	62	2	3	1	88	0	48	1	2	2
89	0	27	1	3	2	90	0	18	2	3	2
91	0	67	1	3	3	92	0	75	1	2	3
93	0	67	1	2	2	94	0	49	1	2	1
95	0	63	1	3	2	96	0	87	1	2	1
97	0	53	2	2	1	98	0	18	1	2	3
99	0	30	1	4	3	100	0	48	1	3	3
101	0	31	2	3	3	102	0	56	1	2	2
103	0	48	1	3	2	104	0	33	2	4	2
105	1	58	1	3	4	106	1	76	1	3	3

107	1	64	1	2	3	108	1	44	1	2	2
109	1	34	1	2	1	110	1	51	1	3	3
111	1	60	1	3	3	112	1	73	1	4	2
113	1	72	1	3	4	114	1	62	1	2	1
115	1	60	2	2	1	116	1	43	1	1	1
117	1	62	1	2	1	118	1	55	1	2	1
119	1	58	2	2	1	120	1	45	1	2	2
121	1	58	1	2	1	122	1	15	1	2	1
123	1	61	2	2	1	124	1	60	1	2	1
125	1	61	2	3	2	126	1	56	2	2	1
127	1	78	1	2	2	128	1	21	1	2	1
129	1	75	1	2	1	130	1	56	2	2	1
131	1	73	2	2	1	132	1	56	1	2	2
133	1	62	2	3	1	134	1	56	1	2	1
135	1	56	2	3	1	136	1	52	2	2	1
137	1	57	2	2	2	138	1	29	2	1	1
139	1	51	1	2	2	140	1	77	1	3	3
141	1	40	2	3	2	142	1	65	1	2	1
143	1	60	1	2	2	144	1	69	2	3	1
145	1	67	1	3	3	146	1	58	1	2	1
147	1	72	1	3	1	148	1	51	2	2	1
149	1	57	1	3	1	150	1	36	2	2	1
151	1	58	1	3	3	152	1	59	1	3	1
153	1	59	1	1	1	154	1	57	2	2	1
155	1	73	1	2	1	156	1	69	1	2	3
157	1	61	1	3	1	158	1	67	1	2	1
159	1	70	1	2	2	160	1	64	1	2	2
161	1	69	1	3	3	162	1	52	1	2	1
163	1	59	1	3	1	164	1	50	2	2	1
165	1	48	1	2	1	166	1	49	1	1	1
167	1	78	1	2	1	168	1	66	1	2	1
169	1	74	1	2	1	170	1	50	1	1	2
171	1	75	2	1	2	172	1	55	2	3	3
173	1	50	2	2	1	174	1	57	2	3	1
175	1	70	2	2	1						

APÊNDICE C

Vírus da Poliomyelitis

Os dados abaixo estão
na ordem: Prep, Dose,
Sexo e Tempo

1	2	M	3	3	3	3	4
		F	3	3	3	3	4
	3	M	3	4	4	5	6
		F	4	4	5	6	8
	4	M	4	5	5	5	6
		F	4	6	6	6	8
	5	M	6	6	7	14	20
		F	6	6	6	9	16
	6	M	7	9	10	11	20
		F	7	8	9	9	20
2	2	M	3	4	5	5	6
		F	3	3	3	3	5
	4	M	5	6	7	8	9
		F	5	5	6	7	20
3	2	M	3	3	3	3	4
		F	3	3	3	3	3
	4	M	4	6	6	6	7
		F	5	6	6	6	6

Tabela C.1 *Dados correspondentes ao vírus da Poliomyelitis*

4	2	M	3	4	5	5	5
		F	4	4	4	4	4
	4	M	6	6	7	8	20
		F	6	6	7	10	10
5	2	M	3	3	3	3	3
		F	3	3	3	3	4
	4	M	4	4	5	6	6
		F	4	4	5	5	5

APÊNDICE D

Estudo da Relação de Abortos com Casamentos Consangüíneos

Tabela D.1 *Dados referentes ao número de abortos em 6358 casos em Shizuoka City no Japão*

Residência	Consangüinidade	Gravidez	Abortos
Distrito	Nenhuma relação	958	27
	primos 2º	160	1
	primos 1º	65	3
Rural	primos 1º	293	12
Distrito	Nenhuma relação	2670	67
	primos 2º	338	11
	primos 1º	237	11
Intermediário	primos 1º	654	23
Distrito	Nenhuma relação	543	7
	primos 2º	70	4
	primos 1º	110	3
Urbano	primos 1º	260	7

APÊNDICE E

Estudo de Câncer Respiratório em Metalúrgicos

E.1 Valores observados

Tabela E.1 *Número observado de mortes para metalúrgicos expostos ao arsênico*

Anos de exposição ao arsênico					
moderado	pesado				
	0	<1	1-4	5 +	total
Nascido em U.S.					
0	28	2	3	6	39
< 1	7	2	1	2	12
1-4	8	4	1	1	15
5-14	4	0	0	0	4
15 +	4	1	0	0	5
total	51	9	5	9	74
Estrangeiro					
0	33	1	0	2	36
< 1	2	0	0	0	2
1-4	4	0	0	0	4
5-14	6	0	0	0	6
15 +	16	3	0	0	19
total	61	4	0	2	67

E.2 Valores esperados

Tabela E.2 *Número esperado de mortes para metalúrgicos expostos ao arsênico*

Anos de exposição ao arsênico					
moderado	pesado				
	0	<1	1-4	5 +	total
Nascido em U.S.					
0	20,86	1,21	0,56	0,60	23,23
< 1	4,91	0,76	0,20	0,29	6,16
1-4	3,10	0,33	0,10	0,11	3,64
5-14	1,58	0,12	0,08	0,01	1,79
15 +	1,14	0,11	0,05	0,03	1,33
total	31,59	2,53	0,99	1,04	36,15
Estrangeiro					
0	7,34	0,39	0,11	0,28	8,12
< 1	1,31	0,10	0,02	0,05	1,48
1-4	0,91	0,01	0,07	0,04	1,03
5-14	1,05	0,02	0,13	0,04	1,24
15 +	1,60	0,20	0,10	0,01	1,91
total	12,21	0,72	0,43	0,42	13,78

APÊNDICE F

**Estudo de Toxidade de Inseticidas em Insetos da
Farinha**

Tabela F.1 *Toxidade de inseticidas nos besouros da farinha*

Dose	DDT	γ -BHC	DDT+(γ -BHC)
2.00	3/50	2/50	28/50
2.64	5/49	14/49	37/50
3.48	19/47	20/50	46/50
4.59	19/50	27/50	48/50
6.06	24/49	41/50	48/50
8.00	35/50	40/50	50/50

APÊNDICE G

Implementação do Algoritmo no S-Plus

A função `'glm.rest'` executa o ajuste de um modelo linear generalizado com restrições nos parâmetros na forma $C\beta = d$. A metodologia utilizada foi descrita por Nyquist (1991). Para usá-la devemos proceder o ajuste do MLG através do comando `'glm'` do S-Plus após a implantação da subrotina abaixo. Deveremos usar o comando `'glm'` com opção `method='glm.rest'`, `Cres` é a matriz de contrastes das restrições, `sol` é o vetor solução. Como ilustração temos:

```
glm(Y~X,family=gaussian,method='glm.rest', Cres=Cres, sol=sol)
```

Os objetos disponíveis após o ajuste são os mesmos gerados pelo comando `glm` (Chambers e Hastie, Cap.6) e mais

- (i) `'coefficients'` $\hat{\beta}$
- (ii) `'residuals'` resíduo tipo `'working'` (Chambers e Hastie, Cap 6)
- (iii) `'fitted.values'` $\hat{\mu}$
- (iv) `'standard.error'` erro padrão condicional
- (v) `'cov.cond.unsc'` matriz de covariância condicional não escalonada
- (vi) `'weights'` $\tilde{w}$
- (vii) `'linear.predictors'` $\tilde{\eta}$
- (viii) `'deviance'` desvio restrito
- (ix) `'dispersion'` $\hat{\phi}^{-1}$.

```

glm.rest <—
function(x=X,y=Y,w = rep(1, length(x[, 1])), start = NULL, offset = 0,
family= gaussian(), maxit=maxit, epsilon = 0.001, trace = F,
null.dev = NULL, qr = F, ...)
{
pass <— 1
X <— x
Y <— y
n <— nrow(X)
p <— ncol(X)
Fam <— as.family(family)
if(any(offset) && dimnames(X)[[2]][1]==‘(Intercept)’&&p==1)
{
deviance <— list(deviance=glm.fit(X[,‘(Intercept)’], drop = F], Y, w,
offset = offset, family = family, maxit = maxit, epsilon = epsilon,
null.dev= NULL)$deviance )
} else
{
fr <— glm.fit(x=X, y=Y, w = w, start = NULL,offset = offset, family= Fam,
maxit = pass, epsilon = epsilon,trace = F, null.dev = T, qr = qr, ...)
we <— fr$weights
W <— diag(as.vector(we))
betai <— matrix(fr$coef,p,1)
const <— solve(t(X)%*%W%*%X)%*%t(Cres)%*% solve(Cres%*%
solve(t(X)%*%W%*%X) %*%t(Cres))%*%(sol-Cres%*%betai)
beta <— betai+const
k <— 0
if(trace) cat(‘GLMREST linear loop’, k, ‘\n’, sep = ‘’, ‘: coef’,
format(round(as.vector(beta), 4))
,‘\n’, sep = ‘’)

```

```

while(any(abs(beta-betai)>epsilon) &&(k>maxit)){
neta <- X%%beta+offset
fr <- glm.fit(x=X, y=Y, w = w, start = neta, offset = offset,family= Fam,
maxit = pass, epsilon = epsilon, trace = F, null.dev = NULL, qr = qr, ...)
betat <- matrix(fr$coef,p,1)
we <- fr$weights
W <- diag(as.vector(we))
betai <- beta
const <- solve(t(X)%*%W%*%X)%*%t(Cres)%*% solve(Cres%*%
solve(t(X)%*%W%*%X)%*%t(Cres))%*%(sol-Cres%*%betat)
beta <- betat+const
k <- k+1
if(trace) cat('GLMREST linear loop', k, '\n', sep = '', ': coef',
format(round(as.vector(beta), 4))
, '\n', sep = '')
}
if(maxit==k)
warning(paste('linear convergence not obtained in', k,'iterations.'))
neta <- X%%beta+offset
mu <- Fam$inverse(neta)
df.residual <- fr$df.residual
if (Fam$family[[1]]=='Binomial') {
if(is.matrix(Y)) {
if(dim(Y)[2] > 2)
stop('only binomial response matrices (2 columns)')
n <- drop(Y%%c(1, 1))
y <- Y[,1]
}
else {
if(is.category(Y))

```

```

y <- Y != levels(Y)[1]
else y <- as.vector(Y)
n <- rep(1, length(Y))
}
y <- y/n
w <- w*n
}
we <- eval(Fam$weight.local=T)
if(!any(is.na(mu))) {
  devr <- Fam$deviance(mu, y, w)
}
else {
  devr <- NA
}
if(nrow(X) > p)
  phi <- devr/nrow(X)
else phi <- NA
famname <- Fam$family['name']
if(is.null(famname))
  famname <- 'Gaussian'
dispersion <- switch(famname,
  Poisson = 1,
  Binomial = 1,
  Gamma = (sum(((y-mu)/mu)**2))/df.residual,
  phi)
names(dispersion) <- famname
cov <- solve(t(X)%*%W%*%X)
cov <- cov%*%(diag(1,ncol(X))-(t(Cres)%*%
  solve(Cres%*%cov%*%t(Cres))%*%Cres%*%cov))
se <- sqrt(diag(cov))*dispersion

```

```
coefs <- as.vector(beta)
lp <- as.vector(neta)-offset
fv <- as.vector(mu)
wei <- as.vector(we)
work <- as.vector((y-mu)*Fam$deriv(mu))
this.call <- match.call()
y <- as.vector(y)
null.dev <- fr>null.deviance
dn <- labels(x)
xn <- dn[[2]]
yn <- dn[[1]]
names(coefs) <- xn
names(work) <- yn
names(fv) <- yn
names(lp) <- yn
names(wei) <- yn
names(se) <- xn
dimnames(cov) <- list(xn, xn)
if(length(attributes(w)) | any(w != w[1])) fit$prior.weights <- w
fit <- list(coefficients = coefs,residuals=work,fitted.values=fv,
standard.error=se,cov.cond.unsc=cov,rank = fr$rank,
assign = attr(X,'assign'),df.residual=df.residual, weights=wei)
if(length(attributes(w)) | any(w != w[1])) fit$prior.weights <- w
if(fr$rank < p) {
  if(df.residual > 0)
    fit$assign.residual <- fr$assign.residual
  fit$R.assign <- fr$R.assign
  fit$x.assign <- attr(X, 'assign')
}
if(qr)
```



```
fit$qr <- qr(X)
c(fit, list(family = Fam$family, linear.predictors = lp,
deviance = devr, null.deviance = null.dev, call = this.call,
iter = k, y = y, contrasts = attr(X, 'contrasts'),
dispersion = dispersion))
}
}
```

APÊNDICE H

PAVA (*pool adjacent violator algorithm*)

O algoritmo PAVA é um caso particular do algoritmo descrito na Seção 3.3 e vale apenas para o caso de ordem simples. Os passos são os seguintes:

- (i) Verificar se há violações entre as médias $\bar{y}_1, \dots, \bar{y}_k$. Se não existir, terminar o processo e fazer $\tilde{\theta}_i = \bar{y}_i$, $i = 1, \dots, k$. Se ocorrer alguma violação, ir para (ii).
- (ii) Supor a violação $\bar{y}_j > \bar{y}_{j+1}$. A estimativa comum corrigida será dada por

$$\bar{y}_{j,j+1} = \frac{n_j \bar{y}_j + n_{j+1} \bar{y}_{j+1}}{n_j + n_{j+1}}.$$

- (iii) Comparar as $(k-1)$ médias resultantes. Se não ocorrer nenhuma violação, terminar o processo iterativo e fazer $\tilde{\theta}_1 = \bar{y}_1, \dots, \tilde{\theta}_j = \tilde{\theta}_{j+1} = \bar{y}_{j,j+1}, \dots, \tilde{\theta}_k = \bar{y}_k$. Se ocorrer alguma violação, corrigir como em (ii) até não ocorrerem mais violações.

Referências

- Amemiya, T. (1985). *Advanced Econometrics*. Cambridge, MA: Harvard University Press.
- Avriel, M. (1976). *Nonlinear Programming: Analysis and Methods*. Englewood Cliffs, NJ: Prentice-Hall.
- Aitkin, M., Anderson, D., Francis, B. e Hinde, J. (1990). *Statistical Modelling in GLIM*. Oxford: Clarendon Press.
- Barlow, R. E.; Bartholomew, D. J.; Bremner, J. N. e Brunk, H. H. (1972). *Statistical Inference under Order Restrictions*. New York: John Wiley.
- Bartholomew, D. J. (1959a). A test of homogeneity for ordered alternatives, I. *Biometrika* 46, 36-48.
- Bartholomew, D. J. (1959b). A test of homogeneity for ordered alternatives, II. *Biometrika* 46, 328-335.
- Bartholomew, D. J. (1961). A test of homogeneity of means under restricted alternatives. *Journal of the Royal Statistical Society B* 23, 239-281.
- Bohrer, R. e Chow, W. (1979). Algorithm AS122. Weights for one-sided multivariate inference. *Applied Statistics* 27, 100-104.
- Breslow, N.E., Lubin, J. H., Marek, P. e Langholz, B. (1983). Multiplicative models and cohort analysis. *Journal of the American Statistical Association* 78, 1-12.
- Chambers, J. H. e Hastie, J. T. (1992). *Statistical Models in S*. California : Wadsworth & Brooks/Cole Advanced Books & Software Pacific Grove.
- Childs, D. P. (1967). Reduction of the multivariate normal integral to characteristic form. *Biometrika* 54, 293-300.
- Collet, D. (1994). *Modelling Binary Data*. London: Chapman and Hall.
- Cordeiro, G. M. (1987). On the corrections to the likelihood ratio statistics. *Bio-*

- metrika* 74, 265-274.
- Cordeiro, G. M. e McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society B* 53, 629-643.
- Cook, R. D. e Weisberg, S. (1982). *Residuals e Influence in Regression*. New York: Chapman and Hall.
- Cox, D. R. e Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall.
- Dachs, J. N. W. e Paula, G. A. (1988). Testing for ordered ratio rates in follow-up studies with incidence density data. *Revista Brasileira de Probabilidade e Estatística* 2, 125-137.
- Fahrmeir, L. e Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics* 13, 342-368.
- Fahrmeir, L. e Klinger, J. (1994). Estimating and testing generalized linear models under inequality restrictions. *Statistical Papers* 35, 211-229.
- Fiacco, A. V. e McCormick, G. P. (1968). *Nonlinear Programming : Sequential Unconstrained Minimization Techniques*, New York : Wiley
- Finney, D. J. (1971). *Probit Analysis*, Third Edition. Cambridge: Cambridge University Press.
- Finney, D. J. (1978). *Statistical Methods in Biological Assay*, Third Edition. London: Griffin.
- Gill, P. E; Murray, W. e Wright, M. H. (1981). *Practical Optimization*. New York: Academic Press.
- Gouriéroux, C.; Holly, A. e Monford, A. (1982). Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica* 50, 63-80.
- Gouriéroux, G. e Monford, A. (1995). *Statistics and Econometric*. Vols. 1 e 2. Cambridge: Cambridge University Press.
- Hildreth, C. (1957). A quadratic programming procedure. *Naval Research Logistics Quarterly* 4, 79-85.

- Hillier, G. (1986). Joint tests of zero restrictions on nonnegative regression coefficients. *Biometrika* 73, 657-669.
- Jørgensen, B. (1987). Exponential dispersion models (with discussion). *Journal of the Royal Statistical Society B* 49, 127-162.
- Kodde, D. A. e Palm, F. C. (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica* 54, 1243-1248.
- Kredler, Ch. (1993). The SQP-method for linearly constrained maximum likelihood problems. *Technical Report Nr.IAMSI1994.5TUM*, Technical University Munich.
- Kudô, A. (1963). A multivariate analogue of the one-sided test. *Biometrika* 50, 403-418.
- Lawless, J. F. (1980). Inference in the generalized gamma and log-gamma distributions. *Technometrics* 22, 409-419.
- Lee, C. C.; Robertson, T. e Wright, F. T. (1993). Bounds on distributions arising in order restricted inferences with restricted weights. *Biometrika* 80, 405-416.
- Lehmann, E. L. (1983). *Theory of Point Estimation*. New York: John Wiley.
- Luenberger, D. G. (1969). *Optimization by Vector Space Methods*. New York: John Wiley.
- Martinez, J. M. e Santos, S. A. (1995). *Métodos Computacionais de Otimização*. 20º Colóquio Brasileiro de Matemática - IMPA - RJ.
- McCullagh, P. e Nelder J. A. (1989). *Generalized Linear Models*, Second Edition. London: Chapman and Hall.
- McDonald, J. M. e Diamond, I. (1983). Fitting generalized linear models with linear inequality constraints. *Glim Newsletter* 6, 29-36.
- McDonald, J. M. e Diamond, I. (1990). On the fitting of generalized linear models with nonnegative parameter constraints. *Biometrics* 46, 201-206.
- Morgan, B. J. T. (1992). *Analysis of Quantal Response Data*. London: Chapman and Hall.
- Nelder, J. A. e Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society A* 135, 370-384.
- Nüesch, P. E. (1964). *Multivariate test of location for restricted alternatives*. Tese

- de doutorado - Swiss Federal Institute of Technology, Zurich.
- Nüesch, P. E. (1966). On the problem of testing location in multivariate populations for restricted alternatives. *Annals of Mathematical Statistics* 37, 113-119.
- Nyquist, H. (1991). Restricted estimation of generalized linear models. *Applied Statistics* 40, 133-141.
- Paula, G. A. (1993). Assessing local influence in restricted regression models. *Computational Statistics and Data Analysis* 16, 63-79.
- Paula, G. A. (1995). Influence and residuals in restricted generalized linear models. *Journal of Statistical Computation and Simulation* 51, 315-331.
- Paula, G. A. (1996). On approximation of the level probabilities for testing ordered parallel regression lines. *Statistics and Probability Letters* 30, 333-338.
- Paula, G. A. (1997). *Estimação e Testes em Modelos de regressão com Parâmetros Restritos*. livro texto do minicurso da 5ª Escola de Modelos de Regressão, Campos do Jordão, SP.
- Paula, G. A. (1997). One-sided test in dose-response models *a ser submetido*.
- Paula, G. A. e Sen, P. K. (1994). Tests of ordered hypotheses in linkage in heredity. *Statistics and Probability Letters* 20, 395-400.
- Paula, G. A. e Sen, P. K. (1995). One-sided tests in generalized linear models with parallel regression lines. *Biometrics* 51, 1494-1501.
- Paula, G. A. e Rojas, O. V. (1997). On restricted hypotheses in extreme value regression models. *Computational Statistics and Data Analysis* (a aparecer).
- Paula, G. A. e Tudor, R. M. (1986). Utilização da regressão logística para aperfeiçoar o diagnóstico de processo infeccioso pulmonar. *Revista Ciência e Cultura* 38, 1046-1050.
- Payne, C.D., (1986). *The GLIM Manual: Release 3.77*. Oxford, NAG.
- Perlman, M. D. (1969). One-sided problems in multivariate analysis. *Annals of Mathematical Statistics* 40, 549-567.
- Peers, H. W. (1995). Invariant hypothesis testing in order-restricted inference. *Revista Brasileira de Probabilidade e Estatística* 9, 99-118.
- Piegorsch, W. (1990). One-sided-significance tests for generalized linear models un-

- der dichotomous response. *Biometrics* 46, 309-316.
- Pregibon, D. (1981). Logistic regression diagnostics. *The Annals of Statistics* 9, 705-724.
- Powell, M. J. D. (1987). Algorithms for nonlinear constraints that use Lagrangian functions. *Math. Programming* 14, 224-228.
- Robertson, T. e Wright, F. T. (1983). On approximation of the level probabilities and associated distributions in order restricted inference. *Biometrika* 70, 597-606.
- Robertson, T.; Wright, F. T. e Dykstra, R. L. (1988). *Order Restricted Statistical Inference*. New York: John Wiley.
- Rojas, O. V. (1996). *Teste para Hipóteses Restritas em Modelos de Regressão Log-gama Generalizado e Estrutural*. Tese de doutorado - IME-USP.
- Ross, S. M. (1984). *A First Course in Probability*. Macmillan Publishing Company.
- Ryan, D. M. (1974). Penalty and barrier functions. In *Numerical Methods for Constrained Optimization* (Eds. P.E. Gill and W. Murray), pg. 175-190. New York: Academic Press.
- Shapiro, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika* 72, 133-144.
- Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review* 56, 49-62.
- Schittkowski K. (1981). The nonlinear programming method of Wilson, Han and Powell with an augmented Lagrangian type line search function. *Numerische Mathematik* 38, 83-114.
- Sen, P. K. e Singer, J. M. (1993). *Large Sample Methods in Statistics: An Introduction with Applications*. New York: Chapman and Hall.
- Silvapulle, M. J. (1991). On limited dependent variable models: maximum likelihood estimation and test of one-sided hypothesis. *Econometric Theory* 7, 385-395.
- Silvapulle, M. J. (1994). On tests against one-sided hypotheses in some generalized linear models. *Biometrics* 50, 853-858.
- Siskind, V. (1976). Approximate probability integrals and critical values for Bartholomew's test of ordered means. *Biometrika* 63, 641-654.

- Sun, H. J. (1988a). A general reduction method for n-variate normal orthant probability. *Communications in Statistics. Theory and Methods* 17, 3913-3921.
- Sun, H. J. (1988b). A Fortran subroutine for computing normal orthant probabilities. *Communications in Statistics, Simula* 17, 1097-1111.
- Theil, H. e Van de Panne, C. (1960). Quadratic programming as an extension of classical quadratic maximization. *Management Science* 7, 1-20.
- Wang, J. (1996). Asymptotics of least-squares estimators for constrained nonlinear regression. *The Annals of Statistics* 24, 1316-1326.
- Waterman, M. S. (1977). Least squares with non negative regression coefficients. *Journal of Statistical Computation and Simulation* 6, 67-70.
- Wedderburn, R. W. M. (1976). On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika* 63, 27-32.
- Wolak, F. A. (1987). An exact test for multiple inequality and equality constraints in the linear regression model. *Journal of the American Statistical Association* 82, 782-793.
- Wolak, F. A. (1989a). Testing inequality constraints in linear econometric models. *Journal of Econometrics* 41, 205-235.
- Wolak, F. A. (1989b). Local and global testing of linear and nonlinear inequality constraints in nonlinear econometric models. *Econometric Theory* 5, 1-35.
- Wolak, F. A. (1991). The local nature of hypothesis tests involving inequality constraints in nonlinear models. *Econometrika* 59, 981-995.
- Wollan, P. G. e Dykstra, R. L. (1987). Algorithm AS 225 Minimizing linear inequality constrained Mahalonobis distances. *Applied Statistics* 36, 234-240.
- Woodbury, M. (1950). Inverting modified matrices. *Memorandum* 42. Princeton University.