

**Uso de transformações em
modelos de regressão logística**

Noemi Ichihara Ishikawa

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO GRAU DE MESTRE
EM
CIÊNCIAS

Área de Concentração: **Estatística**
Orientadora: **Profa. Dra. Silvia Nagib Elian**

Durante o desenvolvimento deste trabalho, a autora recebeu apoio financeiro do CNPq.

- São Paulo, janeiro de 2007 -

Uso de Transformações em Modelos de Regressão Logística

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Noemi Ichihara Ishikawa e aprovada pela comissão julgadora.

São Paulo, 07 de maio de 2007.

Banca Examinadora:

- Profa. Dra. Silvia Nagib Elian (Orientadora) – IME - USP
- Prof. Dr. Gilberto Alvarenga Paula – IME – USP
- Prof. Dr. Edwin Moises Marcos Ortega – ESALQ - USP

Aos meus pais,
à minha família,
ao meu marido Mauricio
e aos meus amigos.

Agradecimentos

Aos meus pais, Mitiko e Shigeo, que me concederam a vida e nunca mediram esforços para que meus sonhos se concretizassem.

À minha orientadora Prof^a. Dr^a. Silvia Nagib Elian, pela dedicação, paciência e carinho.

Ao meu marido Mauricio, que sempre esteve ao meu lado nos momentos difíceis e com paciência e carinho me ajuda a alcançar meus objetivos.

À minha família, especialmente ao meu irmão Célio e aos meus tios Toshio e Kazumi que me apóiam em todas as decisões.

Às minhas amigas de infância Dayana e Katia, pela eterna amizade.

Aos amigos do IME-USP pelo apoio e estudos compartilhados: Elisa, Eirilânia, Kátia, Marcia, Karina, Ronaldo, Cezar, Fábio e muitos outros.

Aos amigos que me ajudaram a amadurecer profissionalmente e confiaram no meu trabalho: Silene, Carolina, Vera, Reinaldo, Diego, Tie, Érika, Vanessa, David, Marcelo, Danielle, Bernardo, Silvana, Aldo e Fabrício.

Aos professores do Departamento de Estatística do IME-USP, pela dedicação com que compartilham seus conhecimentos.

Finalmente, a todos que de alguma forma contribuíram para que esse trabalho fosse realizado.

Resumo

Modelos para dados binários são bastante utilizados em várias situações práticas. Transformações em Análise de Regressão podem ser aplicadas para linearizar ou simplificar o modelo e também para corrigir desvios de suposições. Neste trabalho, descrevemos o uso de transformações nos modelos de regressão logística para dados binários e apresentamos modelos envolvendo parâmetros adicionais de modo a obter um ajuste mais adequado. Posteriormente, analisamos o custo da estimação quando são adicionados parâmetros aos modelos e apresentamos os testes de hipóteses relativos aos parâmetros do modelo de regressão logística de Box-Cox. Finalizando, apresentamos alguns métodos de diagnóstico para avaliar a influência das observações nas estimativas dos parâmetros de transformação da covariável, com aplicação a um conjunto de dados reais.

Abstract

Binary data models have a lot of utilities in many practical situations. In Regression Analysis, transformations can be applied to linearize or simplify the model and correct deviations of the suppositions. In this dissertation, we show the use of the transformations in logistic models to binary data models and models involving additional parameters to obtain more appropriate fits. We also present the cost of the estimation when parameters are added to models, hypothesis tests of the parameters in the Box-Cox logistic regression model and finally, diagnostics methods to evaluate the influence of the observations in the estimation of the transformation covariate parameters with their applications to a real data set.

Sumário

1	Introdução	1
2	Modelos para dados binários	
2.1	Introdução	3
2.2	Modelos multi-paramétricos para dados binários	10
2.3	Modelos com transformações nas covariáveis	21
2.4	Transformações de Box-Cox	25
3	Custo na estimação do parâmetro adicional	
3.1	Introdução	31
3.2	Custo na estimação do parâmetro adicional em modelos logísticos generalizados	32
3.3	Custo na estimação do parâmetro adicional na transformação de Box-Cox	43
4	Comparação entre grupos no modelo logístico com transformação de Box-Cox	
4.1	Introdução	58
4.2	Custo da estimação de λ	60
4.3	Testes de hipóteses dos efeitos de tratamentos	65
5	Testes de hipóteses dos parâmetros do modelo de regressão logística de Box-Cox	
5.1	Introdução	72
5.2	Teste da necessidade do uso da transformação	73
5.3	Teste do efeito da covariável	74

6 Análise de diagnóstico no modelo de regressão

logística de Box-Cox

6.1 Introdução.	81
6.2 Métodos de diagnósticos.	82
6.3 Aplicação.	84
6.4 Considerações finais.	90

Referências Bibliográficas.	91
--	-----------

Capítulo 1

Introdução

Modelos para dados binários são aqueles em que a variável de interesse (resposta) possui somente dois valores possíveis. Geralmente atribui-se o valor “1” para o acontecimento de interesse (“sucesso”) e o valor “0” para o acontecimento complementar (“fracasso”). Como existem várias situações práticas onde variáveis resposta binárias são encontradas, o estudo sobre o assunto é bastante amplo.

Em Análise de Regressão, transformações podem ser aplicadas para linearizar ou simplificar o modelo e também para corrigir desvios das suposições.

Este trabalho consiste em descrever o uso de transformações nos modelos de regressão logística para dados binários em busca de modelos mais apropriados. No Capítulo 2, serão apresentados modelos para dados binários. Serão descritos os modelos logísticos e algumas transformações e também modelos envolvendo parâmetros adicionais de forma, introduzidos na função de ligação, na função de tolerância ou ainda no preditor linear, que têm a finalidade de aperfeiçoar o ajuste do modelo logístico.

O Capítulo 3 se destina a analisar o custo na estimação quando são adicionados parâmetros aos modelos.

No Capítulo 4 será introduzido o modelo logístico de Box-Cox com uma variável explicativa e dois tratamentos ou grupos a serem comparados. Será analisado o custo da estimação do parâmetro associado à transformação e apresentado um teste de comparação dos grupos.

O Capítulo 5 descreve alguns testes de hipóteses dos parâmetros do modelo de regressão logística de Box-Cox.

Finalmente, no Capítulo 6, serão apresentados métodos de diagnósticos para avaliar a influência das observações nas estimativas dos parâmetros de transformação da covariável e a aplicação dos mesmos em um conjunto de dados com resposta binária.

Capítulo 2

Modelos para dados binários

2.1 Introdução

Modelos para a análise de dados cuja variável de interesse (Y), habitualmente designada por variável resposta, admite apenas dois resultados, e pode depender de outras variáveis (x), designadas variáveis explicativas ou covariáveis, são chamados de modelos binários. Devido à dicotomização da variável resposta, é comum atribuir o código “1” ao resultado mais importante da resposta ou àquele que se pretende relacionar ao acontecimento de interesse (“sucesso”) e o código “0” ao resultado complementar (“fracasso”).

Existem várias situações práticas onde variáveis resposta binárias são encontradas. Áreas de aplicação onde modelos binários são bastante utilizados são:

Agronomia, Biologia, Farmacologia, Engenharia, Zoologia, Medicina, entre outras. Como exemplos temos:

- i) Resultado positivo ou negativo de um exame médico;
- ii) Germinação ou não de uma semente submetida a um ambiente úmido;
- iii) Quebra ou não de um componente de uma máquina submetida à alta temperatura;
- iv) A morte ou não de um inseto exposto a uma dose particular de inseticida.

Nestes casos, a probabilidade de ocorrência dos valores da variável Y é denotada por:

$$p = P(Y = 1) \text{ e } 1 - p = P(Y = 0).$$

Podemos analisar modelos de dados binários através de uma perspectiva de Regressão, admitindo-se que $p = P(Y = 1)$ está relacionada com uma ou mais variáveis explicativas X_1, X_2, \dots, X_k . No entanto, como p assume valores no intervalo $]0,1[$, não é possível admitir uma relação linear da forma $p(\tilde{x}) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$, sendo $\tilde{x} = (x_1 \ x_2 \ \dots \ x_k)$.

Para contornar o problema, determina-se uma transformação h, de modo que $h(p(\tilde{x}))$ pertença ao intervalo $]-\infty, +\infty[$, podendo assim ser modelada pela função linear $\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$.

No contexto de modelos lineares generalizados, a função h é denominada função de ligação.

De modo geral, as transformações têm a particularidade de permitirem a inclusão de parâmetros adicionais, na função de ligação, na função de tolerância (que será definida a seguir), ou mesmo no preditor linear, para modelar melhor os dados, resultando em modelos mais flexíveis.

Apresentaremos os principais modelos de dados binários, iniciando com a transformação logito.

Sabemos que os estudos mais simples de dados binários envolvem uma variável resposta Y que é binária e uma covariável x. A probabilidade de sucesso nesse caso é:

$$P(Y = 1 | x) = p(x).$$

Notando que p representa o valor esperado de Y dado x , temos em um modelo linear simples $\beta_0 + \beta_1 x$ que:

$$E(Y | x) = \beta_0 + \beta_1 x, \quad (2.1)$$

onde β_0 e β_1 são constantes reais desconhecidas.

A expressão (2.1) não é apropriada para dados binários já que $E(Y | x) = p(x)$ se encontra no intervalo $]0,1[$ e $\beta_0 + \beta_1 x$ assume valores no intervalo $]-\infty, +\infty[$.

A função logística resolve esta limitação. Sua expressão analítica é dada por:

$$F(u) = \frac{e^u}{1 + e^u} = (1 + e^{-u})^{-1}, \quad -\infty < u < +\infty. \quad (2.2)$$

Observa-se que esta função tem contradomínio no intervalo $(0,1)$. Assim, tomando $u = \beta_0 + \beta_1 x$ e adaptando (2.1) à função (2.2), temos:

$$p(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

e que é equivalente à expressão:

$$\log \left[\frac{p(x)}{1 - p(x)} \right] = \beta_0 + \beta_1 x,$$

que por sua vez, é conhecida como transformação logística ou logito, ou seja,

$$\text{logit}(p) = \log \left[\frac{p(x)}{1 - p(x)} \right].$$

Esta é uma forma alternativa de definição do modelo de regressão logística linear simples, ou modelo logístico simples. A transformação logito pode também ser aplicada ao modelo logístico linear múltiplo, onde $p(\tilde{x}) = \tilde{x}^T \beta$, com $\tilde{x}^T = (x_1 \ x_2 \ \dots \ x_k)$., vetor contendo os valores das k variáveis explicativas.

Curvas de dose resposta

Modelos de dose resposta, que descreveremos a seguir, constituem-se um caso particular importante dos modelos para dados binários.

Para fixar idéias, vamos considerar um grupo de insetos que foram submetidos à exposição de diferentes e conhecidas concentrações de inseticidas. Se eles forem

expostos a uma pequena concentração de inseticida, nenhum dos insetos irá morrer, mas se forem expostos a uma grande concentração, todos morrerão. Agora, se doses intermediárias de concentrações de inseticidas forem utilizadas, os insetos irão ou não morrer, dependendo de sua tolerância ao inseticida. Aqueles insetos com menor tolerância ao inseticida são mais propensos a morrer que aqueles com maior tolerância.

Associado ao grupo de insetos descritos, existe uma distribuição de probabilidades dos níveis de tolerância. Insetos com tolerância menor que d_i morrerão quando a dose d_i for aplicada, ou seja, a tolerância de uma unidade experimental é a dose crítica que provoca a resposta em estudo, no caso, a morte. Sendo Y a variável aleatória associada à distribuição de tolerância e y a tolerância de cada inseto, a função densidade de probabilidade de Y é denotada por $f(y)$ e a probabilidade de morte quando um determinado inseto é exposto a uma dose d_i é dada por:

$$p_i = P(Y \leq d_i) = \int_{-\infty}^{d_i} f(y) dy .$$

Supondo que a tolerância tem distribuição normal, então Y tem distribuição normal com média μ e variância σ^2 , ou seja,

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\}, \quad -\infty < y < +\infty$$

então, a probabilidade de morte de cada inseto exposto a uma dose d_i é:

$$p_i = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{d_i} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\} dy = \Phi\left(\frac{d_i - \mu}{\sigma}\right),$$

em que $\Phi(\cdot)$ é a função distribuição de probabilidades de uma variável aleatória com distribuição normal padrão.

Considerando $\beta_0 = -\frac{\mu}{\sigma}$ e $\beta_1 = \frac{1}{\sigma}$, temos que

$$p_i = \Phi(\beta_0 + \beta_1 d_i),$$

logo,

$$\beta_0 + \beta_1 d_i = \Phi^{-1}(p_i) = \text{probito}(p_i),$$

ou seja, o probito surge do resultado direto da suposição de que a distribuição da tolerância é normal.

Observamos que para qualquer valor de p em $]0,1[$, o valor probito(p) estará no intervalo $(-\infty, +\infty)$.

Já para o caso em que a tolerância tem distribuição logística, a função densidade de probabilidade da variável aleatória Y é dada por

$$f(y) = \frac{\exp\{(y - \mu)/\tau\}}{\tau[1 + \exp\{(y - \mu)/\tau\}]^2}, \quad -\infty < y < +\infty, \quad -\infty < \mu < +\infty, \quad \tau > 0.$$

Esta distribuição tem média μ e variância $\pi^2\tau^2/3$. Portanto a tolerância de cada indivíduo é:

$$p_i = \int_{-\infty}^{d_i} \frac{\exp\{(y - \mu)/\tau\}}{\tau[1 + \exp\{(y - \mu)/\tau\}]^2} dy = \frac{\exp\{(d_i - \mu)/\tau\}}{1 + \exp\{(d_i - \mu)/\tau\}}.$$

Considerando $\beta_0 = -\frac{\mu}{\tau}$ e $\beta_1 = \frac{1}{\tau}$, temos que

$$p_i = \frac{\exp(\beta_0 + \beta_1 d_i)}{1 + \exp(\beta_0 + \beta_1 d_i)},$$

ou seja,

$$\log \left[\frac{p_i}{1 - p_i} \right] = \beta_0 + \beta_1 d_i,$$

que caracteriza o logito. Logo,

$$\log it(p_i) = \log \left[\frac{p_i}{1 - p_i} \right].$$

Temos então que o logito surge como consequência da suposição de que a variável tolerância tem distribuição logística.

O modelo logístico é certamente o mais utilizado para dados binários pela simplicidade, pela disponibilidade em vários softwares estatísticos e principalmente devido à facilidade de interpretação dos coeficientes de regressão em termos da razão de chances.

Voltemos a considerar o grupo de insetos citado. Se neste grupo, houver a morte, ou seja, a resposta positiva (ocorrência do evento crítico em estudo) de uma proporção de x indivíduos expostos a uma dose d , significa que a proporção de x indivíduos tem tolerância menor ou igual a d e uma proporção $1-x$ de indivíduos tem tolerância maior que d (respondem a uma dose maior que d). Esta situação caracteriza um ensaio

biológico com resposta quantal. Para estes tipos de ensaios biológicos, a tolerância não pode ser estabelecida individualmente.

Se relacionarmos a proporção esperada de respostas positivas com o logaritmo na base 10 das doses aplicadas (Finney, 1978), obteremos uma curva que é denominada curva de dose-resposta. A curva de dose-resposta em um ensaio quantal corresponde à função distribuição da variável aleatória tolerância medida nos ensaios diretos. Na verdade, como foram consideradas o logaritmo das doses, esta curva de dose-resposta corresponde à função distribuição da variável “logaritmo da tolerância”. Para facilitar a nomenclatura, esta variável será denotada simplesmente por tolerância. Evidências experimentais mostram que a curva dose-resposta é geralmente uma sigmóide (curva em forma de “S”) já que o logaritmo da tolerância é mais simetricamente distribuído que a tolerância.

Como vimos, o probito corresponde à tolerância com distribuição normal e logito, à distribuição logística, ou seja, ambas são distribuições de probabilidade simétricas. As curvas associadas às suas distribuições são sigmóides e são simétricas aproximadamente em $p = 0,5$ (ou seja, as curvas tendem a zero com mesma “velocidade” que tendem a 1), sendo que a curva probito tem caudas mais pesadas.

Vamos supor agora que a tolerância dos indivíduos tem uma distribuição Gumbel (que é utilizada como uma distribuição limite na teoria de valores extremos), conhecida também como distribuição de valor extremo, cuja função densidade de probabilidades é dada por:

$$f(y) = \frac{1}{\kappa} e^{(y-\alpha)/\kappa} \exp\{-e^{(y-\alpha)/\kappa}\}, \quad -\infty < y < +\infty$$

onde $-\infty < \alpha < +\infty$ e $\kappa > 0$ são parâmetros desconhecidos. Esta distribuição é assimétrica com moda igual a α , média maior que α e variância dada por $1,645\kappa^2$. Neste caso, a probabilidade de uma resposta positiva de um indivíduo quando exposto a uma dose d_i é dada por:

$$p_i = \int_{-\infty}^{d_i} f(y)dy = 1 - \exp\{-e^{(d_i-\alpha)/\kappa}\},$$

então, considerando $\beta_0 = -\frac{\alpha}{\kappa}$ e $\beta_1 = \frac{1}{\kappa}$, sendo estes parâmetros desconhecidos, temos que:

$$\log\{-\log(1 - p_i)\} = \beta_0 + \beta_1 d_i$$

que caracteriza a transformação complemento log-log. Esta função também transforma uma probabilidade, que está no intervalo (0,1), para um valor no intervalo $(-\infty, +\infty)$ e fornece uma curva assimétrica.

Outra transformação que gera curva dose-resposta assimétrica é a transformação log-log, que é caracterizada por:

$$-\log\{-\log(p_i)\} = \beta_0 + \beta_1 d_i.$$

As transformações complemento log-log e log-log são relacionadas, já que se a primeira delas estiver relacionada à probabilidade de fracasso (1 - p) então a segunda estará relacionada com a probabilidade de sucesso (p).

As transformações apresentadas modelam os dados binários em várias situações mas às vezes, podem não fornecer um bom ajuste dos dados. Há vários modelos alternativos para descrever melhor os dados.

Para solucionar o problema, a sugestão mais comum apresentada na literatura é a adição de parâmetros extras aos modelos. Estes parâmetros podem ser adicionados tanto na função de distribuição acumulada da tolerância, quando se tratarem de modelos de dose-resposta, quanto na generalização das ligações usuais, definindo ligações mais ricas.

Outra sugestão é escolher uma ligação específica e transformar a covariável, em busca de um modelo mais apropriado.

Há também casos em que as duas situações descritas são aplicadas.

Cada uma destas soluções apresenta vantagens e desvantagens, sendo que algum tipo de critério deve ser adotado para escolher qual a melhor delas. Uma estratégia de utilização de transformações é modelar os dados segundo algumas e escolher aquela que produz melhor ajuste.

Vamos descrever nas seções seguintes algumas dessas transformações.

2.2 Modelos multi-paramétricos para dados binários

Nesta seção, apresentaremos transformações nos modelos usuais correspondendo à adição de parâmetros extras. Como foi mencionado anteriormente, estes parâmetros podem ser adicionados à função de distribuição acumulada ou na generalização das ligações usuais, resultando em ligações mais ricas.

De acordo com Silva (1992), em muitos problemas de dose-reposta, o ajuste através de modelos de regressão logística não é satisfatório. Como consequência, a estimação das doses letais extremas, que é muito sensível ao grau de ajuste do modelo, pode ser prejudicada. Portanto, nestes casos, a utilização de modelos de dose-resposta mais flexíveis seria muito vantajosa.

Modelo de Prentice

A probabilidade de morte quando um determinado inseto é exposto a uma dose d é dada pela função distribuição:

$$p(d) = \int_{-\infty}^d f(y) dy, \quad (2.3)$$

sendo que a função densidade de probabilidades $f(y)$ é geralmente normal ou logística, resultando em funções de distribuição probito e logito respectivamente.

Mas, segundo considerações feitas por Berkson (1944), Chambers e Cox (1967) e Schneiderman (1974), citados em Prentice (1976), seria preciso uma amostra muito grande de dados para poder diferenciar qual dos dois modelos (logito ou probito) é mais eficiente. Além disso, muitas vezes, curvas de porcentagem de morte em função da dose são assimétricas. No entanto, nestes casos, testes de falta de ajuste podem não detectar inadequações dos modelos logito ou probito. Assim alternativas mais gerais deveriam ser consideradas.

Prentice (1976) sugeriu então uma função densidade de probabilidades para modelar dados binários com adição de dois novos parâmetros (m_1 e m_2) aos parâmetros

da relação linear ($\eta = x^T\beta$) de modo a incluir os modelos citados (logito e probito) em um modelo paramétrico mais geral. Esta densidade é dada por:

$$f(y) = \frac{(\exp y m_1)(1 + \exp y)^{-(m_1+m_2)}}{B(m_1, m_2)} \quad (2.4)$$

em que $B(m_1, m_2)$ é a função beta com parâmetros m_1 e m_2 , tal que

$$B(m_1, m_2) = \int_0^1 x^{m_1-1} (1-x)^{m_2-1} dx,$$

e $m_1 > 0, m_2 > 0$.

Esta expressão tem utilização bastante vasta, pois gera vários modelos alternativos.

Se $m_1 = m_2 = 1$, a expressão (2.4) resulta em:

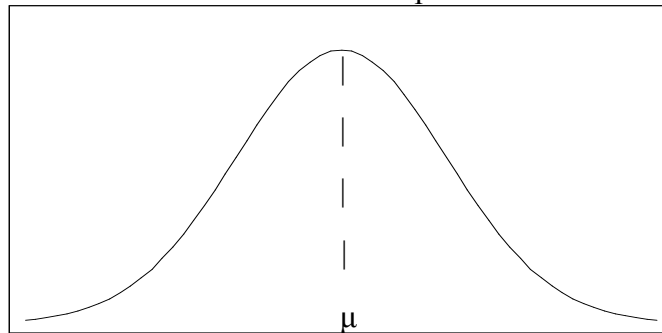
$$f(y) = (\exp y)(1 + \exp y)^{-2}$$

correspondendo ao modelo logístico.

No caso em que $m_1 \rightarrow \infty$ e $m_2 \rightarrow \infty$, verifica-se que a expressão (2.4) se reduz ao modelo probito (ou seja, a expressão converge para a densidade de uma distribuição normal). O modelo complemento log-log será obtido quando $m_1 \rightarrow \infty$ e $m_2 = 1$ na expressão (2.4). O modelo de Prentice também se reduz à distribuição exponencial quando $m_1 \neq 0$ e $m_2 \rightarrow \infty$. Quando $m_1 \rightarrow 0$ e $m_2 \rightarrow 0$ obtemos a função densidade de probabilidade de uma variável aleatória com distribuição exponencial dupla.

Se tivermos $m_1 = m_2$ na expressão (2.4), obteremos os modelos chamados simétricos, ou seja, modelos em que $f(\mu-c) = f(\mu+c)$, sendo μ a média da distribuição, que é igual à mediana nesse caso.

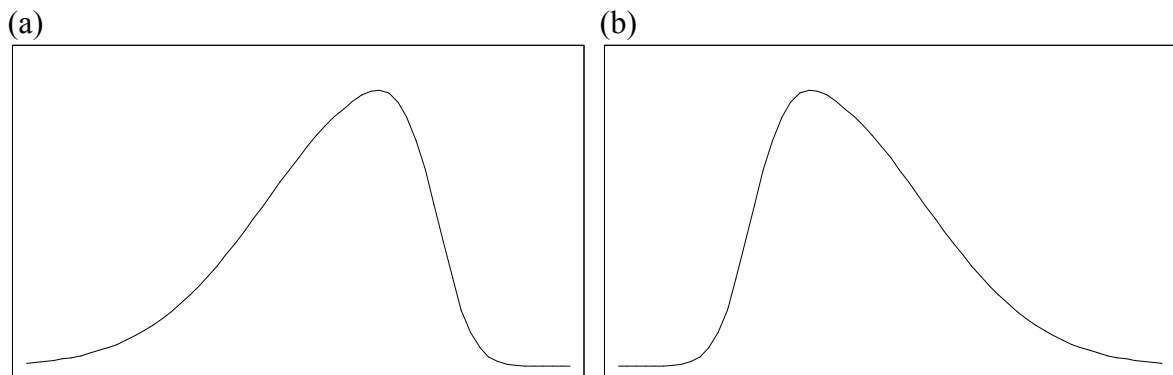
Figura 2.1 - Gráfico de uma densidade de probabilidades simétrica



Quando $m_1 < m_2$, a distribuição é assimétrica à esquerda, correspondendo ao caso em que a média é menor que a mediana (Figura 2.2 (a)). Se $m_1 > m_2$, o modelo tem assimetria positiva, de modo que a média é maior que a mediana como na Figura 2.2 (b).

Apesar da expressão (2.4) apresentar grande utilização, existem problemas computacionais para ajustar este modelo. Uma solução freqüentemente adotada consiste em fixar o valor de um dos parâmetros adicionais. Com isso, os modelos mais usuais são os Modelos de Prentice- m_1 e de Prentice- m_2 .

Figura 2.2 - Gráficos de densidades de probabilidades assimétricas



O Modelo de Prentice- m_1 consiste em fixarmos $m_2 = 1$ na expressão (2.4), sendo que a densidade resultante é dada por:

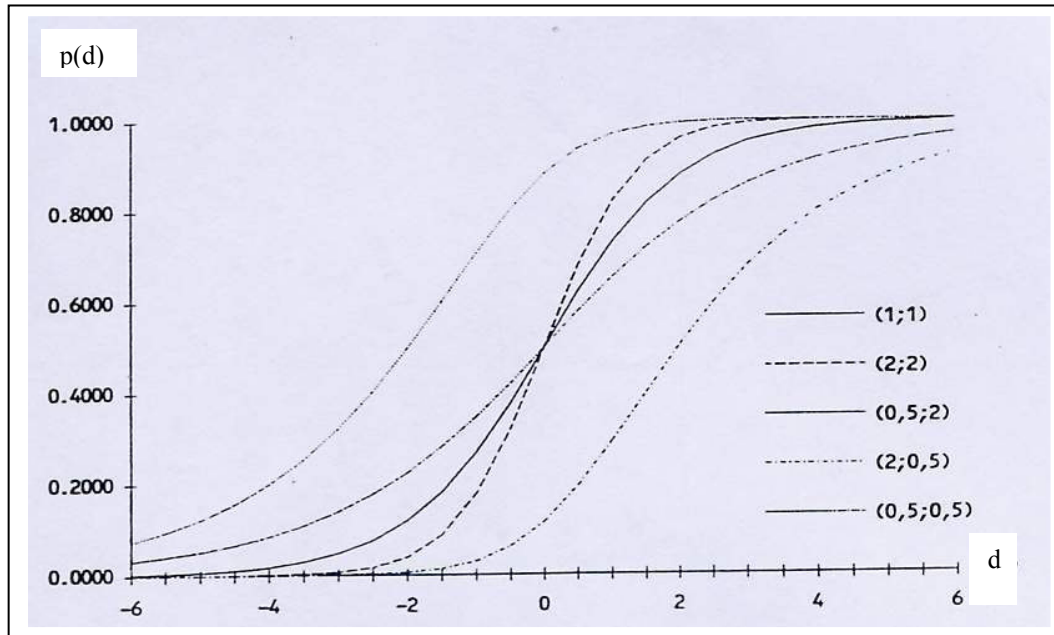
$$f(y) = m_1(\exp y m_1)(1 + \exp y)^{-(m_1+1)}.$$

Já no Modelo de Prentice- m_2 , fixando $m_1 = 1$, obtemos a densidade definida por:

$$f(y) = m_2(\exp y)(1 + \exp y)^{-(m_2+1)}.$$

A Figura 2.3 mostra a probabilidade de sucesso $p(d)$ expressa na função (2.3), onde d assume valores pertencentes ao intervalo $(-6; 6)$ para alguns valores de m_1 e m_2 na função (2.4).

Figura 2.3 – Gráfico de p(d) no modelo de Prentice



Quantit Analysis (Modelo Ômega)

Quantit Analysis é um método paramétrico introduzido por Copenhaver e Mielke (1977) para analisar ensaios com resposta quantal, baseado em uma família de distribuições de tolerância com três parâmetros.

Tal distribuição de probabilidades é designada como distribuição ômega. A função distribuição acumulada $F(x)$ e a função densidade de probabilidade são caracterizadas por:

$$F(x(q)) = q,$$

$$f(x(q)) = 1 - |2q - 1|^{v+1}$$

sendo $x(q)$ o quantil de ordem q da variável aleatória X , f é a função densidade de X e

$$x(q) = \int_{\frac{1}{2}}^q \frac{dz}{f(x(z))},$$

ou seja,

$$x(q) = F^{-1}(q) = \int_{\frac{1}{2}}^q \frac{dz}{1 - |2z - 1|^{v+1}}, \quad 0 < q < 1, \quad v > -1$$

e v é um parâmetro adicional.

Quando $v = 1$, temos que

$$f(x(q)) = 1 - |2q - 1|^2 = 4q(1 - q)$$

e

$$F^{-1}(q) = \int_{\frac{1}{2}}^q \frac{dz}{4z(1-z)} = \frac{1}{4} \ln\left(\frac{q}{1-q}\right),$$

ou seja,

$$F(q) = \frac{e^{4q}}{1 + e^{4q}},$$

que caracteriza a função de distribuição acumulada da ligação logito de q . A função densidade de probabilidades neste caso é

$$f(q) = \frac{dF(q)}{dq} = \frac{4e^{4q}}{(1 + e^{4q})^2}.$$

No caso em que $v = 0$,

$$f(x(q)) = 1 - |2q - 1|$$

e

$$F^{-1}(q) = \int_{\frac{1}{2}}^q \frac{dz}{1 - |2z - 1|} = \begin{cases} \frac{1}{2} \ln 2q, & q < \frac{1}{2} \\ -\frac{1}{2} \ln(2q - 2), & q > \frac{1}{2} \end{cases}$$

ou seja, a função distribuição acumulada é dada por

$$F(q) = \begin{cases} \frac{1}{2} e^{2q}, & q < \frac{1}{2} \\ 1 + \frac{1}{2} e^{-2q}, & q > \frac{1}{2} \end{cases}$$

e a função densidade de probabilidades é

$$f(q) = \frac{dF(q)}{dq} = e^{-2|q|},$$

o que caracteriza uma função densidade de probabilidades de uma distribuição exponencial dupla.

Se $v \rightarrow \infty$, temos a distribuição uniforme pois

$$\lim_{v \rightarrow \infty} F^{-1}(q) = \lim_{v \rightarrow \infty} \int_{\frac{1}{2}}^q \frac{dz}{1 - |2z - 1|^{v+1}} = \int_{\frac{1}{2}}^q dz = q - \frac{1}{2},$$

então

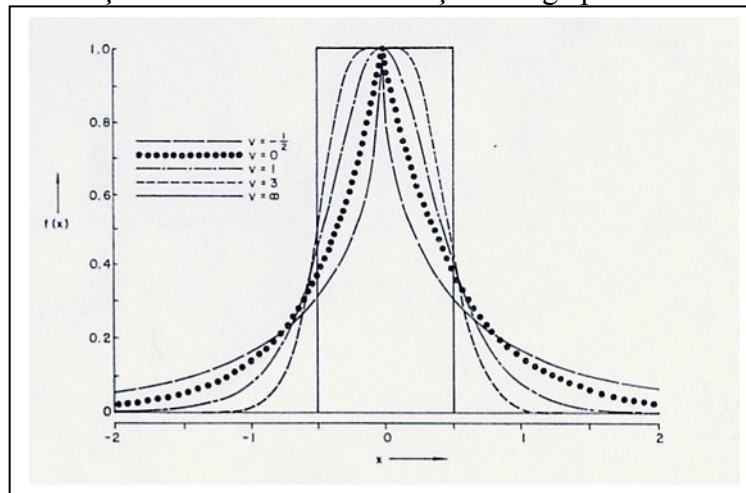
$$F(q) = q + \frac{1}{2},$$

conseqüentemente,

$$f(q) = \frac{dF(q)}{dq} = 1.$$

Nas proximidades de $v = 2$, a distribuição ômega tem comportamento similar ao da distribuição normal, como podemos verificar na Figura 2.4.

Figura 2.4 - Gráfico da função densidade da distribuição ômega para diferentes valores de v



No Modelo de Prentice, há casos em que a distribuição de tolerância é simétrica, como por exemplo, quando $m_1 = m_2$. Neste modelo, o subgrupo onde a distribuição de tolerância é simétrica corresponde aproximadamente ao subgrupo onde $0 < v < 2$ na distribuição ômega. Como consequência, não há distribuições de tolerância simétricas equivalentes no Modelo de Prentice quando a distribuição ômega apresenta $v \geq 2$, caso em que ela apresenta caudas leves, ou para $-1 < v \leq 0$, quando apresenta caudas pesadas. Em compensação, ao contrário do Modelo de Prentice, a família de distribuições ômega não inclui distribuições de tolerância assimétricas.

Modelo de Aranda-Ordaz

Aranda-Ordaz (1981) sugere duas famílias de transformações (ou famílias de funções de ligação) que generalizam a função logito e relacionam a probabilidade de sucesso p ($0 < p < 1$) com a combinação linear de k covariáveis, ou seja, com $\eta = X\beta$ onde X é a matriz determinada pelos valores das variáveis explicativas e β é o vetor de coeficientes das covariáveis. Cada uma dessas famílias de transformações associa a probabilidade de sucesso p à característica de interesse.

O objetivo do autor é construir modelos mais gerais, que incluem o de regressão logística e outros modelos conhecidos com caso particular.

São consideradas inicialmente transformações simétricas. Por transformações simétricas entende-se aquelas que levam à mesma análise se sucesso é trocado por fracasso.

A família de transformações simétricas é dada por:

$$T_\lambda(p) = \frac{2 p^\lambda - (1-p)^\lambda}{\lambda p^\lambda + (1-p)^\lambda} \quad (2.5)$$

em que λ é o parâmetro de transformação.

A função (2.5) satisfaz às condições $T_\lambda(p) = -T_\lambda(1-p)$ e $T_\lambda(p) = T_{-\lambda}(p)$, caracterizando assim, uma função simétrica. Quando $\lambda \rightarrow 0$ na função (2.5), $T_\lambda(p)$ é a transformação logística.

Invertendo-se a função (2.5), obtemos:

$$p(\eta) = \begin{cases} 0 & , \frac{1}{2} \lambda \eta \leq -1 \\ \frac{(1 + \frac{1}{2} \lambda \eta)^{\frac{1}{\lambda}}}{(1 + \frac{1}{2} \lambda \eta)^{\frac{1}{\lambda}} + (1 - \frac{1}{2} \lambda \eta)^{\frac{1}{\lambda}}}, & \left| \frac{1}{2} \lambda \eta \right| < 1 \\ 1 & , \frac{1}{2} \lambda \eta \geq 1 \end{cases} \quad (2.6)$$

em que η é um número real.

O Modelo de Aranda-Ordaz simétrico é caracterizado pela função (2.6).

Há situações em que desejamos tratar sucessos e fracassos assimetricamente. Neste caso, utiliza-se o modelo:

$$\eta = \log \left[\frac{(1-p)^{-\lambda} - 1}{\lambda} \right], \quad (2.7)$$

em que λ é o parâmetro de transformação e η é uma expressão linear.

Se $\lambda \rightarrow 0$ na função (2.7), teremos a transformação complemento log-log e quando $\lambda = 1$, a função (2.7) é a transformação logística.

A inversão da função (2.7) é dada por:

$$p(\eta) = \begin{cases} 1 - (1 + \lambda e^\eta)^{-\frac{1}{\lambda}}, & \lambda e^\eta > -1 \\ 1 & , \lambda e^\eta \leq -1 \end{cases} \quad (2.8)$$

caracterizando o Modelo de Aranda-Ordaz assimétrico.

O modelo de Aranda-Ordaz pode ser visto como um elemento da classe de Modelos Lineares Generalizados uma vez identificadas três componentes: a estrutura dos dados, a parte linear sistemática e a função de ligação.

Considerando que a distribuição dos dados é Binomial $B(n_i, p_i)$, onde n_i é o número de tentativas com probabilidade p_i de sucesso na i -ésima observação ($i = 1, 2, \dots, m$), obtemos então, o parâmetro correspondente ao momento $\mu_i = n_i p_i$. A parte linear sistemática da forma $\eta = X\beta$ e as funções de ligação são dadas pelas transformações inversas de (2.6) e (2.8), assim, o parâmetro do momento pode ser associado à parte linear do modelo dependendo da particular família escolhida.

Portanto, para cada valor do parâmetro de transformação, os modelos propostos podem ser ajustados através de pacotes computacionais que ajustam Modelos Lineares Generalizados (como o S-Plus).

Modelo de Stukel

Stukel (1988) introduziu uma classe de modelos indexados por dois parâmetros de forma α_1 e α_2 ($\alpha = (\alpha_1, \alpha_2)^T$). Estes parâmetros de forma são introduzidos com o objetivo de estender o alcance do modelo logístico para curvas de probabilidade assimétricas e melhorar o ajuste dessas curvas, especialmente nos extremos.

A forma generalizada do modelo proposto é:

$$p_{\alpha}(\eta) = \frac{\exp(h_{\alpha}(\eta))}{1 + \exp(h_{\alpha}(\eta))}$$

ou

$$h_{\alpha}(\eta) = \log\left(\frac{p}{1-p}\right),$$

onde $h_{\alpha}(\eta)$ é uma função estritamente crescente e não linear em η , indexada por dois parâmetros, α_1 e α_2 . Eles são definidos da seguinte maneira:

Para $\eta \geq 0$ e $p \geq \frac{1}{2}$,

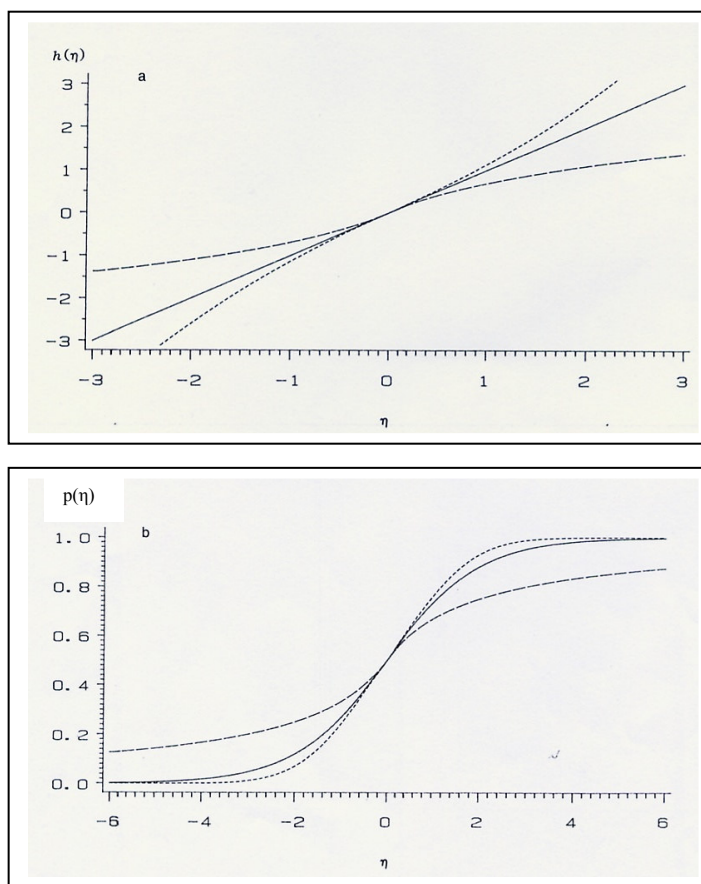
$$h_{\alpha} = \begin{cases} \alpha_1^{-1} (\exp \alpha_1 |\eta|) - 1, & \alpha_1 > 0 \\ \eta & , \alpha_1 = 0 \\ -\alpha_1^{-1} \log(1 - \alpha_1 |\eta|), & \alpha_1 < 0 \end{cases}$$

e para $\eta \leq 0$ e $p \leq \frac{1}{2}$,

$$h_{\alpha} = \begin{cases} -\alpha_2^{-1} (\exp \alpha_2 |\eta|) - 1 & , \alpha_2 > 0 \\ \eta & , \alpha_2 = 0 \\ \alpha_2^{-1} \log(1 - \alpha_2 |\eta|) & , \alpha_2 < 0. \end{cases}$$

Para $h(\eta) = \eta$, ou seja, $\alpha_1 = \alpha_2 = 0$, temos o modelo logístico. No caso em que $\alpha_1 = \alpha_2$, a correspondente curva de probabilidade $p(\eta)$ terá comportamento simétrico. Quando $\alpha_1 \neq \alpha_2$, a curva de probabilidade $p(\eta)$ será assimétrica.

Figura 2.5 – Curvas de $h(\eta)$ e $p(\eta)$



Modelo de Pregibon

Pregibon (1980) sugere o uso da família de funções de ligação da forma

$$g(\eta, \alpha, \delta) = \frac{(\eta + \alpha)^\delta - 1}{\delta},$$

com $\eta = x\beta$ e α e δ parâmetros adicionais.

Na análise de curvas de dose resposta, caso em que a distribuição de probabilidade da tolerância é usualmente o mecanismo que gera a variável Y , o autor

propõe a família de distribuições para $d = \frac{Y - \mu}{\sigma}$,

$$f(d) = \frac{1}{p^{\alpha - \delta - 1} + (1 - p)^{\alpha + \delta - 1}},$$

sendo que $p = P(\text{reação à dose } y = d\sigma + \mu)$ e α e δ são parâmetros adicionais.

Para $\delta = 0$, há simetria no sentido de que p e $1 - p$ podem ser permutados na expressão de f , sem alterar o seu valor.

Verifica-se ainda que o peso das caudas da distribuição é controlado pelo parâmetro α , aumentando com o decréscimo do mesmo.

Teste de adequação da função de ligação

Pregibon (1980) definiu um método para testar a adequação da função de ligação assumida na modelagem dos dados, já que a identificação dos modelos para dados binários pode ser feita através dela.

O autor verifica as conseqüências de ajustar um modelo com a função de ligação $g_0(\mu)$, quando na verdade, a correta mas desconhecida função de ligação é dada por $g_*(\mu)$. Admite-se que tanto $g_0(\mu)$ quanto $g_*(\mu)$ pertencem a uma família de funções de ligação $g(\mu)$, com dois parâmetros (λ_1, λ_2) a serem estimados.

Desta forma, a função de ligação correta é $g_*(\mu) = g(\mu; \lambda_1^*, \lambda_2^*)$, enquanto que a função de ligação adotada é $g_0(\mu) = g(\mu; \lambda_1^0, \lambda_2^0)$. Temos então que os parâmetros λ_i^* ($i = 1$ ou 2) são associados à função de ligação $g_*(\mu)$ e os parâmetros λ_i^0 ($i = 1$ ou 2) são associados à função de ligação $g_0(\mu)$.

Expandindo-se $g_*(\mu)$ em torno de $g_0(\mu)$ através de séries de Taylor de primeira ordem obtém-se aproximadamente:

$$g_*(\mu) = g_0(\mu) + (\lambda_1^* - \lambda_1^0)D_{\lambda_1}(g_0) + (\lambda_2^* - \lambda_2^0)D_{\lambda_2}(g_0)$$

em que $D_{\lambda_1}(g_0) = \left. \frac{\partial}{\partial \lambda_1} g(\mu; \lambda_1, \lambda_2) \right|_{\lambda = \lambda^0}$, e $D_{\lambda_2}(g_0) = \left. \frac{\partial}{\partial \lambda_2} g(\mu; \lambda_1, \lambda_2) \right|_{\lambda = \lambda^0}$,

$$\lambda = (\lambda_1, \lambda_2)^T, \quad \lambda^0 = (\lambda_1^0, \lambda_2^0)^T.$$

Assim, podemos aproximar a função de ligação $g_*(\mu) = x^T \beta$ por:

$$g_0(\mu) = x^T \beta + z^T \gamma$$

em que $z^T = (D_{\lambda_1}(g_0), D_{\lambda_2}(g_0))$ e $\gamma^T = (\lambda_1^* - \lambda_1^0, \lambda_2^* - \lambda_2^0)$.

Desta forma, considerando um ajuste inicial do modelo com função de ligação dada por $g_0(\mu) = x^T \beta$, estima-se β por $\hat{\beta}$ e obtêm-se os valores ajustados $g_0(\hat{\mu}) = x^T \hat{\beta}$, permitindo determinar $\hat{z} = z \Big|_{\beta=\hat{\beta}}$. Para verificar a adequação da função de ligação, basta reajustar o modelo adicionando \hat{z} às variáveis explicativas. Caso haja uma redução significativa na função desvio ou *deviance*, a função de ligação $g_0(\mu)$ não é adequada. Este procedimento também fornece uma estimativa para γ e assim, para λ_1 e λ_2 , que por sua vez, permitem a estimação da função de ligação.

Como podemos verificar, os modelos apresentados nessa seção têm como característica a presença de parâmetros adicionais de forma e geralmente influenciam na simetria e nas caudas (regiões não centrais) das curvas de probabilidade $p(\eta)$. Além disso, os modelos apresentados incluem o modelo logístico como um caso especial, mas geralmente resultam em distribuições mais específicas. Algumas delas conseguem também acomodar dados assimétricos.

2.3 Modelos com transformações nas covariáveis

A transformação da variável explicativa x foi proposta por Finney (1978) para análise de dados com resposta binária, como por exemplo, o uso do valor de logaritmo da dose ao invés de apenas usar o valor da dose em modelos de dados binários.

Muitos métodos considerando a transformação nas covariáveis foram propostos e desenvolvidos, principalmente para a ligação logito. Apresentaremos alguns deles nesta seção.

Modelo Cúbico Logístico

Morgan (1985), motivado pela *Quantit Analysis*, sugeriu dois modelos que utilizam transformações das covariáveis.

Foi visto na *Quantit Analysis*, que quando $v \geq 1$, $F(x)$ gerava uma distribuição de caudas leves bastante útil para análise de dados de resposta quantal. Uma boa aproximação para $F(x)$ neste caso é a função de distribuição acumulada de uma logística cúbica, que foi usado por Page (1977) como uma aproximação da função de distribuição acumulada da Normal. A função distribuição acumulada da logística cúbica é dada por:

$$F(x) = (1 + e^{-(\theta_1 x + \theta_2 x^3)})^{-1},$$

para convenientes parâmetros $\theta_1, \theta_2 \geq 0$, estimados separadamente para cada valor de v .

Baseado nisto, Morgan (1985) apresenta os seguintes modelos logísticos cúbicos:

$$p(x) = \{1 + \exp[\alpha + \beta x + \gamma x^2 + \delta x^3]\}^{-1}, \quad -\infty < x < +\infty$$

e

$$p(x) = \{1 + \exp[-\beta(x - \mu) + \gamma(x - \mu)^3]\}^{-1}, \quad -\infty < x < +\infty$$

sendo μ o percentil de ordem 0,5 desta distribuição.

O segundo modelo resulta em um modelo logístico quando $\gamma = 0$.

Transformação da covariável pelo método de Kay e Little

Kay e Little (1987) propuseram um método para obter melhor ajuste dos modelos de dados binários através da transformação do vetor de variáveis explicativas do modelo de regressão logística, levando em consideração a distribuição dessas variáveis condicionado ao grupo de resultados (a resposta positiva ou negativa da variável Y). Nesta abordagem, ao contrário das anteriores, os autores admitem que as variáveis explicativas são aleatórias.

Considera-se um conjunto de dados com resposta binária onde a variável Y assume valores 0 ou 1 e um grupo de variáveis explicativas $x^T = (x_1 \dots x_k)$ que são os valores observados das variáveis $X^T = (X_1 \dots X_k)$. Denotando por $f_j(x)$ a densidade condicional de X dado $Y = j$ e considerando que o logaritmo da razão destas funções de

densidades é linear em x , ou seja, que $\log \frac{f_1(x)}{f_0(x)} = \alpha_0 + \alpha'x$, uma vez que:

$$\log \frac{f_1(x)}{f_0(x)} = \log \frac{P(Y=1|x)}{P(Y=0|x)} - \log \frac{p_1}{p_0},$$

em que $p_j = P(Y = j)$, $j = 0, 1$, temos que:

$$\log \frac{P(Y=1|x)}{P(Y=0|x)} = \log \frac{f_1(x)}{f_0(x)} + \log \frac{p_1}{p_0} = \log \frac{p_1}{p_0} + \alpha_0 + \alpha'x = \beta_0 + \beta x,$$

onde $\beta_0 = \alpha_0 + \log \frac{p_1}{p_0}$, $\beta = \alpha'$. Desta maneira, se o logaritmo da razão da densidade

condicional de X dado y é uma função linear de x, o modelo logístico é válido para a distribuição condicional de Y dado X = x. Usando essa idéia, consideremos agora que o logaritmo da razão da densidade condicional é uma função linear não em x mas em algum vetor função de x, como por exemplo, g(x). Usando o mesmo processo, temos que:

$$\log \frac{P(Y=1|x)}{P(Y=0|x)} = \beta_0 + \beta' g(x).$$

Logo, o modelo logístico é apropriado quando usamos g(x) como um vetor de variáveis explicativas.

Consideremos agora algumas distribuições específicas de X dado y e determinaremos quais as transformações g(x) necessárias para satisfazer o modelo logístico. Se a densidade condicional de X dado y pertence à família exponencial, sua função densidade de probabilidades pode ser escrita como:

$$f(x|\theta_j) = B(\theta_j)h(x) \exp\left\{\sum_{i=1}^t R_i(x)Q_i(\theta_j)\right\}, \quad \theta' = (\theta_1 \dots \theta_t).$$

Denotando a densidade condicional de X|Y = j por $f_j(x)$ como anteriormente, conclui-se que:

$$\log \frac{f_1(x)}{f_0(x)} = \log \frac{B(\theta_1)}{B(\theta_0)} + \sum_{i=1}^t R_i(x)\{Q_i(\theta_1) - Q_i(\theta_0)\}.$$

Portanto, as transformações necessárias para as variáveis explicativas satisfazerem o modelo logístico são $R_i(x)$, $i = 1, \dots, t$.

A Tabela 2.1 exhibe as transformações necessárias para a validade do modelo de regressão logística para membros da família exponencial.

Tabela 2.1 – Transformações necessárias para membros da família exponencial

$X Y = j$	$R_1(x)$	$Q_1(\theta_j)$	$R_2(x)$	$Q_2(\theta_j)$
Binomial (n, p_j)	x	$\log\{p_j/(1 - p_j)\}$	_____	_____
Poisson (λ_j)	x	$\log \lambda_j$	_____	_____
Normal (μ_j, σ_j^2)	x	μ_j / σ_j^2	x^2	$-1/2\sigma_j^2$
Gama (α_j, β_j)	x	$-\beta_j$	$\log x$	α_j
Beta (α_j, β_j)	$1 - x$	α_j	$\log(1 - x)$	β_j

Observamos que estes membros da família exponencial forneceram uma razoável gama de distribuições para modelar X dado y : as distribuições Binomial e Poisson são apropriadas para variáveis explicativas discretas, a distribuição Normal para variáveis contínuas simétricas, a distribuição Gama para variáveis contínuas assimétricas positivamente e a distribuição Beta para variáveis contínuas que possuem domínio restrito.

Hauck (1990), motivado por Aranda-Ordaz (1981), Guerrero Johnson(1982), Pregibon (1980) e Prentice (1976) são autores que propuseram generalizações do modelo logístico, sugerindo mudança de escala nas variáveis independentes já que, para modelar dados assimétricos, o modelo logístico pode não ser muito viável. Tal modelo, no entanto, não deve ser descartado, já que, de modo geral, é bastante flexível.

O autor sugeriu um método gráfico para a escolha da melhor transformação para as variáveis independentes, com a restrição de não ser facilmente aplicável especialmente em casos de variáveis independentes múltiplas.

Foram apresentados nesta seção alguns modelos que utilizam transformações na covariável. Experiências sugerem que este é um procedimento bastante satisfatório, porém, Hauck (1990) e Siqueira (1992) apontam alguns problemas no ajuste desses modelos quanto a resposta é quantal.

2.4 Transformações de Box Cox

Nas seções anteriores, as transformações sugeridas são as que adicionam parâmetros extras nos modelos ou as que transformamos a covariável escolhendo uma ligação específica. Nesta seção apresentaremos o procedimento desenvolvido por Box e Cox (1964) que define uma família de transformações junto a modelos comuns, mostrando uma abordagem mais sistemática que as anteriores. Estas transformações também introduzem parâmetro na função de ligação do modelo.

Tukey (1957) sugere um modelo com classes de funções do tipo potência dado por:

$$x^{(\lambda)} = \begin{cases} x^\lambda, & \lambda \neq 0 \\ \log x, & \lambda = 0. \end{cases} \quad (2.9)$$

Se $\lambda = 0$, temos a transformação logarítmica, já que:

$$x^{(0)} = \log x.$$

Para se $\lambda = \frac{1}{2}$, utiliza-se transformação raiz quadrada, pois:

$$x^{(1/2)} = x^{1/2} = \sqrt{x},$$

e se $\lambda = -1$, temos:

$$x^{(-1)} = x^{-1} = \frac{1}{x},$$

resultando em uma transformação recíproca.

O modelo de Tukey é bastante útil pois inclui transformações usuais dependendo do valor de λ . O uso destas transformações nos dados antes do ajuste do modelo é comum na análise de modelos pois, entre outras utilidades, corrigem a não linearidade da equação de regressão.

Para evitar a descontinuidade quando $\lambda = 0$ no modelo (2.9), Box e Cox (1964) o alteraram para a seguinte família de transformações:

$$x^{(\lambda)} = \begin{cases} \frac{x^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log x, & \lambda = 0. \end{cases} \quad (2.10)$$

Guerrero e Johnson (1982) aplicam esta transformação à razão de chances de sucesso $\frac{p}{1-p}$ como forma de generalização do modelo logístico.

Vamos assumir que alguma transformação potência da razão de chances de sucesso satisfaz o modelo linear. Assim:

$$\frac{p^{(\lambda)}}{1-p^{(\lambda)}} = \eta = X^T \beta,$$

onde $p = P(Y = 1)$, ou seja, em um modelo dose-reposta, x é a dose e p é a probabilidade de sucesso.

Se considerarmos $\frac{p^{(\lambda)}}{1-p^{(\lambda)}}$ em (2.10), temos:

$$\frac{p^{(\lambda)}}{1-p^{(\lambda)}} = \begin{cases} \frac{(p/(1-p))^\lambda - 1}{\lambda}, \lambda \neq 0 \\ \log\left(\frac{p}{1-p}\right), \lambda = 0. \end{cases}$$

No caso, em que $\lambda = 0$, temos:

$$\frac{p^{(0)}}{1-p^{(0)}} = \log \frac{p}{1-p},$$

que corresponde exatamente ao modelo logístico.

Em particular, o modelo com apenas uma covariável é dado por:

$$\log\left[\frac{p}{1-p}\right] = \beta_0 + \beta_1 x^{(\lambda)} \quad (2.11)$$

que equivale à expressão

$$p = \frac{\exp(\beta_0 + \beta_1 x^{(\lambda)})}{1 + \exp(\beta_0 + \beta_1 x^{(\lambda)})}, \quad (2.12)$$

em que $x^{(\lambda)}$ é definido por (2.10), com $x > 0$.

Se considerarmos um modelo de dose-resposta, x é a dose e p a probabilidade de sucesso (resposta positiva).

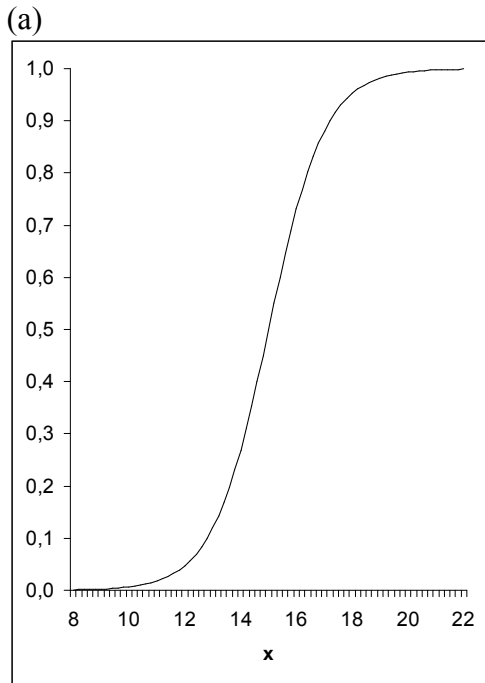
A curva de probabilidades definida na expressão (2.12) é monótona em x e o gráfico de p contra $x^{(\lambda)}$, é geralmente uma sigmóide. A Figura 2.6 mostra diversas curvas

de probabilidade com as quatro possíveis combinações de sinais de β_1 e λ . Para $\beta_1 > 0$ e $\lambda > 0$, temos o gráfico (a) da Figura 2.6. Em (b), temos $\beta_1 < 0$ e $\lambda < 0$, em (c), é considerado $\beta_1 > 0$ e $\lambda < 0$ e em (d), $\beta_1 < 0$ e $\lambda > 0$.

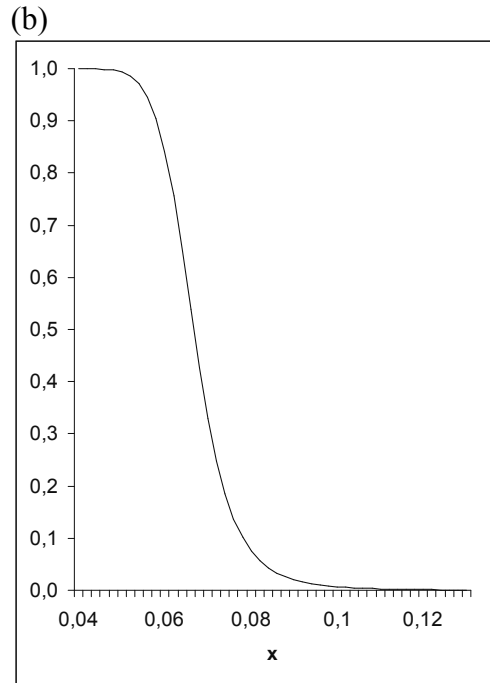
Curvas de probabilidade são também exibidas na Figura 2.7 (a), (b), (c) e (d), representando casos peculiares. Nas curvas (a) e (b), p nunca será 0 para valores positivos de x e nos gráficos (c) e (d), p nunca terá valor 1, pois alcançará no máximo valor próximo de 0,88.

Sob (2.11), o modelo logístico seria apropriado após uma transformação da covariável, mas não sabemos antecipadamente, qual seria ela. A estimação do parâmetro λ será discutida no próximo capítulo.

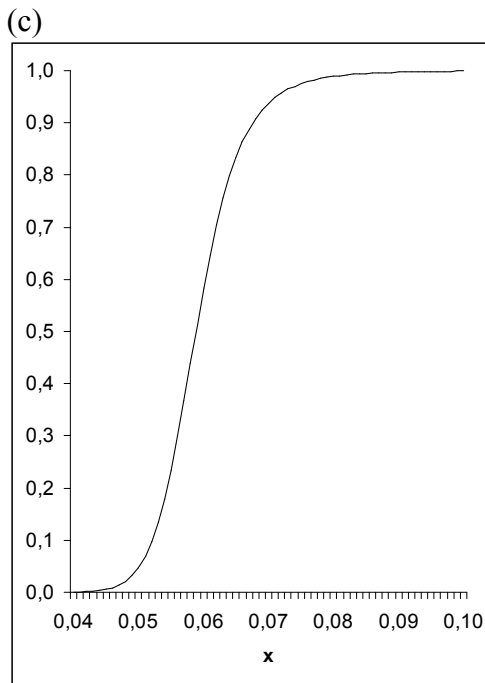
Figura 2.6 - Curvas de probabilidade do modelo logístico Box-Cox



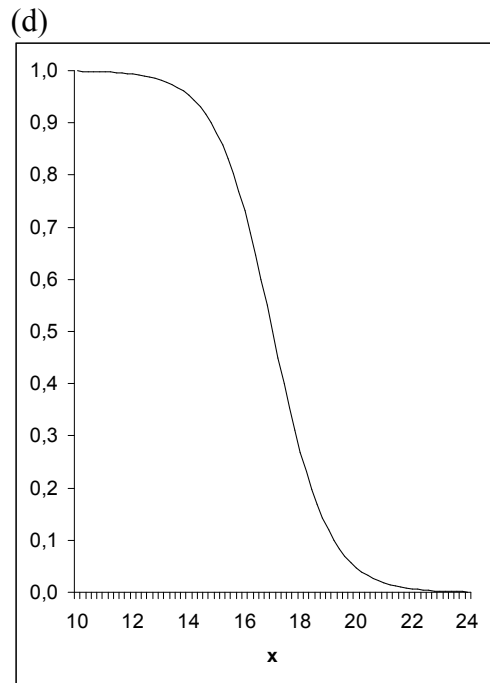
$$\begin{aligned}\beta_0 &= -14 \\ \beta_1 &= 1 \\ \lambda &= 1\end{aligned}$$



$$\begin{aligned}\beta_0 &= -14 \\ \beta_1 &= -1 \\ \lambda &= -1\end{aligned}$$

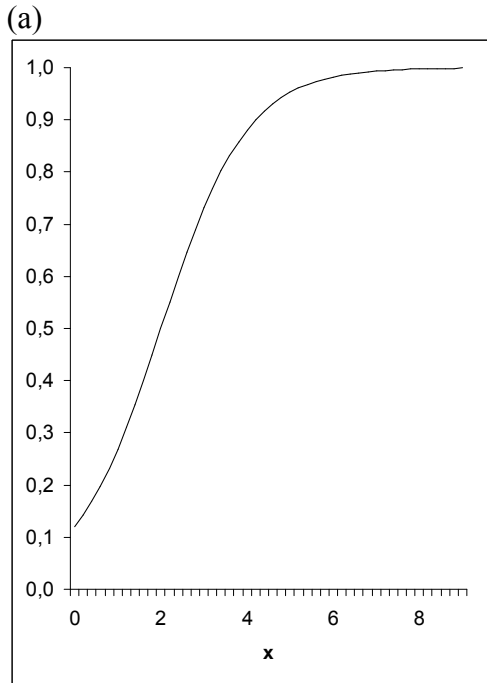


$$\begin{aligned}\beta_0 &= 16 \\ \beta_1 &= 1 \\ \lambda &= -1\end{aligned}$$

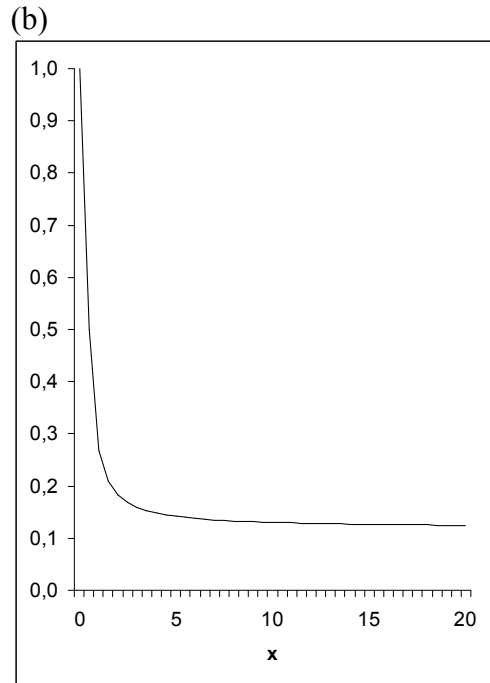


$$\begin{aligned}\beta_0 &= 16 \\ \beta_1 &= -1 \\ \lambda &= 1\end{aligned}$$

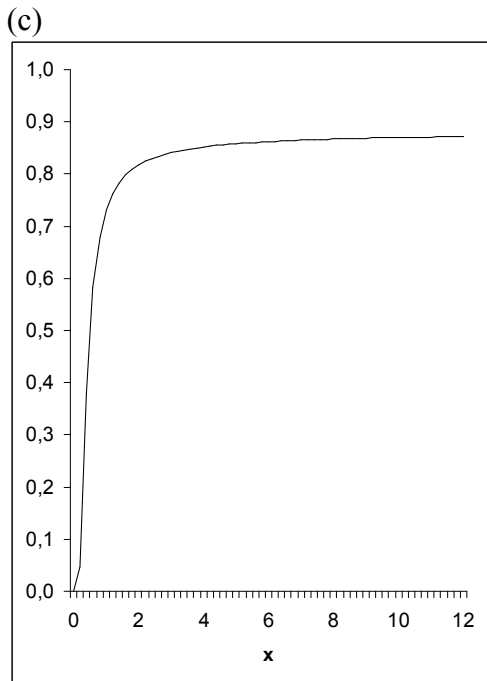
Figura 2.7: Curvas de probabilidade do modelo logístico Box-Cox



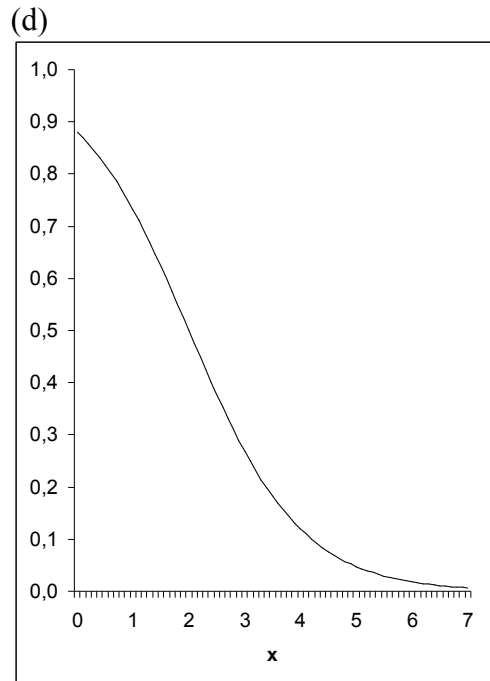
$$\begin{aligned}\beta_0 &= -1 \\ \beta_1 &= 1 \\ \lambda &= 1\end{aligned}$$



$$\begin{aligned}\beta_0 &= -1 \\ \beta_1 &= -1 \\ \lambda &= -1\end{aligned}$$



$$\begin{aligned}\beta_0 &= 1 \\ \beta_1 &= 1 \\ \lambda &= -1\end{aligned}$$



$$\begin{aligned}\beta_0 &= 1 \\ \beta_1 &= -1 \\ \lambda &= 1\end{aligned}$$

Neste capítulo foram apresentados modelos para dados binários, modelos logísticos, algumas transformações e modelos que envolveram parâmetros adicionais para aperfeiçoar o ajuste do modelo logístico.

Como dito anteriormente, no capítulo seguinte serão apresentados o custo da estimação de parâmetros adicionais em modelos multi-paramétricos e transformações da covariável.

Capítulo 3

Custo na estimação do parâmetro adicional

3.1 Introdução

Os modelos de dados binários podem ser analisados através de uma perspectiva de Regressão, relacionando uma ou mais variáveis explicativas com a resposta binária. Um método comum de análise é assumir que uma combinação linear dessas variáveis determinam a probabilidade de resposta através de uma função de ligação específica como as funções logito, probito ou complemento log-log.

Vários autores sugeriram métodos de análise mais flexíveis através da adição de parâmetros nos modelos tanto na função de distribuição acumulada da tolerância, quanto na generalização das ligações usuais. Dentre os autores, temos Copenhaver e Mielke

(1977), Aranda-Ordaz (1981), Guerrero e Johnson (1982) e Morgan (1985) que apresentaram famílias com um parâmetro adicional e Prentice (1976) e Stukel (1988), que apresentaram modelos com dois parâmetros adicionais, como vimos no Capítulo 2.

Quando os modelos são mais flexíveis, temos a vantagem de adquirir melhor ajuste dos dados. Geralmente, estes métodos fornecem uma redução do viés das estimativas das quantidades de interesse, que é uma propriedade desejada, mas com isso, há um aumento da variância, devido à variabilidade da estimativa dos parâmetros adicionais.

Esta inflação da variância é interpretada como o custo na estimação do parâmetro adicional. A redução do viés pela estimação do parâmetro também deve ser levada em consideração.

Uma maneira de interpretar a inflação da variância é como uma medida de quanto a variabilidade está sendo subestimada se a análise é feita condicionalmente no modelo selecionado, ignorando o fato de que o parâmetro extra é desconhecido e foi estimado pelos dados. Outra interpretação possível seria considerar o fator de inflação como um custo devido à adição de um parâmetro desnecessário no modelo.

Neste capítulo, será investigada a inflação da variância devido à inclusão de parâmetros extras nas duas situações descritas anteriormente: modelos multi-paramétricos e transformações da covariável. Iniciaremos o estudo considerando os modelos logísticos generalizados.

3.2 Custo na estimação do parâmetro adicional em modelos logísticos generalizados

Para verificar o custo da adição dos parâmetros nos modelos logísticos generalizados, Taylor (1988) propõe o uso de métodos assintóticos, sendo que o primeiro passo dessa análise é a estimação do parâmetro adicional. O autor se restringe a modelos para famílias de Aranda-Ordaz, embora alguns dos resultados obtidos possam ser aplicados a outras famílias.

Como vimos anteriormente, Aranda-Ordaz (1981) sugeriu uma família de funções assimétricas e simétricas relacionando combinações lineares das covariáveis ($\eta = x\beta$) com a probabilidade de sucesso de um evento, denotado por $P(x)$.

A família de funções assimétricas é definida por:

$$P(x) = \begin{cases} 1 - (1 + \lambda e^\eta)^{-\frac{1}{\lambda}}, & \lambda e^\eta > -1 \\ 1 & \lambda e^\eta \leq -1. \end{cases} \quad (3.1)$$

Como cada Y_i (variável resposta) é uma variável aleatória com distribuição de Bernoulli, então a distribuição de probabilidade de Y_i será:

$$P(Y_i = y_i | x_i) = f(y_i | x_i) = [P(x_i)]^{y_i} [1 - P(x_i)]^{1-y_i}, \quad i=1, 2, \dots, n,$$

em que, $P(x_i)$ é dada por (3.1) e y_i é igual a 1 quando a resposta é sucesso e 0 se a resposta é fracasso.

Admitindo-se adicionalmente independência entre as variáveis aleatórias Y_i , $i = 1, 2, \dots, n$, a função de verossimilhança é expressa por:

$$L = \prod_{i=1}^n f(y_i | x_i) = \prod_{i=1}^n [P(x_i)]^{y_i} [1 - P(x_i)]^{1-y_i},$$

e, o logaritmo da função de verossimilhança é:

$$\begin{aligned} \log L &= \log \prod_{i=1}^n f(y_i | x_i) \\ &= \sum_{i=1}^n [y_i \log P(x_i) + (1 - y_i) \log(1 - P(x_i))]. \end{aligned}$$

Os estimadores de máxima verossimilhança $\hat{\lambda}$ e $\hat{\beta}$ dos parâmetros do modelo são aqueles que maximizam o logaritmo da função de verossimilhança dada anteriormente. Assim, devemos calcular a derivada parcial de $\log L$ em relação a λ e posteriormente igualá-la a zero. Em seguida, repetimos o procedimento em relação a β . Teremos assim um sistema de duas equações para o qual não será possível obter a solução explícita. Utilizamos nesse caso, métodos iterativos para resolver o sistema, determinando as estimativas de máxima verossimilhança de λ e β .

Vamos denotar por $\theta = (\beta, \lambda)$, o vetor de parâmetros, por $\theta_0 = (\beta_0, \lambda_0)$ os verdadeiros valores destes parâmetros e $\hat{\theta} = (\hat{\beta}(\hat{\lambda}), \hat{\lambda})$ a estimativa de máxima verossimilhança de θ . Devido às propriedades dos estimadores de máxima

verossimilhança, θ_0 é o valor limite de $\hat{\theta}$ quando $n \rightarrow \infty$. Denotemos como $\hat{\beta}(\lambda_0)$ o estimador de máxima verossimilhança de β quando λ é fixado no verdadeiro valor λ_0 . Embora $\hat{\beta}(\lambda_0)$ não possa ser calculado com base em um conjunto de observações, é um valor importante a ser considerado para o problema de estimação do aumento da variância devido à seleção de λ .

A matriz de covariância assintótica de $\sqrt{n}\hat{\theta}$ é dada pela inversa da matriz informação I, onde o elemento (j, k) de I é dado por:

$$I_{jk} = -E \left[\frac{\partial^2 \log L}{\partial \theta_j \partial \theta_k} \right]_{\theta=\theta_0} = \sum_{i=1}^n \frac{\partial P_i}{\partial \theta_j} \frac{\partial P_i}{\partial \theta_k} \frac{1}{P_i(1-P_i)}. \quad (3.2)$$

Se λ é conhecido e igual a λ_0 , o vetor de parâmetros se reduz a β e a matriz de informação J será dada por:

$$J_{jk} = -E \left[\frac{\partial^2 \log L}{\partial \beta_j \partial \beta_k} \right]_{\theta=\theta_0}. \quad (3.3)$$

Assim, a matriz de covariância assintótica de $\sqrt{n}\hat{\beta}(\lambda_0)$ é dada por J^{-1} . Para grandes amostras (caso em que $\hat{\lambda}$ é próximo de λ_0), J^{-1} também seria uma aproximação da covariância assintótica de $\sqrt{n}\hat{\beta}(\hat{\lambda})$, condicionado ao valor fixo de $\hat{\lambda}$.

Conforme sugerido por Carroll e Ruppert (1981) e Bickel e Doksum (1981) (citados em Taylor (1988)), o custo da estimação de λ_0 pode ser obtido através da comparação dos termos relevantes de I^{-1} e J^{-1} .

A família de probabilidades simétrica para modelos de dose-resposta, proposta por Aranda-Ordaz, é dada por:

$$P(x) = \begin{cases} 0, & \frac{1}{2}\lambda\eta \leq -1 \\ \frac{(1 + \frac{1}{2}\lambda\eta)^{\frac{1}{\lambda}}}{(1 + \frac{1}{2}\lambda\eta)^{\frac{1}{\lambda}} + (1 - \frac{1}{2}\lambda\eta)^{\frac{1}{\lambda}}}, & \left| \frac{1}{2}\lambda\eta \right| < 1 \\ 1, & \frac{1}{2}\lambda\eta \geq 1. \end{cases} \quad (3.4)$$

Esta família de probabilidades não é inteiramente satisfatória, pois a matriz de informação I é próxima de uma matriz singular quando λ se aproxima de zero.

Se considerarmos um valor fixo da covariável, $X = X_0$, sejam $\hat{P}(X_0, \hat{\lambda}, \hat{\beta}(\hat{\lambda}))$ a estimativa da probabilidade $P(x)$ de (3.1) ou (3.4) quando λ e β são estimados e $\hat{P}(X_0, \lambda_0, \hat{\beta}(\lambda_0))$ a probabilidade estimada quando β é estimado através dos dados e $\lambda = \lambda_0$. Neste caso, as variâncias assintóticas dos estimadores são dadas por:

$$\hat{Var}\left\{\sqrt{n} \hat{P}(X_0, \hat{\lambda}, \hat{\beta}(\hat{\lambda}))\right\} = \left\{ \frac{\hat{P}(X_0, \hat{\lambda}, \hat{\beta}(\hat{\lambda}))}{\partial \theta} \right\}^T I^{-1} \left\{ \frac{\hat{P}(X_0, \hat{\lambda}, \hat{\beta}(\hat{\lambda}))}{\partial \theta} \right\} \Bigg|_{\theta=\theta_0} \quad (3.5)$$

e

$$\hat{Var}\left\{\sqrt{n} \hat{P}(X_0, \lambda_0, \hat{\beta}(\lambda_0))\right\} = \left\{ \frac{\hat{\partial P}(X_0, \lambda_0, \hat{\beta}(\lambda_0))}{\partial \beta} \right\}^T J^{-1} \left\{ \frac{\hat{\partial P}(X_0, \lambda_0, \hat{\beta}(\lambda_0))}{\partial \beta} \right\} \Bigg|_{\theta=\theta_0} \quad (3.6)$$

Desta forma, a razão entre as expressões (3.5) e (3.6) é a inflação da variância do estimador de $P(x)$ devido à estimação de λ , ou seja, é a medida do custo da estimação de λ .

O custo global da estimação de λ poderia ser obtido como uma média ponderada dessas razões para os diferentes valores de X_0 . O autor utiliza como peso da i -ésima observação, o inverso da variância de Y_i para $x = X_i$ que é $w_i = \frac{1}{P(X_i)(1-P(X_i))}$, assumindo $P(X_i)$ diferente a 0 ou 1, dado pelas expressões (3.1) ou (3.4), com $\theta = \theta_0$ e X_i ($i = 1, 2, \dots, m$) os m possíveis diferentes valores de X . Estes pesos são maiores para os P_i 's dos extremos, de modo a compensar a menor variabilidade das estimativas dos extremos nesses pontos.

Seja D a matriz diagonal $m \times m$ com elementos w_i na diagonal principal e V a matriz de ordem $m \times (q + 1)$ fornada pelos elementos $\frac{\partial P(X_i)}{\partial \theta_j}$, com $q = \dim(\beta)$, para $j = 1, 2, \dots, q+1$. Como consequência de (3.2), a matriz I é dada por $I = V'DV$.

Nessas condições, a média ponderada de inflação na variância devido à estimação de λ é:

$$C = \frac{\sum_{i=1}^m \frac{1}{P(X_i)(1-P(X_i))} \widehat{Var}\{\sqrt{n} \widehat{P}(X_i, \widehat{\lambda}, \widehat{\beta}(\widehat{\lambda}))\}}{\sum_{i=1}^m \frac{1}{P(X_i)(1-P(X_i))} \widehat{Var}\{\sqrt{n} \widehat{P}(X_i, \lambda_0, \widehat{\beta}(\lambda_0))\}} = \frac{1+q}{q} = 1 + \frac{1}{q}.$$

Tal fato pode ser verificado pois, com a notação introduzida, o numerador de C se reduz a:

$$\begin{aligned} & \sum_{i=1}^m \frac{1}{P(X_i)(1-P(X_i))} \widehat{Var}\{\sqrt{n} \widehat{P}(X_i, \widehat{\lambda}, \widehat{\beta}(\widehat{\lambda}))\} \\ &= \sum_{i=1}^m w_i (VI^{-1}V')_{ii}, \end{aligned}$$

com $(VI^{-1}V')_{ii}$ representando o i -ésimo elemento da diagonal de $VI^{-1}V'$.

Como $I = V'DV$, e $\text{tr}(AB) = \text{tr}(BA)$ quando ambos os produtos são possíveis, a expressão fica:

$$\begin{aligned} & \text{tr}\{DV(V'DV)^{-1}V'\} \\ &= \text{tr}(I_{q+1}) \\ &= q+1, \end{aligned}$$

em que I_{q+1} é a matriz identidade de ordem $q+1$.

Da mesma maneira, se T é a matriz de ordem $m \times q$ com elementos $\frac{\partial P(X_i)}{\partial \beta_j}$,

devido a (3.3), $J = T'DT$, portanto, o denominador de C é dado por:

$$\begin{aligned} & \sum_{i=1}^m \frac{1}{P(X_i)(1-P(X_i))} \widehat{Var}\{\sqrt{n} \widehat{P}(X_i, \lambda_0, \widehat{\beta}(\lambda_0))\} \\ &= \sum_{i=1}^m w_i (TJ^{-1}T')_{ii} \\ &= \text{tr}\{DT(T'DT)^{-1}T'\} \\ &= \text{tr}(I_q) \\ &= q. \end{aligned}$$

Podemos perceber que este resultado independe do fato da família ser simétrica ou assimétrica, e vale para qualquer outra família definida por um parâmetro extra. Além disso, para famílias com dois parâmetros adicionais, o custo médio da estimação de ambos os parâmetros é igual a $1 + 2/q$, resultado que pode ser mostrado a partir de prova similar.

Numa etapa seguinte, Taylor (1988) considerou o problema de estimar o valor de uma covariável que origina uma probabilidade fixa de sucesso, admitindo o modelo com uma única covariável. Dessa forma, $x\beta$ pode ser escrito como $a + bd$, com a e b parâmetros e d o valor da covariável. Suponha que d_1, d_2, \dots, d_m são os m diferentes valores da covariável e $P^{(i)}$ a verdadeira probabilidade de sucesso em d_i .

Em famílias assimétricas, admitindo $\lambda_0 \neq 0$, a probabilidade de sucesso em d_i é dado por:

$$P^{(i)} = 1 - [1 + \lambda_0 \exp(a_0 + b_0 d_i)]^{-1/\lambda_0}.$$

Suponha que P é a probabilidade de interesse. O estimador de d que fornece uma probabilidade de sucesso P é:

$$\hat{d}(P, \hat{\lambda}, \hat{\beta}(\hat{\lambda})) = \frac{1}{\hat{b}(\hat{\lambda})} \left[\log \left\{ \frac{(1-P)^{-\hat{\lambda}} - 1}{\hat{\lambda}} \right\} - \hat{a}(\hat{\lambda}) \right],$$

em que λ é estimado pelos dados.

No caso em que $\lambda = \lambda_0$ temos que a estimativa de d é uma expressão análoga, descrita por:

$$\hat{d}(P, \lambda_0, \hat{\beta}(\lambda_0)) = \frac{1}{\hat{b}(\lambda_0)} \left[\log \left\{ \frac{(1-P)^{-\lambda_0} - 1}{\lambda_0} \right\} - \hat{a}(\lambda_0) \right].$$

As variâncias assintóticas desses estimadores são dadas respectivamente por:

$$\hat{Var}\{\sqrt{n} \hat{d}(P, \hat{\lambda}, \hat{\beta}(\hat{\lambda}))\} = \left\{ \frac{\partial \hat{d}(P, \hat{\lambda}, \hat{\beta}(\hat{\lambda}))}{\partial \theta} \right\}^T I^{-1} \left\{ \frac{\partial \hat{d}(P, \hat{\lambda}, \hat{\beta}(\hat{\lambda}))}{\partial \theta} \right\} \Bigg|_{\theta=\theta_0} \quad (3.7)$$

e

$$\hat{Var}\{\sqrt{n} \hat{d}(P, \lambda_0, \hat{\beta}(\lambda_0))\} = \left\{ \frac{\partial \hat{d}(P, \lambda_0, \hat{\beta}(\lambda_0))}{\partial \beta} \right\}^T J^{-1} \left\{ \frac{\partial \hat{d}(P, \lambda_0, \hat{\beta}(\lambda_0))}{\partial \beta} \right\} \Bigg|_{\theta=\theta_0}.$$

De maneira análoga, para famílias simétricas, se $P^{(i)}$ é a verdadeira probabilidade de sucesso em d_i , seja $\hat{d}(P, \hat{\lambda}, \hat{\beta}(\hat{\lambda}))$ a estimativa do valor da covariável que origina a probabilidade de sucesso P quando o parâmetro λ é estimado a partir dos dados e de $\hat{d}(P, \lambda_0, \hat{\beta}(\lambda_0))$ o correspondente estimador quando λ é conhecido e igual a λ_0 .

Temos então que para famílias simétricas e assimétricas:

$$C_i = \frac{\widehat{Var}\{\sqrt{n} \widehat{d}(P^{(i)}, \widehat{\lambda}, \widehat{\beta}(\widehat{\lambda}))\}}{\widehat{Var}\{\sqrt{n} \widehat{d}(P^{(i)}, \widehat{\lambda}_0, \widehat{\beta}(\widehat{\lambda}_0))\}} = \frac{\widehat{Var}\{\sqrt{n} \widehat{P}(d_i, \widehat{\lambda}, \widehat{\beta}(\widehat{\lambda}))\}}{\widehat{Var}\{\sqrt{n} \widehat{P}(d_i, \widehat{\lambda}_0, \widehat{\beta}(\widehat{\lambda}_0))\}}, \quad i=1, 2, \dots, m. \quad (3.8)$$

Devido à similaridade entre (3.5) e (3.7), basta mostrar que:

$$\frac{\partial \widehat{P}(d_i, \widehat{\lambda}, \widehat{\beta}(\widehat{\lambda}))}{\partial \theta_j} = K_i \frac{\partial \widehat{d}(P^{(i)}, \widehat{\lambda}, \widehat{\beta}(\widehat{\lambda}))}{\partial \theta_j}, \quad (3.9)$$

em que K_i é uma constante que independe de j , para que a expressão (3.8) seja verdadeira.

Como $\widehat{P}(d_i, \lambda, \beta(\lambda)) = f(\widehat{d}(P^{(i)}, \lambda, \beta(\lambda)))$ para uma função f ,
conseqüentemente $\frac{\partial \widehat{P}(d_i, \widehat{\lambda}, \widehat{\beta}(\widehat{\lambda}))}{\partial \theta_j} = \frac{\partial f}{\partial d} \frac{\partial \widehat{d}(P^{(i)}, \widehat{\lambda}, \widehat{\beta}(\widehat{\lambda}))}{\partial \theta_j} = K_i \frac{\partial \widehat{d}(P^{(i)}, \widehat{\lambda}, \widehat{\beta}(\widehat{\lambda}))}{\partial \theta_j}$, com
 $K_i = \frac{\partial f}{\partial d_i}$, comprovando a afirmação (3.9).

A expressão (3.8) mostra que a inflação da variância nas estimativas das probabilidades é a mesma que a inflação da variância das estimativas dos valores das covariáveis que geraram probabilidades fixas.

Através de um estudo de simulação, o autor analisa o aumento na variância de \widehat{P} com o uso da transformação. Foi considerada inicialmente a família assimétrica (3.1), com $\lambda = 1$, $\beta = (0,1)$ e uma covariável com 5 valores, pertencentes aos seguintes grupos:

Ajuste 1 = $\{-4,60; -2,20; 0,00; 2,20; 4,60\}$;

Ajuste 2 = $\{-2,20; -0,85; 0,00; 0,85; 2,20\}$;

Ajuste 3 = $\{-2,94; -0,85; -0,41; 1,39; 2,20\}$.

A Tabela 3.1 exibe, para cada um dos ajustes, o valor x da covariável, a respectiva probabilidade $P(x)$ obtida a partir de (3.1) e a razão de variâncias do

$$\text{estimador } \hat{P}(x), \left(\frac{\widehat{Var}(\sqrt{n} \hat{P}(X, \hat{\lambda}, \hat{\beta}(\hat{\lambda})))}{\widehat{Var}(\sqrt{n} \hat{P}(X, \lambda_0, \hat{\beta}(\lambda_0)))} \right) \text{ para } \lambda_0 = 1.$$

Tabela 3.1 – Inflação da estimativa da variância devido à estimação de λ na família assimétrica

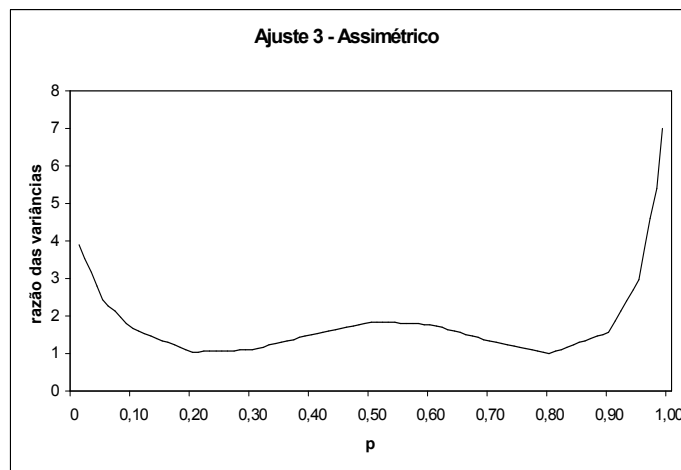
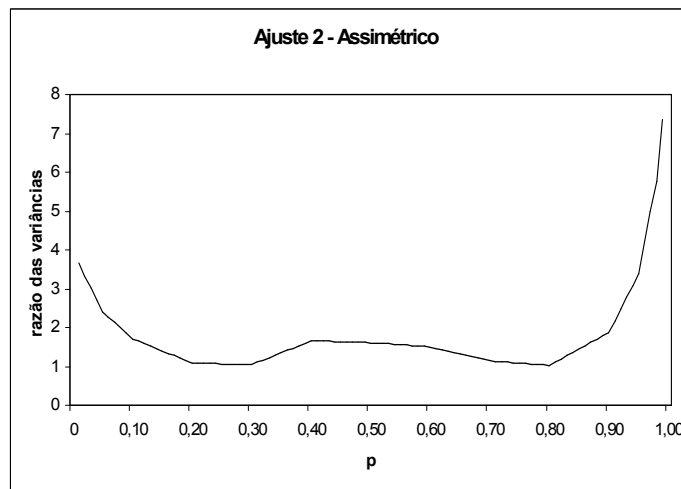
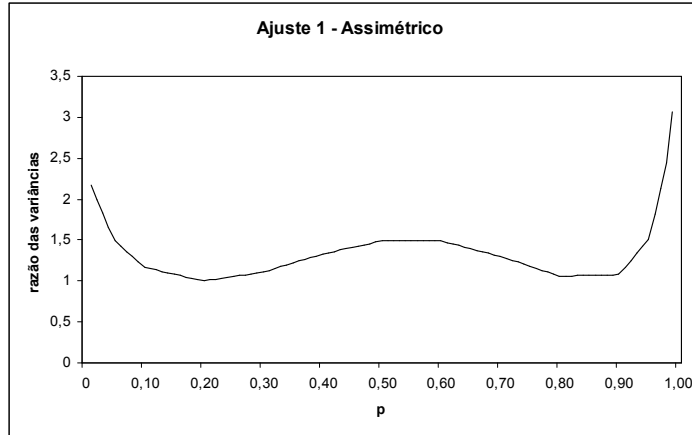
p	Ajuste 1		Ajuste 2		Ajuste 3	
	Covariável	Razão das variâncias	Covariável	Razão das variâncias	Covariável	Razão das variâncias
0,01	-4,60	2,17		3,65		3,89
0,05		1,49		2,40	-2,94	2,43
0,10	-2,20	1,17	-2,20	1,71		1,66
0,20		1,00		1,09		1,04
0,30		1,11	-0,85	1,04	-0,85	1,10
0,40		1,33		1,33	-0,41	1,50
0,50	0,00	1,49	0,00	1,61		1,85
0,60		1,49		1,50		1,77
0,70		1,30	0,85	1,14		1,34
0,80		1,06		1,03	1,39	1,00
0,90	2,20	1,08	2,20	1,88	2,20	1,57
0,95		1,51		3,40		2,98
0,99	4,60	3,06		7,35		7,01

Fonte: Taylor (1988)

Através dela, notamos que embora o custo da estimação de λ tenha média 1,5, esta quantidade pode variar sensivelmente pelos pontos ajustados.

A Figura 3.1 apresenta os gráficos das razões das estimativas das variâncias em função de $p(x)$ para cada ajuste. Observa-se que os gráficos têm forma de W e que para probabilidades extremas, a inflação da variância é grande. O pico central de W geralmente ocorre para probabilidade 0,5, com razão das variâncias próxima de 1,5. A exata posição e altura do pico e depressões de W dependem dos pontos ajustados e dos valores dos parâmetros.

Figura 3.1 – Razão das estimativas das variâncias das estimativas das probabilidades para os ajustes 1, 2 e 3



A Tabela 3.2 mostra a análise feita pelo autor para dados da família simétrica (3.4) com valor de $\lambda = 0,5$ e $\beta = (0, 1)$ e uma covariável com 5 valores, pertencentes aos seguintes grupos:

Ajuste 1 = $\{-3,27; -2,00; 0,00; 2,00; 3,27\}$;

Ajuste 2 = $\{-2,00; -0,83; 0,00; 0,83; 2,00\}$;

Ajuste 3 = $\{-2,51; -0,83; -0,40; 1,33; 2,00\}$.

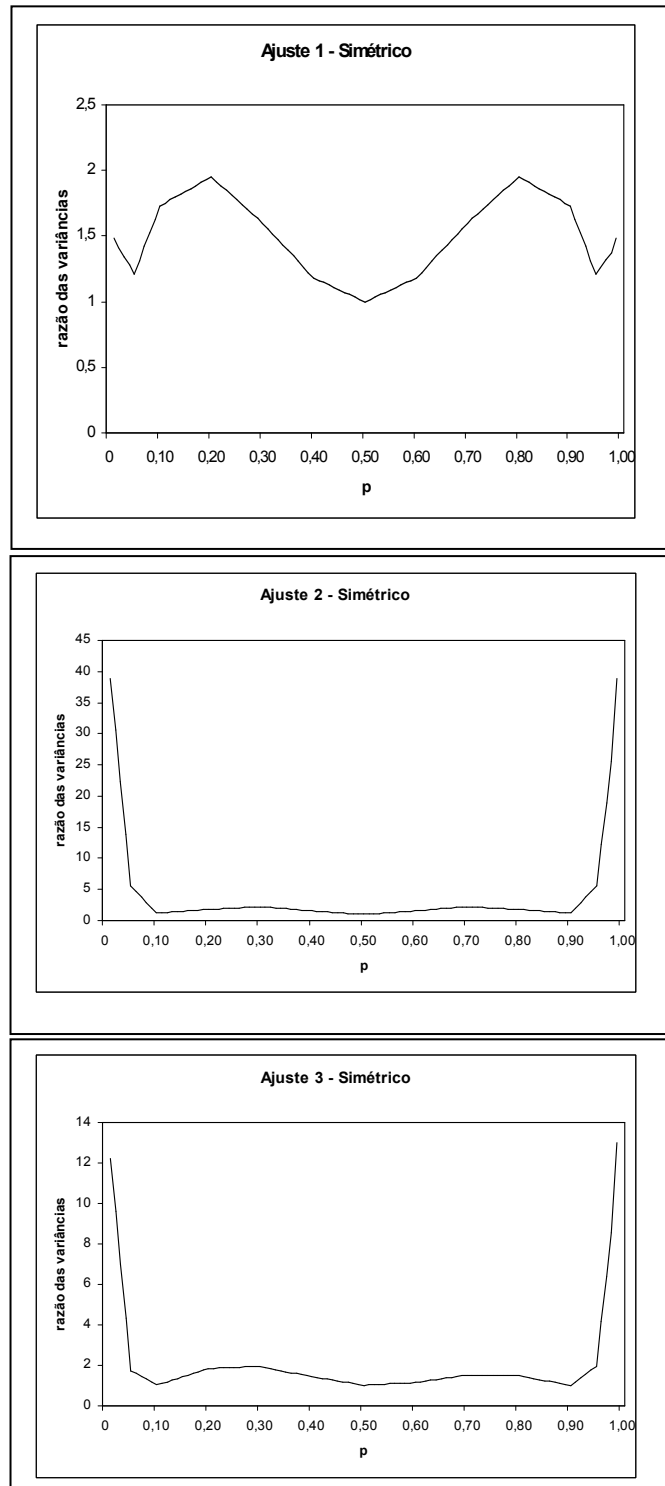
Tabela 3.2 – Inflação das estimativas das variâncias devido à estimação de λ na família simétrica

p	Ajuste 1		Ajuste 2		Ajuste 3	
	Covariável	Razão das variâncias	Covariável	Razão das variâncias	Covariável	Razão das variâncias
0,01	-3,27	1,48		38,91		12,22
0,05		1,21		5,58	-2,51	1,74
0,10	-2,00	1,73	-2,00	1,18		1,03
0,20		1,95		1,78		1,84
0,30		1,60	-0,83	2,24	-0,83	1,95
0,40		1,18		1,54	-0,40	1,43
0,50	0,00	1,00	0,00	1,00		1,02
0,60		1,18		1,54		1,15
0,70		1,60	0,83	2,24		1,52
0,80		1,95		1,78	1,33	1,50
0,90	2,00	1,73	2,00	1,18	2,00	1,00
0,95		1,21		5,58		1,97
0,99	3,27	1,48		38,91		12,68

Fonte: Taylor (1988)

As razões das variâncias para famílias simétricas, exibidas na Tabela 3.2, variam sensivelmente pelos pontos ajustados. Assim como nos modelos assimétricos, a inflação da variância é grande nas probabilidades extremas, como podemos verificar pela Figura 3.2.

Figura 3.2 – Razão das estimativas das variâncias das estimativas das probabilidades para os ajustes 1, 2 e 3



3.3 Custo na estimação do parâmetro adicional da Transformação de Box e Cox

Siqueira e Taylor (1999) analisam o custo da estimação de λ no modelo logístico de Box-Cox incluindo apenas uma covariável, da forma:

$$\log \left[\frac{p(x)}{1-p(x)} \right] = \beta_0 + \beta_1 x^{(\lambda)} \quad (3.10)$$

ou

$$p(x) = \frac{\exp(\beta_0 + \beta_1 x^{(\lambda)})}{1 + \exp(\beta_0 + \beta_1 x^{(\lambda)})} = \{1 + \exp[-(\beta_0 + \beta_1 x^{(\lambda)})]\}^{-1}$$

em que

$$x^{(\lambda)} = \begin{cases} \frac{x^\lambda - 1}{\lambda} & , \lambda \neq 0 \\ \log x & , \lambda = 0. \end{cases}$$

De maneira análoga aos modelos logísticos generalizados, o cálculo do custo da estimação do parâmetro adicional λ é posterior à sua estimativa.

Como a resposta Y_i é uma variável aleatória com distribuição de Bernoulli, o logaritmo da função de verossimilhança L indexado pelos parâmetros $\theta = (\beta_0, \beta_1, \lambda)$ é dado por

$$\begin{aligned} L &= \sum_{i=1}^n [y_i \log p_i + (1 - y_i) \log(1 - p_i)] \\ &= \sum_{i=1}^n y_i (\beta_0 + \beta_1 x_i^{(\lambda)}) - \sum_{i=1}^n \log(1 + \exp(\beta_0 + \beta_1 x_i^{(\lambda)})) \end{aligned}$$

com $p_i = p(x_i)$.

Após cálculos, verifica-se que a matriz de informação definida por

$$I = E \left[- \frac{\partial^2 L}{\partial \theta^2} \right], \text{ é dada por:}$$

$$I^{(\lambda)} = \begin{bmatrix} \sum_{i=1}^n w_i & \sum_{i=1}^n w_i u_i & \sum_{i=1}^n w_i v_i \\ \sum_{i=1}^n w_i u_i & \sum_{i=1}^n w_i u_i^2 & \sum_{i=1}^n w_i u_i v_i \\ \sum_{i=1}^n w_i v_i & \sum_{i=1}^n w_i u_i v_i & \sum_{i=1}^n w_i v_i^2 \end{bmatrix}, \text{ para } \lambda \neq 0$$

e

$$I^{(0)} = \begin{bmatrix} \sum_{i=1}^n w_i^0 & \sum_{i=1}^n w_i^0 \log x_i & \frac{\beta_1}{2} \sum_{i=1}^n w_i^0 (\log x_i)^2 \\ \sum_{i=1}^n w_i^0 \log x_i & \sum_{i=1}^n w_i^0 (\log x_i)^2 & \frac{\beta_1}{2} \sum_{i=1}^n w_i^0 (\log x_i)^3 \\ \frac{\beta_1}{2} \sum_{i=1}^n w_i^0 (\log x_i)^2 & \frac{\beta_1}{2} \sum_{i=1}^n w_i^0 (\log x_i)^3 & \frac{\beta_1}{4} \sum_{i=1}^n w_i^0 (\log x_i)^4 \end{bmatrix}$$

quando $\lambda = 0$, em que

$$u_i = \frac{x_i^\lambda - 1}{\lambda}, \quad v_i = \frac{\beta_1}{\lambda^2} (\lambda x_i^\lambda \log x_i - x_i^\lambda + 1), \quad w_i = p_i(1 - p_i), \quad p_i = [1 + \exp(-(\beta_0 + \beta_1 u_i))]^{-1}$$

e $w_i^0 = p_i^0(1 - p_i^0)$, $p_i^0 = [1 + \exp(-(\beta_0 + \beta_1 \log x_i))]^{-1}$, $i = 1, 2, \dots, n$.

A matriz I pode ser reescrita como

$$I^{(\lambda)} = V' W V, \text{ para } \lambda \neq 0$$

ou

$$I^{(0)} = V^{(0)'} W^{(0)} V^{(0)}, \text{ para } \lambda = 0,$$

com

$$V' = \begin{bmatrix} 1 & 1 & \dots & 1 \\ u_1 & u_2 & \dots & u_n \\ v_1 & v_2 & \dots & v_n \end{bmatrix}, \quad V^{(0)'} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \log x_1 & \log x_2 & \dots & \log x_n \\ \frac{\beta_1}{2} (\log x_1)^2 & \frac{\beta_1}{2} (\log x_2)^2 & \dots & \frac{\beta_1}{2} (\log x_n)^2 \end{bmatrix},$$

e W e $W^{(0)}$ são matrizes diagonais com w_i e w_i^0 , $i = 1, 2, \dots, n$, respectivamente na diagonal principal.

Para obter as estimativas de máxima verossimilhança de β_0 , β_1 e λ , Siqueira (1992) sugere o seguinte procedimento de estimação em dois estágios:

Inicialmente, fixado um particular valor de λ , obtém-se as estimativas de máxima verossimilhança de β_0 e β_1 através do procedimento de Newton-Raphson. Calcula-se então o logaritmo da função de verossimilhança com base no valor de λ fixado e $\hat{\beta}_0$, $\hat{\beta}_1$, estimadores de máxima verossimilhança obtidos. Repete-se o procedimento para cada valor de λ pertencente a um particular conjunto de possíveis valores. Sugere-se que λ seja escolhido no intervalo $[-2, 2]$, que pode ser ampliado se necessário. A estimativa de máxima verossimilhança de λ é o valor desse parâmetro que maximiza o logaritmo da função de verossimilhança.

Verifica-se que esse método de estimação, embora bastante simples, geralmente é executado com sucesso.

Ajustado o modelo, duas quantidades de primordial importância são: a estimativa \hat{p} da probabilidade $p(x)$ para algum valor fixo x da covariável e \hat{x} , estimativa do valor da covariável que fornece probabilidade p especificada. No contexto de dose-resposta, o valor da covariável que produz resposta p é chamado de dose letal (LD(p)) ou de dose eficiente (ED(p)) e é a dose suficiente para causar reação em 100 p % da população em estudo.

No modelo (3.10), a probabilidade de resposta positiva correspondente a um valor fixo de x , por exemplo, x_0 , é dada por:

$$p(x_0) = \frac{\exp(\beta_0 + \beta_1 x_0^{(\lambda)})}{1 + \exp(\beta_0 + \beta_1 x_0^{(\lambda)})},$$

em que $x_0^{(\lambda)}$ é a transformação de Box-Cox no ponto x_0 .

Para uma probabilidade fixa p_0 de ocorrer um evento, o valor da covariável correspondente quando $\lambda \neq 0$ é:

$$x(p_0) = \left\{ \frac{\lambda}{\beta_1} \left[\log\left(\frac{p_0}{1-p_0}\right) - \beta_0 \right] + 1 \right\}^{\frac{1}{\lambda}} \quad (3.11)$$

e para $\lambda = 0$, temos:

$$x(p_0) = \exp\left\{ \frac{1}{\beta_1} \left[\log\left(\frac{p_0}{1-p_0}\right) - \beta_0 \right] \right\}. \quad (3.12)$$

A expressão de $x(p_0)$ em (3.11) pode ser calculada para todo λ somente se $\left\{ \frac{\lambda}{\beta_1} \left[\log \left(\frac{p_0}{1-p_0} \right) - \beta_0 \right] + 1 \right\} > 0$. Verifica-se através de manipulações algébricas simples que para isso,

a) se β_1 e λ são ambos positivos ou negativos, devemos ter

$$p_0 > p_b = \frac{\exp(\beta_0 - \beta_1 / \lambda)}{1 + \exp(\beta_0 - \beta_1 / \lambda)} = \frac{1}{1 + \exp(\beta_1 / \lambda - \beta_0)}.$$

b) se β_1 e λ têm sinais opostos, é necessário que $p_0 < p_b$.

Desta forma, concluímos que o alcance de p é limitado. Em termos práticos, isso significa que não podemos estimar o valor da dose letal (ou genericamente, o valor da covariável) para certos valores de p . O problema ocorre apenas para probabilidades nas caudas das curvas ajustadas, que não estão necessariamente fora do conjunto formado pelos pontos amostrais. Dependendo do objetivo do estudo, tal fato impossibilitaria a estimação do valor de interesse da covariável, o que pode ser considerado uma limitação do modelo logístico de Box-Cox.

A estimativa da probabilidade, \hat{p} , para um valor fixo da covariável e a estimativa do valor da covariável, \hat{x} , para uma probabilidade fixa podem ser obtidas substituindo-se o vetor de estimativas dos parâmetros $\hat{\theta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\lambda})$ nas expressões (3.10) e (3.11) respectivamente. As variâncias assintóticas de \hat{p} e de \hat{x} são calculadas pelo método delta. A partir daí, os intervalos de confiança para a probabilidade p e para o valor da covariável x podem ser construídos.

Para um valor fixo de x , temos que:

$$Var(\hat{p}) = g' I^{-1} g \Big|_{\theta=\theta_0}, \quad (3.13)$$

em que I é a matriz informação (quando $\lambda \neq 0$, $I = I^{(\lambda)}$ e quando $\lambda = 0$, $I = I^{(0)}$) e g é dado por:

$$g = \frac{\partial \hat{p}}{\partial \theta} = w(x)[1, u(x), v(x)]$$

para $\lambda \neq 0$, em que $v(x) = \frac{\beta_1}{\lambda^2}(\lambda x^2 \log x - x^\lambda + 1)$, $u(x) = \frac{x^\lambda - 1}{\lambda}$, $w(x) = p(x)(1 - p(x))$, e $p(x) = \{1 + \exp[-(\beta_0 + \beta_1 u(x))]\}^{-1}$ e $\theta_0 = (\beta_0, \lambda_0)$ os verdadeiros valores destes parâmetros.

Para $\lambda = 0$, temos que:

$$g = \frac{\partial \hat{p}}{\partial \theta} = w^{(0)}(x)[1, \log x, \beta_1 (\log x)^2 / 2],$$

em que $w^{(0)}(x) = p^{(0)}(x)(1 - p^{(0)}(x))$, $p^{(0)}(x) = \{1 + \exp[-(\beta_0 + \beta_1 \log x)]\}^{-1}$.

Analogamente, para p fixo:

$$Var(\hat{x}) = t' I^{-1} t \Big|_{\theta=\theta_0}, \quad (3.14)$$

em que I é a matriz informação e $t = \frac{\partial \hat{x}}{\partial \theta}$. É possível escrever t explicitamente, mas a expressão não é tão simples quanto a de g. Como no modelo logístico de Box-Cox, a correspondência entre p e x é única, então, x é uma função de p, ou seja, $x = h(p)$ dado por (3.11) ou (3.12). Com isso, a expressão (3.14) pode ser calculada por:

$$Var(\hat{x}) = (h'(p))^2 Var(\hat{p})$$

onde $h'(p) = \frac{\partial h(p)}{\partial p}$ é dado por:

$$h'(p) = \frac{\left[\frac{\lambda}{\beta_1} \left(\log \frac{p}{1-p} - \beta_0 \right) + 1 \right]^{\frac{1-\lambda}{\lambda}}}{\beta_1 p(1-p)}, \text{ para } \lambda \neq 0$$

e

$$h'(p) = \frac{\left[\frac{1}{\beta_1} \left(\log \frac{p}{1-p} - \beta_0 \right) \right]}{\beta_1 p(1-p)}, \text{ para } \lambda = 0.$$

Desta forma, a variância do estimador \hat{x} correspondente a uma probabilidade fixa p_0 pode ser obtida calculando-se $Var(\hat{p}_0)$, onde \hat{p}_0 é a estimativa da probabilidade

correspondente a \hat{x} (como se \hat{x} fosse um valor fixo) e posteriormente, $Var(\hat{x}) = (h'(p_0))^2 Var(\hat{p}_0)$.

A magnitude de $Var(\hat{p})$ depende do quão perto o valor da covariável fixa está do alcance dos pontos ajustados. Se está distante, esperamos uma alta variância, no entanto, sabe-se que neste caso, o procedimento de estimação não é recomendado.

Voltando ao estudo do custo da estimação de λ , vamos denotar por $\theta = (\beta, \lambda)$ o vetor de parâmetros, por $\theta_0 = (\beta_0, \lambda_0)$ os verdadeiros valores destes parâmetros e $\hat{\theta} = (\hat{\beta}, \hat{\lambda})$ a estimativa de máxima verossimilhança de θ . Devido às propriedades dos estimadores de máxima verossimilhança, $\hat{\theta}_0$ é o valor limite de $\hat{\theta}$ quando $n \rightarrow \infty$. Denotemos como $\hat{\theta}_{\lambda_0} = (\hat{\beta}_0, \hat{\beta}_1)$ o estimador de máxima verossimilhança de θ quando λ é fixado no verdadeiro valor λ_0 .

De modo geral, quando se utiliza $\hat{\theta}$, a inferência estatística é baseada na matriz de informação I de ordem 3 ($I^{(\lambda)}$ ou $I^{(0)}$). Para λ conhecido, $\lambda = \lambda_0$, a matriz informação $I^*(I^{(\lambda)*}$ ou $I^{(0)*}$) de ordem 2, que é a matriz $I^{(\lambda)}$, de ordem 3, sem a terceira linha e sem a terceira coluna, deve ser usada.

O custo da estimação de λ é definido como a razão dos termos de I^{-1} e I^{*-1} . Temos então que o custo da estimação de β_0 devido à estimação de λ é dado por:

$$RV(\hat{\beta}_0) = \frac{I_{(11)}^{-1}}{I_{(11)}^{*-1}}$$

e de β_1 é:

$$RV(\hat{\beta}_1) = \frac{I_{(22)}^{-1}}{I_{(22)}^{*-1}},$$

sendo que $I_{(kk)}^{-1}$ e $I_{(kk)}^{*-1}$ são, respectivamente, o k -ésimo elemento da diagonal das matrizes I^{-1} e I^{*-1} .

Normalmente não há interesse nestas variâncias assintóticas, pois a interpretação delas depende do parâmetro λ . O interessante seria obter quantidades definidas sem depender do parâmetro λ .

De maneira análoga, o custo da estimação de λ no modelo (3.10) na estimação das probabilidades $p(x)$ ou dos valores das covariáveis pode ser avaliado pela razão das correspondentes variâncias assintóticas, comparando o caso onde λ é conhecido com o caso em que λ é estimado.

Temos que a razão das variâncias das estimativas da probabilidade $p(x_0)$ correspondente a um valor x_0 da covariável é definida por:

$$\frac{Var[\hat{p}(x_0, \hat{\beta}_0, \hat{\beta}_1, \hat{\lambda})]}{Var[\hat{p}(x_0, \hat{\beta}_0^*, \hat{\beta}_1^*)]} \quad (3.15)$$

e a razão das variâncias da estimativa do valor da covariável para uma probabilidade fixada p_0 é dada por:

$$\frac{Var[\hat{x}(p_0, \hat{\beta}_0, \hat{\beta}_1, \hat{\lambda})]}{Var[\hat{x}(p_0, \hat{\beta}_0^*, \hat{\beta}_1^*)]} \quad (3.16)$$

sendo que as variâncias presentes foram definidas em (3.13) e (3.14).

Se x_0 é o valor da covariável correspondente à probabilidade p_0 na curva $p(x)$, então os valores numéricos de (3.15) e (3.16) serão os mesmos. Dessa forma, o aumento da variância de uma probabilidade estimada é o mesmo que o aumento da variância para o valor da covariável estimada correspondente. Isso acontece porque a variância da estimativa do valor da covariável difere da variância da estimativa da probabilidade pela mesma constante em ambos os casos (λ conhecido ou desconhecido) e tal constante é cancelada quando a razão das variâncias é calculada. Este resultado encontra-se na seguinte proposição.

Proposição 3.1:

Suponha que x_1, \dots, x_m são os pontos amostrais correspondentes às probabilidades p_1, \dots, p_m , de modo que $p(x_i) = p_i$. Nestas condições, para cada x_i , $i = 1, \dots, m$:

$$\frac{Var[\hat{p}_i(x_i, \hat{\beta}_0, \hat{\beta}_1, \hat{\lambda})]}{Var[\hat{p}_i(x_i, \hat{\beta}_0^*, \hat{\beta}_1^*)]} = \frac{Var[x_i(p_i, \hat{\beta}_0, \hat{\beta}_1, \hat{\lambda})]}{Var[x_i(p_i, \hat{\beta}_0^*, \hat{\beta}_1^*)]}. \quad (3.17)$$

Prova:

Seja $\theta = (\beta_0, \beta_1, \lambda)$ o vetor de parâmetros. Conforme comentado anteriormente, $\hat{x}(p_i, \theta) = h[\hat{p}(x_i, \theta)]$, para alguma função h . Portanto, para cada elemento de θ (θ_j , $j = 1, 2, 3$), temos:

$$\frac{\partial \hat{x}(p_i, \theta)}{\partial \theta_j} = \frac{\partial h}{\partial p} \frac{\partial \hat{p}(x_i, \theta)}{\partial \theta_j} = h'(p_i) \frac{\partial \hat{p}(x_i, \theta)}{\partial \theta_j} = K_i \frac{\partial \hat{p}(x_i, \theta)}{\partial \theta_j}. \quad (3.18)$$

Note que K_i é uma constante que não depende de j . A expressão (3.18) vale ainda para λ conhecido, caso em que $\theta = (\beta_0, \beta_1)$, para o mesmo valor da constante K_i .

Devido à similaridade entre $Var[\hat{p}_i(x_i, \hat{\beta}_0, \hat{\beta}_1, \hat{\lambda})]$ e $Var[x_i(p_i, \hat{\beta}_0, \hat{\beta}_1, \hat{\lambda})]$, a igualdade da expressão (3.17) pode ser provada.

De (3.13) e (3.14) temos que:

$$Var(\hat{p}) = \left(\frac{\partial \hat{p}}{\partial \theta} \right)^T I^{-1} \frac{\partial \hat{p}}{\partial \theta} \Bigg|_{\theta=\theta_0}$$

e

$$Var(\hat{x}) = \left(\frac{\partial \hat{x}}{\partial \theta} \right)^T I^{-1} \frac{\partial \hat{x}}{\partial \theta} \Bigg|_{\theta=\theta_0}.$$

Como (3.18) vale (para todo $j = 1, 2, 3$ se λ é desconhecido e para $j = 1, 2$ se λ é conhecido), então:

$$\frac{\partial \hat{x}}{\partial \theta} = K_i \frac{\partial \hat{p}}{\partial \theta},$$

para $\theta = (\beta_0, \beta_1, \lambda)$ ou $\theta = (\beta_0, \beta_1)$.

Portanto,

$$Var[x_i(p_i, \hat{\beta}_0, \hat{\beta}_1, \hat{\lambda})] = K_i^2 Var[\hat{p}_i(x_i, \hat{\beta}_0, \hat{\beta}_1, \hat{\lambda})]$$

e

$$Var[\hat{x}_i(p_i, \hat{\beta}_0^*, \hat{\beta}_1^*)] = K_i^2 Var[\hat{p}_i(x_i, \hat{\beta}_0^*, \hat{\beta}_1^*)],$$

o que prova o resultado.

Como foi mostrado anteriormente, no Modelo de Aranda-Ordaz, o custo de estimação de um parâmetro adicional é dado por $1 + \frac{1}{q}$.

Siqueira (1992) mostra que este resultado também é válido para o Modelo de Box-Cox e apresenta uma generalização do mesmo.

O modelo considerado é da forma:

$$\log \frac{p}{1-p} = \beta_0 + \sum_{j=1}^r \beta_j \frac{x_j^{\lambda_j} - 1}{\lambda_j} + \sum_{k=1}^s \gamma_k z_k, \quad (3.19)$$

em que $\lambda_j \neq 0$ ($j = 1, 2, \dots, r$), x_{ij} , $j = 1, 2, \dots, r$, são valores de r variáveis explicativas quantitativas e z_k , $k = 1, 2, \dots, s$, valores de s variáveis indicadoras.

Neste modelo, a probabilidade da resposta positiva para o i -ésimo indivíduo ($i = 1, 2, \dots, n$) é dada por:

$$\begin{aligned} p_i &= \frac{\exp \left[- \left(\beta_0 + \sum_{j=1}^r \beta_j \frac{x_j^{\lambda_j} - 1}{\lambda_j} + \sum_{k=1}^s \gamma_k z_k \right) \right]}{1 + \exp \left[- \left(\beta_0 + \sum_{j=1}^r \beta_j \frac{x_j^{\lambda_j} - 1}{\lambda_j} + \sum_{k=1}^s \gamma_k z_k \right) \right]} \\ &= \frac{1}{1 + \exp \left[- \left(\beta_0 + \sum_{j=1}^r \beta_j \frac{x_j^{\lambda_j} - 1}{\lambda_j} + \sum_{k=1}^s \gamma_k z_k \right) \right]}. \end{aligned} \quad (3.20)$$

Sejam $\theta = (\beta_0, \beta_1, \dots, \beta_r, \gamma_1, \dots, \gamma_s, \lambda_1, \dots, \lambda_r)$ o vetor de parâmetros do modelo e $I^{(\lambda)}$ a matriz de informação. Denotemos por $\hat{\theta}$ o estimador de máxima verossimilhança de θ para λ_j , $j = 1, 2, \dots, r$ desconhecidos e por $\hat{\theta}^*$ o estimador de máxima verossimilhança de θ para λ_j , $j = 1, 2, \dots, r$ conhecidos.

De maneira análoga ao obtido no modelo (3.10), verifica-se que neste caso, a matriz de informação $I^{(\lambda)}$ é da forma:

$$I = G'DG,$$

com

$$G' = (g'_1, \dots, g'_n), \quad g'_i = \frac{\partial p_i}{\partial \theta} = w_i(1, u_{i1}, \dots, u_{ir}, z_{i1}, \dots, z_{is}, v_{i1}, \dots, v_{ir}), \quad w_i = p_i(1 - p_i),$$

$$u_{ij} = \frac{x_{ij}^{\lambda_j} - 1}{\lambda_j}, \quad v_{ij} = \frac{\beta_j}{\lambda_j^2} (\lambda_j x_{ij}^{\lambda_j} \log x_{ij} - x_{ij}^{\lambda_j} + 1) \text{ e } D = \text{diag} \left(\frac{1}{w_i} \right).$$

Se $\lambda_j, j = 1, 2, \dots, r$ são conhecidos, a matriz de informação $I^{(\lambda)*}$ é a matriz $I^{(\lambda)}$ sem as últimas r linhas e r colunas. Ainda é válida a expressão $I^{(\lambda)} = G'DG$, mas para $I^{(\lambda)*}$, $g'_i = w_i(1, u_{i1}, \dots, u_{ir}, z_{i1}, \dots, z_{is})$.

Uma medida geral do custo de estimação de λ utilizada é a razão das médias ponderadas das variâncias, calculada para todos os pontos amostrais.

O custo total da estimação de λ é definido pela razão de variâncias ponderadas:

$$RVP(\hat{p}) = \frac{\sum_{i=1}^n \frac{1}{w_i} \text{var} \left\{ \hat{p}_i \left(\underset{\sim}{c}_i, \hat{\theta} \right) \right\}}{\sum_{i=1}^n \frac{1}{w_i} \text{var} \left\{ \hat{p}_i \left(\underset{\sim}{c}_i, \hat{\theta}^* \right) \right\}}$$

em que $w_i = p_i(1 - p_i)$ e $\underset{\sim}{c}_i = (x_{i1}, \dots, x_{ir}, z_{i1}, \dots, z_{is})$ é o vetor de covariáveis associado ao i -ésimo elemento amostral e $\hat{p}_i(\underset{\sim}{c}_i, \hat{\theta})$ e $\hat{p}_i(\underset{\sim}{c}_i, \hat{\theta}^*)$ são as estimativas de p_i em (3.20) após obtenção de $\hat{\theta}$ e $\hat{\theta}^*$, respectivamente.

O numerador de $RVP(\hat{p})$ é:

$$\begin{aligned} C_i &= \sum_{i=1}^n \frac{1}{w_i} g'_i I^{(\lambda)-1} g'_i \\ &= \sum_{i=1}^n \frac{1}{w_i} (GI^{(\lambda)-1}G')_{ii} \\ &= \text{tr}\{DG(G'DG)^{-1}G'\} \\ &= \text{tr}(I_{2r+s+1}) \\ &= \text{tr}(I_{r+q}) \\ &= r + q, \end{aligned}$$

em que $q = r + s + 1$ e I é a matriz identidade.

Analogamente, é possível mostrar que o denominador de $RVP(\hat{p})$ é igual a q e consequentemente:

$$RV(\hat{p}) = \frac{r+q}{q} = 1 + \frac{r}{q}.$$

Desta forma, quando os pesos são dados como $\frac{1}{p_i(1-p_i)}$, que é o inverso da variância da variável aleatória com distribuição de Bernoulli, o custo médio total devido à estimação dos parâmetros extras com relação ao caso que λ_j é conhecido é exatamente igual a $1+r/q$, onde r é o número de variáveis transformadas e q é o número de parâmetros excluindo-se os λ_j , ou seja, $q = r + s + 1$. Isto significa que a medida total do custo depende apenas do número de parâmetros e o não conhecimento dos λ_j aumenta o custo médio em $100(r/q)\%$. Por exemplo, para o modelo logístico de Box-Cox com apenas uma covariável, $RVP(\hat{p})$ é exatamente 1,5, ou seja, há um custo adicional de 50% com a estimação de λ . A Tabela 3.3 mostra valores de $RVP(\hat{p})$ para alguns modelos.

A prova do resultado foi feita utilizando a transformação de Box-Cox com $\lambda_j \neq 0$, mas prova análoga pode ser feita quando $\lambda_j = 0$.

Este resultado, já obtido no modelo de Aranda-Ordaz, vale para qualquer função de ligação no modelo de resposta binária, desde que a matriz de informação possa ser decomposta na forma $I = G'DG$.

Tabela 3.3 - $\widehat{RV}(p)$ para casos especiais do modelo (3.19)

Logit(p) =	r	s	q	$\widehat{RV}(p) = 1+r/q$
$\beta_0 + \beta_1 x^{(\lambda)}$	1	0	2	$1 + 1/2 = 1,5$
$\beta_0 + \beta_1 x^{(\lambda)} + \gamma$	1	1	3	$1 + 1/3 \approx 1,33$
$\beta_0 + \beta_1 x^{(\lambda)} + \sum_{k=1}^s \gamma_k z_k$	1	s	s + 2	$1 + \frac{1}{s+2}$
$\beta_0 + \sum_{j=1}^r \beta_j x^{(\lambda_j)}$	r	0	r + 1	$1 + \frac{r}{r+1}$
$\beta_0 + \sum_{j=1}^r \beta_j x^{(\lambda_j)} + \sum_{k=1}^s \gamma_k z_k$	r	s	r + s + 1	$1 + \frac{r}{r+s+1}$

r: número de variáveis transformadas

s: número de variáveis não transformadas

q: número de parâmetros excluindo λ_j

Fonte: Siqueira (1992)

Através de um estudo de simulação, Siqueira (1992) analisa o aumento na variância de \widehat{P} com o uso da transformação. Foi considerado o modelo (3.10), com os parâmetros $\beta = (-9, 1)$ e $\lambda = 1$, e uma covariável com 5 valores, pertencentes aos seguintes grupos:

Ajuste 1 = {5,405; 7,803; 10,000; 12,197; 14,595};

Ajuste 2 = {7,803; 9,153; 10,000; 10,847; 12,197};

Ajuste 3 = {7,056; 9,153; 9,595; 11,386; 12,197}.

A Tabela 3.4 exhibe, para cada um dos ajustes, o valor x da covariável, a respectiva probabilidade P(x) obtida a partir de (3.10) e a razão de variâncias do

$$\text{estimador } \widehat{P}(x), \frac{\text{Var}[\widehat{x}(p_0, \widehat{\beta}_0, \widehat{\beta}_1, \widehat{\lambda})]}{\text{Var}[\widehat{x}(p_0, \widehat{\beta}_0^*, \widehat{\beta}_1^*)]}.$$

Conforme os dados da Tabela 3.4, podemos notar que há uma variação considerável na razão de variâncias, especialmente no segundo ajuste. Além disso, esta

quantidade pode ficar bem distante de 1,5, o custo médio total da estimação de λ , obtido para $r = 1$ e $s = 0$. Isso também foi observado no ajuste feito por Taylor (1988) no modelo de Aranda-Ordaz.

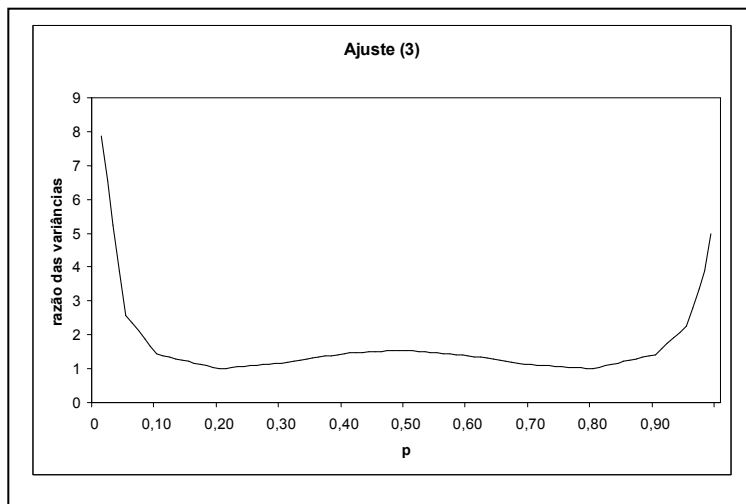
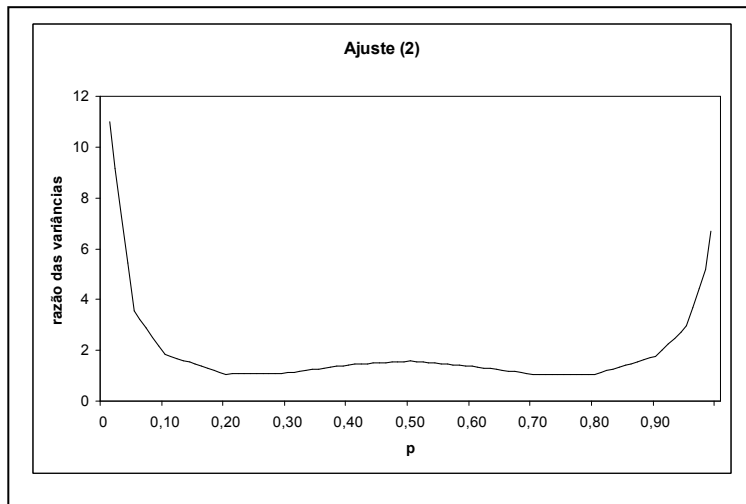
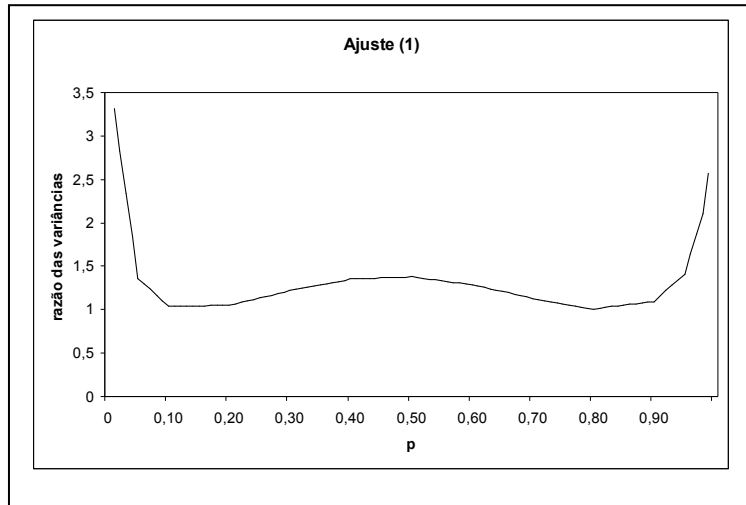
Tabela 3.4 – Inflação da variância pela estimação de λ no modelo de Box-Cox

p	Ajuste 1		Ajuste 2		Ajuste 3	
	Covariável	Razão das variâncias	Covariável	Razão das variâncias	Covariável	Razão das variâncias
0,01	5,405	3,317		11,005		7,875
0,05		1,362		3,556	7,056	2,576
0,10	7,803	1,039	7,803	1,838		1,428
0,20		1,051		1,026		1,000
0,30		1,219	9,153	1,112	9,153	1,174
0,40		1,353		1,430	9,595	1,449
0,50	10,000	1,379	10,000	1,572		1,552
0,60		1,287		1,366		1,389
0,70		1,131	10,847	1,066		1,116
0,80		1,009		1,049	11,386	1,005
0,90	12,197	1,094	12,197	1,766	12,197	1,422
0,95		1,407		2,952		2,250
0,99	14,595	2,575		6,692		4,987

Fonte: Siqueira (1992)

A Figura 3.3 mostra que o gráfico da razão das variâncias, em função das probabilidades tem forma de W nos três ajustes. O pico central de W é próximo de $p = 0,5$, correspondendo a uma razão de variâncias de aproximadamente 1,5.

Figura 3.3 – Razão das Variâncias das estimativas das probabilidades para os ajustes 1, 2 e 3



Pudemos observar neste capítulo o custo da estimação do parâmetro adicional nos modelos para famílias de Aranda-Ordaz e de Box-Cox, através da inflação da variância devido à adição de um parâmetro adicional no modelo.

No capítulo seguinte será introduzido o modelo logístico de Box-Cox com uma variável explicativa e dois tratamentos ou dois grupos a serem comparados. Além disso, serão apresentados o custo na estimação do parâmetro adicional neste modelo e o teste de hipóteses do efeito de tratamentos.

Capítulo 4

Comparação entre grupos no modelo logístico com transformação de Box-Cox

4.1 Introdução

Nesse capítulo analisaremos a situação em que a variável resposta é binária (sucesso ou fracasso), há uma variável explicativa e há dois tratamentos ou dois grupos para serem comparados. Uma maneira de modelar esses dados é assumir que uma função de ligação específica (por exemplo, a função logito) da probabilidade de sucesso (p) pode ser escrita como a combinação linear das variáveis explicativas. Um modelo comum para este caso é dado por:

$$\log \frac{p}{1-p} = \beta_0 + \beta_1 x + \mathcal{A},$$

em que x é uma variável contínua e I é uma variável indicadora (variável binária que identifica o tratamento recebido ou o grupo que o elemento pertence), ou seja, $I = 1$ para o grupo de referência e $I = 0$, caso contrário.

Este tipo de modelo pode ser bastante útil para ensaios clínicos com efeitos de tratamento a serem estimados após controle da covariável x . Usualmente são comparadas respostas nos dois grupos, um deles recebendo tratamento e o outro recebendo um placebo.

Uma generalização do modelo anterior é:

$$\log \frac{p}{1-p} = \beta_0 + \beta_1 x^{(\lambda)} + \gamma I, \quad (4.1)$$

em que,

$$x^{(\lambda)} = \begin{cases} \frac{x^\lambda - 1}{\lambda} & , \lambda \neq 0 \\ \log x & , \lambda = 0 \end{cases}$$

ou seja, $x^{(\lambda)}$ é uma transformação de Box-Cox para x .

Este modelo corresponde a duas curvas relacionando x e p , uma para cada valor de I . Como $x^{(\lambda)}$ é definida para alguns valores de λ somente para quando x é positivo, existirão valores de λ , para os quais as curvas não são definidas para todos os valores de p no intervalo $(0, 1)$. Ou seja, essa limitação para os valores de p pode ocasionar problemas se tentarmos estimar os valores das covariáveis fora dos limites de p . Para que isso não seja um empecilho na estimação das covariáveis, Bickel e Doksum (1981)

(citados em Siqueira e Taylor (1999)), sugere usar a transformação $\frac{|x|^\lambda \text{ sinal}(x) - 1}{\lambda}$ ao

invés de $\frac{x^\lambda - 1}{\lambda}$.

Além disso, o modelo (4.1) pode ser visto como um modelo de regressão logístico padrão ajustado a observações que já foram transformadas ou como um modelo não-linear onde λ é considerado um parâmetro. Se o considerarmos como um modelo de regressão logístico padrão, a interpretação de β_1 é mais simples, porém ignora o fato de que λ é estimado a partir dos dados. Nas duas interpretações do modelo, o parâmetro γ

tem uma interpretação simples. A primeira, que é chamada de condicional, é considerada mais simples que a segunda, que é conhecida como incondicional.

O objetivo deste capítulo é apresentar o teste de hipóteses do efeito de tratamentos no modelo (4.1) e será estudado o custo na estimação do parâmetro λ .

4.2 Custo na estimação de λ

Consideremos $\theta = (\beta_0, \beta_1, \gamma, \lambda)$ como o vetor de parâmetros do modelo (4.1) e $\hat{\theta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\gamma}, \hat{\lambda})$ o estimador de máxima verossimilhança de θ . Conforme introduzido em Siqueira e Taylor (1999), sob este modelo, a matriz informação esperada é dada por:

$$I = \begin{bmatrix} \sum w_i & \sum w_i u_i & \sum w_i I_i & \sum w_i v_i \\ \sum w_i u_i & \sum w_i u_i^2 & \sum w_i u_i I_i & \sum w_i u_i v_i \\ \sum w_i I_i & \sum w_i u_i I_i & \sum w_i I_i^2 & \sum w_i v_i I_i \\ \sum w_i v_i & \sum w_i u_i v_i & \sum w_i v_i I_i & \sum w_i v_i^2 \end{bmatrix}, \quad (4.2)$$

em que $w_i = p_i(1-p_i)$, $p_i = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 u_i + \gamma)]}$, $u_i = \frac{x_i^\lambda - 1}{\lambda}$, para $\lambda \neq 0$ e

$u_i = \log x_i$ para $\lambda = 0$, e $v_i = \frac{\beta_1(\lambda x_i^\lambda \log x_i - x_i^\lambda + 1)}{\lambda^2}$, para $\lambda \neq 0$ e $v_i = \frac{\beta_1(\log x_i)^2}{2}$, para

$\lambda = 0$ e as somatórias variam de $i = 1$ a n .

Suponha que $\lambda = \lambda^*$ conhecido e que $\hat{\theta}^* = (\hat{\beta}_0^*, \hat{\beta}_1^*, \hat{\gamma}^*)$ é o estimador de máxima verossimilhança dos demais parâmetros. A matriz de informação nesse caso é I^* , de dimensão 3×3 . A matriz I^* é a matriz I dada em (4.2), sem a última coluna e a última linha.

O custo da estimação de λ na inferência sobre o efeito do tratamento, que é a medida de quanto a variância do estimador de γ aumenta quando λ é estimado a partir dos dados, pode ser medido pela seguinte razão de variâncias:

$$RV(\hat{\gamma}) = \frac{I_{(33)}^{-1}}{I_{(33)}^{*-1}},$$

em que $I_{(33)}^{-1}$ e $I_{(33)}^{*-1}$ são os elementos que ocupam a posição 33 nas matrizes Γ^{-1} e Γ^{*-1} respectivamente.

Geralmente não há expressão analítica simples para determinar $R\hat{V}(\gamma)$. Quando $R\hat{V}(\gamma) = 1$, concluímos que não há custo adicional da estimação de λ na inferência sobre o efeito de tratamento.

Analisaremos a seguir sob que condições $R\hat{V}(\gamma) = 1$.

Para isso, consideremos a seguinte partição da matriz de informação I:

$$I = \begin{bmatrix} A & B \\ B' & D \end{bmatrix},$$

$$\text{em que } A = \begin{bmatrix} \sum w_i & \sum w_i u_i & \sum w_i I_i \\ \sum w_i u_i & \sum w_i u_i^2 & \sum w_i u_i I_i \\ \sum w_i I_i & \sum w_i u_i I_i & \sum w_i I_i^2 \end{bmatrix}, B = \begin{bmatrix} \sum w_i v_i \\ \sum w_i u_i v_i \\ \sum w_i v_i I_i \end{bmatrix} \text{ e } D = \left[\sum w_i v_i^2 \right].$$

Proposição 4.1: Seja $F = A^{-1}B$. Se $(FF')_{33} = 0$, então $R\hat{V}(\gamma) = 1$.

Prova:

Através de resultados para inversa de matrizes na forma particionada temos que:

$$I^{-1} = \begin{bmatrix} A^{-1} + FE^{-1}F' & -FE^{-1} \\ -E^{-1}F' & E^{-1} \end{bmatrix},$$

em que $E = D - B'A^{-1}B$.

Como E é escalar, $(FE^{-1}F')_{33} = (E^{-1}FF')_{33}$ e se $(FF')_{33} = 0$, então $(FE^{-1}F')_{33} = 0$.

Segue portanto que se $(FF')_{33} = 0$, então $[A^{-1} + FE^{-1}F']_{33} = A^{-1}_{33}$ e com isso, $R\hat{V}(\gamma) = 1$, pois $\Gamma^{-1}_{33} = A^{-1}_{33}$.

Proposição 4.2: Se $\gamma = 0$ e x é a mesma em cada grupo (isto é, os valores de x são os mesmos com mesma quantidade de observações a cada nível de x em cada grupo), então $R\hat{V}(\gamma) = 1$.

A demonstração desse fato encontra-se em Siqueira (1992, página 59).

A Proposição 4.2 é um resultado importante, porque é diretamente aplicável a ensaios clínicos. De acordo com ela, o teste de inexistência de efeito de tratamento pode ser executado mesmo quando a covariável é transformada usando um método baseado nos dados. No entanto, o resultado não contempla a situação em que $\gamma \neq 0$ e o planejamento da variável x não é o mesmo em ambos os grupos.

É desejável estabelecer condições mais gerais que propiciem guias práticos para a inflação da variância no efeito do tratamento, $Var(\hat{\gamma})$, causada pela estimação de λ . Por exemplo, pode-se querer estimar como a distribuição de x nos dois grupos afetaria $RV(\hat{\gamma})$. Entretanto, a formulação para o caso geral quando $\gamma \neq 0$ e o ajuste nos dois grupos não é necessariamente o mesmo, não leva a um padrão geral de comportamento.

Por esses motivos, Siqueira (1992) fez um estudo do custo da estimação de λ , variando o vetor de parâmetros, utilizando novamente uma covariável com 5 valores, pertencentes aos seguintes grupos:

Ajuste 1 = $\{-4,60; -2,20; 0,00; 2,20; 4,60\}$;

Ajuste 2 = $\{-2,20; -0,85; 0,00; 0,85; 2,20\}$;

Ajuste 3 = $\{-2,94; -0,85; -0,41; 1,39; 2,20\}$.

O experimento é balanceado e se $p_{(1)}$ e $p_{(0)}$, correspondem às probabilidades da variável indicadora I ser 1 ou 0, respectivamente então $p_{(1)} = p_{(0)}$.

Foram considerados 5 valores de γ (0, 0,1, 1, 10 e 100) e os outros parâmetros foram considerados de tal maneira que o valor da covariável podia ser estimado para um p fixo no intervalo (0,01; 0,99).

As razões de variâncias nas estimativas de β_0 , β_1 e γ são apresentadas na Tabela 4.1.

Os resultados da Tabela 4.1 mostram que não há custo adicional na estimação de λ na inferência sobre o efeito de tratamento (γ) quando $\gamma = 0$ e é praticamente nulo exceto para um valor muito alto de γ . As razões de variâncias $RV(\hat{\beta}_0)$ e $RV(\hat{\beta}_1)$ demonstram claramente a inflação da variância causada pelo desconhecimento de λ .

Tabela 4.1 – Razão de Variâncias nas estimativas de β_0 , β_1 e γ para três ajustes e cinco vetores de parâmetros.

	Covariáveis	Parâmetros $\theta = (\beta_0, \beta_1, \lambda, \gamma)$	Razão de variâncias		
			β_0	β_1	γ
Ajuste 1	{-4,60; -2,20; 0,00; 2,20; 4,60}	(0; 7; 1; 0)	1,180	1,034	1
	{-4,60; -2,20; 0,00; 2,20; 4,60}	(0; 7,1; 1; 0,1)	1,177	1,047	1,0000006
	{-4,60; -2,20; 0,00; 2,20; 4,60}	(0; 8; 1; 1)	1,137	1,234	1,0000822
	{-4,60; -2,20; 0,00; 2,20; 4,60}	(0; 17; 1; 10)	1,005	1,917	1,0208943
	{-4,60; -2,20; 0,00; 2,20; 4,60}	(0; 107; 1; 100)	1,000	1,989	1,1688381
Ajuste 2	{-2,20; -0,85; 0,00; 0,85; 2,20}	(0; 7; 1; 0)	1,284	1,014	1
	{-2,20; -0,85; 0,00; 0,85; 2,20}	(0; 7,1; 1; 0,1)	1,279	1,032	1,0000003
	{-2,20; -0,85; 0,00; 0,85; 2,20}	(0; 8; 1; 1)	1,173	1,388	1,0000013
	{-2,20; -0,85; 0,00; 0,85; 2,20}	(0; 17; 1; 10)	1,003	1,974	1,0200708
	{-2,20; -0,85; 0,00; 0,85; 2,20}	(0; 107; 1; 100)	1,000	1,997	1,1664425
Ajuste 3	{-2,94; -0,85; -0,41; 1,39; 2,20}	(0; 7; 1; 0)	1,262	1,016	1
	{-2,94; -0,85; -0,41; 1,39; 2,20}	(0; 7,1; 1; 0,1)	1,258	1,032	1,0000002
	{-2,94; -0,85; -0,41; 1,39; 2,20}	(0; 8; 1; 1)	1,177	1,324	1,0000005
	{-2,94; -0,85; -0,41; 1,39; 2,20}	(0; 17; 1; 10)	1,004	1,962	1,0197431
	{-2,94; -0,85; -0,41; 1,39; 2,20}	(0; 107; 1; 100)	1,000	1,995	1,1667499

Fonte: Siqueira (1992)

Foi repetido o estudo para outro grupo de valores dos parâmetros, exceto os de γ , fixados em 0, 0,1, 1, 10 e 100 e observou-se valores de $R\hat{V}(\gamma)$ muito próximos dos valores apresentados na Tabela 4.1. Com base nessa experiência, parece não haver efeito dos parâmetros β_0 , β_1 e λ na razão de variância $R\hat{V}(\gamma)$. Verificou-se no entanto um grande acréscimo devido à estimação de λ na variância das estimativas de β_0 e β_1 .

A Tabela 4.2 mostra a razão de variâncias das estimativas das probabilidades para valores fixos das covariáveis, para o vetor de parâmetros $\theta = (\beta_0, \beta_1, \lambda, \gamma) = (0, 8, 1, 1)$.

Tabela 4.2 – Razão de Variâncias para as estimativas das probabilidades para x fixo correspondente a p no modelo de transformação logística de Box-Cox com uma covariável e uma variável indicadora (I) e com média ponderada nos pontos ajustados.

	Ajuste 1		Ajuste 2		Ajuste 3	
	Razão das Variâncias		Razão das Variâncias		Razão das Variâncias	
p	$I = 1$	$I = 0$	$I = 1$	$I = 0$	$I = 1$	$I = 0$
0,01	4,340	1,937	13,571	5,066	9,951	3,818
0,05	1,611	1,034	4,482	1,519	3,246	1,288
0,10	1,135	1,011	2,397	1,047	1,796	1,009
0,20	1,001	1,116	1,208	1,062	1,070	1,087
0,30	1,062	1,179	1,000	1,208	1,012	1,121
0,40	1,138	1,181	1,092	1,254	1,122	1,243
0,50	1,183	1,137	1,228	1,173	1,230	1,177
0,60	1,181	1,070	1,264	1,044	1,254	1,066
0,70	1,132	1,011	1,171	1,009	1,177	1,000
0,80	1,051	1,014	1,030	1,248	1,048	1,112
0,90	1,002	1,217	1,072	2,096	1,028	1,679
0,95	1,091	1,575	1,449	3,190	1,277	2,488
0,99	1,688	2,692	3,161	6,148	2,556	4,771

Fonte: Siqueira (1992)

Os resultados da Tabela 4.2 são muito similares àqueles encontrados no modelo que apresenta apenas a covariável sem a presença dos grupos.

Outra importante verificação pode ser feita se observarmos que há influência da distribuição dos pontos ajustados dos dois grupos no valor de $RV(\hat{\gamma})$. Isto pode sugerir que a intervenção nos ajustes é possível e vantajosa em diversas ocasiões.

Há muitas possíveis configurações: o uso dos mesmos pontos ajustados nos dois grupos, coincidência de algum ou diversos valores da covariável ou nenhuma coincidência.

Denotemos por $x_{(1)}$ e $x_{(0)}$ o conjunto dos valores das covariáveis correspondentes a $I = 1$ e a $I = 0$, respectivamente. Para $x_{(1)}$ fixo, variou-se $x_{(0)}$ permitindo algum grau de

coincidência, que é medido como a proporção dos valores de x comuns aos dois grupos. Considerando $x_{(1)} = \{0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 1,0\}$ e $x_{(0)} = x_{(1)} + \delta$, onde $\delta = \{0, 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 1,0\}$, na Tabela 4.3 são mostrados os resultados obtidos.

Tabela 4.3 – Razão de Variâncias em função das porcentagens de coincidência para cinco grupos de parâmetros.

Coincidência		Parâmetros: $(\beta_0, \beta_1, \gamma, \lambda)$				
δ	%	(0, 1,5, 0, 0,5)	(0, 2,5, 0,5, 1)	(0, 3,5, 0, 0,5)	(0, 7,5, 0,5, 1)	(0, 8, 1, -1)
0	100	1	1,000	1	1,001	1,004
0,1	90	1,000	1,000	1,008	1,024	1,046
0,2	80	1,000	1,001	1,015	1,034	1,055
0,3	70	1,002	1,000	1,014	1,026	1,043
0,4	60	1,007	1,000	1,009	1,013	1,029
0,5	50	1,020	1,001	1,005	1,004	1,017
0,6	40	1,040	1,002	1,004	1,000	1,005
0,7	30	1,068	1,004	1,006	1,000	1,001
0,8	20	1,105	1,010	1,014	1,001	1,008
0,9	10	1,151	1,023	1,033	1,020	1,002
1,0	0	1,207	1,046	1,072	1,115	1,013

Fonte: Siqueira e Taylor (1999)

Observa-se que as razões das variâncias da estimativa de γ são próximas de 1 se existem coincidências e tendem a crescer com o decréscimo dessa característica, embora $R\hat{V}(\lambda)$ não seja estritamente monótona em δ .

4.3 Testes de hipóteses dos efeitos de tratamentos

A comparação entre dois grupos pode ser realizada através dos testes das hipóteses $H_0: \gamma = 0$ contra $H_1: \gamma \neq 0$. Podem ser usados os testes de Wald ou da razão de

verossimilhanças entre outros, desde que seja utilizada a matriz de informação de ordem 4 associada ao vetor de parâmetros $\theta = (\beta_0, \beta_1, \lambda, \gamma)$.

A questão que surge é se os testes padrão podem ser usados para dados transformados, ignorando o fato de que a transformação foi selecionada através destes mesmos dados.

Deseja-se estabelecer condições que permitam que testes condicionais baseados no estimador de $\hat{\lambda}$ como se este fosse conhecido sejam assintoticamente equivalentes aos testes padrão. Condições no planejamento, como a distribuição de x nos dois grupos, serão investigadas.

O desempenho do teste condicional comparado a um teste incondicional foi avaliado por Siqueira e Taylor (1999) através de um estudo de simulação.

Os autores realizaram o estudo para diversas distribuições dos valores de x nos dois grupos, permitindo vários graus de coincidência variando os pontos estimados. Os dados foram gerados a partir do modelo descrito em (4.1) com $\beta_0 = 0$, $\beta_1 = 7$, $\gamma = 0$ e $\lambda = 1$, usando a sub-rotina RANUNI do pacote computacional SAS.

Os cálculos foram repetidos 1000 vezes. Experimentos balanceados (k indivíduos em cada nível de x) foram simulados com amostras de tamanhos 10, 20, 40 e 80.

Os pontos ajustados foram $x_{(1)} = \{0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 1\}$ e $x_{(0)} = x_{(1)} + \delta$, para δ com valores 0, 0,4, 0,8, 1,0.

Foram calculadas as estatísticas do teste de Wald, com $\hat{Var}(\hat{\gamma})$ obtida através da matriz de informação 4 x 4 (teste incondicional) e da matriz de informação 3 x 3 (teste condicional) admitindo $\lambda = \hat{\lambda}$ (teste condicional). Os níveis de significância obtidos no teste incondicional (TI) condicional (TC) para as amostras de $k = 10, 20, 40$ e 80 elementos em cada nível x e diversos níveis de significância fixados são apresentados na Tabela 4.4.

Os resultados da simulação, apresentados na Tabela 4.4, demonstram concordância nos níveis de significância dos testes condicional e incondicional e correspondência no nível nominal nos ajustes com 100%, 60% e 20% de coincidência. Para os ajustes com nenhuma coincidência, ou seja, com $\delta = 1$, a maioria das discrepâncias nos níveis de significância dos testes condicional e incondicional é

pequena. A maior discrepância observada ocorreu quando o nível nominal (nível de significância fixado) é de 0,05, para a amostra com 10 indivíduos.

Tabela 4.4 – Estimativas dos níveis de significância do teste de Wald condicional e incondicional para $H_0: \gamma = 0$ contra $H_1: \gamma \neq 0$.

Nível de significância fixado	k = 10		k = 20		k = 40		k = 80	
	TI	TC	TI	TC	TI	TC	TI	TC
$\delta = 0$, 100% de coincidência								
0,01	0,011	0,011	0,011	0,011	0,011	0,011	0,011	0,011
0,02	0,018	0,018	0,023	0,023	0,025	0,025	0,021	0,021
0,05	0,048	0,048	0,052	0,052	0,046	0,046	0,043	0,043
0,10	0,091	0,091	0,091	0,091	0,098	0,098	0,083	0,083
$\delta = 0,4$, 60% de coincidência								
0,01	0,010	0,012	0,010	0,011	0,011	0,012	0,018	0,018
0,02	0,014	0,017	0,023	0,024	0,018	0,020	0,027	0,029
0,05	0,044	0,047	0,053	0,056	0,041	0,043	0,068	0,068
0,10	0,103	0,107	0,094	0,098	0,087	0,091	0,113	0,117
$\delta = 0,8$, 20% de coincidência								
0,01	0,012	0,012	0,013	0,013	0,008	0,008	0,009	0,009
0,02	0,021	0,022	0,023	0,024	0,023	0,023	0,024	0,024
0,05	0,048	0,049	0,045	0,045	0,060	0,060	0,053	0,054
0,10	0,101	0,103	0,091	0,091	0,114	0,114	0,092	0,092
$\delta = 1$, 0% de coincidência								
0,01	0,008	0,016	0,013	0,018	0,013	0,018	0,007	0,010
0,02	0,021	0,033	0,028	0,033	0,022	0,031	0,019	0,024
0,05	0,045	0,062	0,056	0,070	0,051	0,063	0,060	0,069
0,10	0,118	0,133	0,115	0,128	0,099	0,114	0,123	0,134

Fonte: Siqueira e Taylor (1999)

As simulações foram repetidas para o caso em que $\gamma = 0,05$ e os demais parâmetros foram mantidos. Os resultados encontram-se na Tabela 4.5, que apresenta uma estimativa do poder de ambos os testes.

Tabela 4.5 –Proporções de rejeições usando o teste de Wald incondicional e condicional para $H_0: \gamma = 0$ contra $H_1: \gamma \neq 0$, quando $\gamma = 0,05$.

Nível de significância fixado	k = 10		k = 20		k = 40		k = 80	
	TI	TC	TI	TC	TI	TC	TI	TC
$\delta = 0$, 100% de coincidência								
0,01	0,043	0,044	0,141	0,144	0,332	0,332	0,670	0,671
0,02	0,070	0,073	0,197	0,200	0,424	0,424	0,742	0,742
0,05	0,163	0,167	0,321	0,321	0,568	0,568	0,847	0,847
0,10	0,264	0,264	0,435	0,435	0,683	0,683	0,908	0,908
$\delta = 0,4$, 60% de coincidência								
0,01	0,058	0,061	0,140	0,145	0,340	0,344	0,694	0,698
0,02	0,092	0,098	0,207	0,210	0,429	0,433	0,773	0,781
0,05	0,179	0,183	0,331	0,336	0,576	0,580	0,867	0,871
0,10	0,262	0,270	0,452	0,458	0,685	0,689	0,926	0,928
$\delta = 0,8$, 20% de coincidência								
0,01	0,049	0,049	0,097	0,097	0,187	0,188	0,486	0,486
0,02	0,076	0,076	0,146	0,146	0,259	0,259	0,578	0,579
0,05	0,151	0,152	0,273	0,274	0,405	0,406	0,709	0,710
0,10	0,252	0,254	0,380	0,380	0,538	0,539	0,815	0,816
$\delta = 1$, 0% de coincidência								
0,01	0,018	0,034	0,050	0,065	0,085	0,109	0,173	0,213
0,02	0,039	0,057	0,078	0,100	0,128	0,155	0,254	0,291
0,05	0,092	0,127	0,146	0,170	0,221	0,246	0,393	0,436
0,10	0,188	0,219	0,230	0,268	0,322	0,325	0,516	0,550

Fonte: Siqueira e Taylor (1999)

Pela Tabela 4.5, podemos perceber que as diferenças de poder entre os testes incondicional e condicional são insignificantes exceto no caso em que não há coincidência.

Os resultados da simulação mostram que, para ajustes com um montante razoável de coincidência entre os dois grupos, o teste incondicional é essencialmente equivalente ao condicional em termos de nível de significância e poder. Além disso, quando não há coincidência, o teste condicional pode ter nível de significância ligeiramente maior, ou seja, pode rejeitar indevidamente a hipótese nula com maior frequência que o teste incondicional.

Estes resultados são consistentes com aqueles apresentados na Seção 4.2, que mostrou uma inflação muito pequena na variância de $\hat{\gamma}$ associada à estimação de λ quando existe coincidência no ajuste dos dois grupos. Portanto, conclui-se que para propósitos práticos, torna-se viável o uso do teste condicional ao invés do teste incondicional para testar o efeito de tratamento.

Para exemplificar, os autores aplicaram os testes ao conjunto de dados descrito por Collet (1991). Os dados presentes na Tabela 4.6 são referentes ao registro da mortalidade de uma espécie de mariposa exposta a um veneno após 72 horas de exposição.

Tabela 4.6 – Número de mortes em uma amostra de 20 mariposas 72 horas após a exposição ao veneno.

Dose de veneno (μg)	Sexo da mariposa	
	Masculino	Feminino
1	1	0
2	4	2
4	9	6
8	13	10
16	18	12
32	20	16

O fator grupo incorporado ao modelo é sexo, com doses idênticas para os dois grupos.

As estimativas dos parâmetros do modelo logístico padrão ($\log \frac{p}{1-p} = \beta_0 + \beta_1 x + \mathcal{M}$) e o modelo usando a transformação de Box-Cox ($\log \frac{p}{1-p} = \beta_0 + \beta_1 x^{(\lambda)} + \mathcal{M}$) são apresentadas na Tabela 4.7.

Tabela 4.7 – Estimativas dos parâmetros para os modelos de resposta binária.

Parâmetro	Modelo 1 – Modelo Logístico Padrão	Modelo 2 – Modelo usando transformação de Box-Cox
β_0 (intercepto)	-2,135	-4,006
β_1 (dose)	0,160	2,390
γ (sexo)	0,969	1,099
λ (transformação)		-0,250

O valor máximo do logaritmo da função de verossimilhança para o modelo saturado é $\hat{L}_s = -103,242$, para o Modelo 1 é $\hat{L}_1 = -117,226$ e $\hat{L}_2 = -106,102$ para o Modelo 2. Verifica-se então uma melhora substancial na qualidade do ajuste quando é usado o modelo 2 (Função Desvio do Modelo ou *deviance* $D = 2[\hat{L}_s - \hat{L}_2] = 5,719$, com nível descritivo do teste de falta de ajuste $p = 0,678$), ao invés do modelo 1 ($D = 27,968$, $p = 0,0009$).

A estimativa da matriz de covariância dos estimadores dos parâmetros ($\beta_0, \beta_1, \lambda, \gamma$) é dada por:

$$\begin{bmatrix} 0,564 & -0,622 & -0,099 & 0,118 \\ & 0,935 & 0,020 & -0,211 \\ & & 0,126 & 0,002 \\ & & & 0,053 \end{bmatrix}$$

e a estimativa dessa matriz considerando $\hat{\lambda} = \lambda$ é:

$$\begin{bmatrix} 0,297 & -0,147 & -0,104 \\ & 0,091 & 0,029 \\ & & 0,126 \end{bmatrix}.$$

Os resultados dos testes incondicional e condicional são dados na Tabela 4.8.

Tabela 4.8 – Resultados do teste de Wald para testes incondicional e condicional.

Parâmetro	Teste incondicional		Teste condicional	
	Estatística de Wald	p-valor	Estatística de Wald	p-valor
β_0 (intercepto)	28,462225	0,000	54,066609	0,00
β_1 (dose)	6,105841	0,0135	63,091249	0,00
γ (sexo)	9,572836	0,002	9,579025	0,002

Observa-se que conforme esperado, há uma diferença muito pequena entre o nível descritivo associado ao teste da influência do fator sexo nas análises condicional e incondicional.

Neste capítulo foi apresentado o modelo logístico com transformação de Box-Cox com dois tratamentos ou dois grupos a serem comparados. Através deste modelo, foi apresentado o custo da estimação do parâmetro adicional no modelo e testes de hipóteses sobre efeitos de tratamentos. No próximo capítulo, serão apresentados testes de hipóteses dos parâmetros do modelo de regressão logística de Box-Cox.

Capítulo 5

Testes de hipóteses dos parâmetros do modelo de regressão logística de Box-Cox

5.1 Introdução

Neste capítulo serão apresentados testes de hipóteses dos parâmetros no modelo de regressão logística de Box-Cox com apenas uma covariável.

Neste caso, o estudo consiste em avaliar os níveis de significância dos testes assintóticos padrão aplicados a dados transformados.

Serão discutidos dois testes de hipóteses relativos aos parâmetros β_1 e λ neste modelo.

5.2 Teste da necessidade do uso da transformação

Uma vez adotada a transformação na covariável, é necessária a verificação da necessidade do uso da mesma. O uso da transformação é necessário se $\lambda \neq 1$ e a análise é feita através da comparação entre o modelo logístico de Box-Cox e o modelo logístico padrão. Formalmente, o problema se reduz a testar as hipóteses $H_0: \lambda = 1$ contra $H_1: \lambda \neq 1$. Um procedimento equivalente seria verificar se o intervalo de confiança para λ contém o valor 1, caso em que a transformação é desnecessária.

Um intervalo de confiança aproximado para λ pode ser facilmente obtido através do gráfico do logaritmo da função de verossimilhança contra λ .

Um intervalo aproximado de $100(1 - \alpha)\%$ de confiança para λ é o conjunto formado pelos valores de λ que satisfazem a inequação

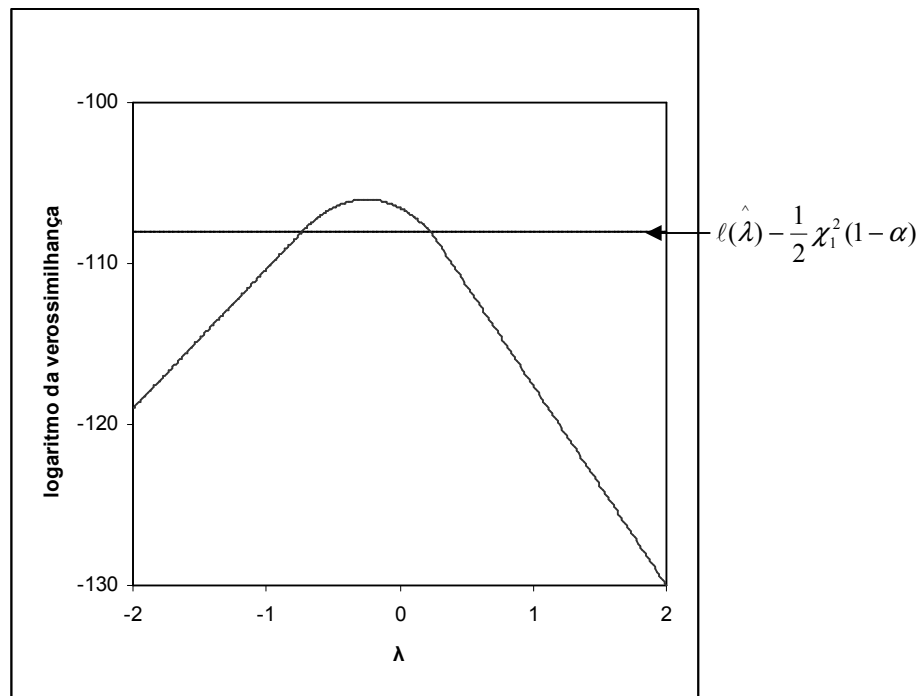
$$\ell(\hat{\lambda}) - \ell(\lambda) \leq \frac{1}{2} \chi_1^2(1 - \alpha)$$

onde $\ell(\lambda)$ é o logaritmo da função de verossimilhança para um valor fixo de λ , $\ell(\hat{\lambda})$ é o valor do logaritmo da função de verossimilhança calculado para $\lambda = \hat{\lambda}$ e $\chi_1^2(1 - \alpha)$ é o quartil de ordem $(1 - \alpha)$ da distribuição qui-quadrado com 1 grau de liberdade. Portanto, um intervalo de confiança para λ pode ser construído através do gráfico de $\ell(\lambda)$ contra λ .

A linha horizontal no nível $\ell(\hat{\lambda}) - \frac{1}{2} \chi_1^2(1 - \alpha)$ da escala vertical encontra a curva em dois valores de λ e as projeções desses pontos no eixo horizontal são os extremos do intervalo de confiança. Como resultado, é determinado um intervalo de confiança aproximado para λ .

No exemplo das mariposas apresentado na Seção 4.3, o intervalo de 95% confiança para λ é $(-0,73, 0,23)$, o que confirma a necessidade do uso da transformação da dose. A Figura 5.1 mostra o gráfico do logaritmo da função de verossimilhança e intervalo de 95% de confiança para λ .

Figura 5.1 – Gráfico do logaritmo da verossimilhança em função de λ e o intervalo de 95% de confiança para λ .



5.3 Teste do efeito da covariável

Consideremos o modelo:

$$\log \frac{p(x)}{1-p(x)} = \beta_0 + \beta_1 x^{(\lambda)} \quad (5.1)$$

em que

$$x^{(\lambda)} = \begin{cases} \frac{x^\lambda - 1}{\lambda} & , \lambda \neq 0 \\ \log x & , \lambda = 0. \end{cases}$$

O efeito da covariável na resposta pode ser testado por $H_0: \beta_1 = 0$ contra $H_1: \beta_1 \neq 0$ (ou como $H_0: \beta_1 > 0$ contra $H_1: \beta_1 < 0$). A hipótese nula refere-se a inexistência de efeito da covariável e a hipótese alternativa corresponde a existência de tal efeito. Note que a hipótese $H_0: \beta_1 = \beta$, em que β é um valor fixo e não nulo, não tem significado prático porque a magnitude de β depende da dimensão de λ . Isto significa que a interpretação usual do coeficiente β seria válida somente se λ fosse conhecido.

Por outro lado, sob a hipótese $H_0: \beta_1 = 0$, o modelo (5.1) não é identificável pois diferentes valores de λ ($\lambda_1 \neq \lambda_2$) resultariam no mesmo valor de $p = \{1 + \exp(-\beta_0)\}^{-1}$ e conseqüentemente, na mesma função de verossimilhança. Desta forma, o parâmetro λ estaria presente somente no modelo sob a hipótese alternativa. Neste caso, estatísticas dos testes assintóticos usuais não são apropriadas devido à dificuldade de obter suas distribuições assintóticas. Por exemplo, se os testes do score ou da razão de verossimilhanças usual forem usados, não fica claro qual seria o número correto de graus de liberdade ou mesmo se a aproximação pela distribuição qui-quadrado é apropriada.

Como não há muitos estudos sobre testes de hipóteses onde um parâmetro extra está presente somente na hipótese alternativa, Siqueira (1992) sugere a abordagem proposta por Davies (1977, 1987). Em ambos os artigos, o autor discute o problema em um caso geral e trata a estatística do teste como um processo estocástico. No primeiro artigo, é considerado que a estatística do teste obedece a uma distribuição normal. No segundo, o autor descreve um meio simples de encontrar o nível de significância do teste e estende as derivações para o caso em que a estatística do teste tem distribuição qui-quadrado. Apresentamos a seguir a descrição dos testes estudados.

Suponha que para um λ fixo, a distribuição assintótica da estatística do teste $S(\lambda)$ é normal padrão. Para λ variando em um particular intervalo ($L \leq \lambda \leq U$), $S(\lambda)$ pode ser considerado um processo Gaussiano.

Para testar as hipóteses $H_0: \beta_1 = 0$ contra $H_1: \beta_1 > 0$, Davies (1977) sugeriu o teste que rejeita H_0 se:

$$S = \sup_{L \leq \lambda \leq U} S(\lambda) > c. \quad (5.2)$$

Embora seja suposto que $S(\lambda)$ tem distribuição normal padrão, não há motivo para que a distribuição de S seja normal padrão. Por outro lado, fixado um nível de significância, o valor de c seria conhecido somente se a distribuição de S fosse conhecida. A distribuição desta estatística pode ser determinada em alguns casos particulares, mas não de modo geral. Conseqüentemente, não é possível calcular o nível de significância exato,

$$P[\sup_{L \leq \lambda \leq U} S(\lambda) > c(\lambda)]. \quad (5.3)$$

Contudo, um limite para esta probabilidade pode ser calculado usando o número esperado de “passagens pelo zero” do processo $S(\lambda) - c(\lambda)$, sob a hipótese nula.

De acordo com Davies (1977), generalizando, a função $c(\lambda)$ ao invés de c é usada em (5.3), assim o processo $S(\lambda) - c(\lambda)$ é considerado.

Definindo os eventos:

$$A = \left\{ \sup_{L \leq \lambda \leq U} S(\lambda) - c(\lambda) > 0 \right\},$$

$$B = S(L) - c(L) < 0,$$

Z = número de “passagens pelo zero” do processo $S(\lambda) - c(\lambda)$,

notamos que $A \cap B$ corresponde ao evento que aconteça pelo menos uma “passagem pelo zero”. O objetivo é calcular $P[A]$.

O número esperado de “passagens pelo zero” é dado por:

$$E[Z] = \sum_{i=0}^{\infty} iP[Z = i] = \sum_{i=1}^{\infty} iP[Z = i] \geq \sum_{i=1}^{\infty} P[Z = i] = P[A \cap B].$$

Logo,

$$P[A \cap B] \leq E[Z]. \quad (5.4)$$

Como $P[A] = \Pr[A \cap B] + P[A \cap B^C]$ e $P[A \cap B^C] \leq P[B^C]$, isto implica que $P[A] \leq P[A \cap B] + P[B^C]$. Além disso, $P[B^C] = P(S(L) \geq c(L)) = \Phi(-c(L))$, onde Φ é a função acumulada da distribuição normal.

$$P[A] \leq E[Z] + \Phi(-c(L)). \quad (5.5)$$

De acordo com resultados para a variável aleatória número de “passagens pelo zero”, presentes em Cramér e Leadbetter (1967), citado em Davies (1977), sob H_0 ,

$$E(Z) = \frac{1}{2\pi} \int_L^U \frac{1}{[-\rho_{11}(\lambda)]^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}[c(\lambda)]^2\right\} \psi\left\{\frac{c_1(\lambda)}{[-\rho_{11}(\lambda)]^{\frac{1}{2}}}\right\} d\lambda, \quad (5.6)$$

em que

$$c_1(\lambda) = \frac{\partial}{\partial \lambda} c(\lambda), \quad \rho_{11}(\lambda) = \frac{\partial^2}{\partial \phi^2} \rho(\phi, \lambda) \Big|_{\phi=\lambda}, \quad \psi(x) = \exp\left(-\frac{1}{2}x^2\right) - x \int_x^{\infty} \exp\left(-\frac{1}{2}t^2\right) dt$$

e $\rho(\phi, \lambda) = \text{cov}(S(\phi), S(\lambda))$.

Por (5.4), (5.5) e (5.6) temos que:

$$P[A] \leq \Phi(-c(L)) + \frac{1}{2\pi} \int_L^U [-\rho_{11}(\lambda)]^{\frac{1}{2}} \exp\left\{-\frac{1}{2}[c(\lambda)]^2\right\} \psi\left\{\frac{c_1(\lambda)}{[-\rho_{11}(\lambda)]^{\frac{1}{2}}}\right\} d\lambda.$$

Assim, fixada uma constante c , sob H_0 ,

$$P\left[\sup_{L \leq \lambda \leq U} S(\lambda) > c\right] \leq \Phi(-c) + \frac{1}{2\pi} \exp\left\{-\frac{c^2}{2}\right\} \int_L^U [-\rho_{11}(\lambda)]^{\frac{1}{2}} d\lambda. \quad (5.7)$$

A expressão (5.7) depende da função de auto-correlação do processo $(\rho_{11}(\lambda))$ e de acordo com Davies (1987), nem sempre pode ser calculada analiticamente. Sugere então que, nos casos mais complicados, se utilize o procedimento a seguir. Se $T(\lambda) = \frac{\partial S(\lambda)}{\partial \lambda}$,

teremos que $Var[T(\lambda)] = -\rho_{11}(\lambda)$ e $[-\rho_{11}(\lambda)]^{\frac{1}{2}} = \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} E[|T(\lambda)|]$. Conseqüentemente,

$\int_L^U [-\rho_{11}(\lambda)]^{\frac{1}{2}} d\lambda$ é proporcional a

$$\int_L^U E[|T(\lambda)|] d\lambda \quad (5.8)$$

que pode ser estimada por

$$V = \int_L^U |T(\lambda)| d\lambda \quad (5.9)$$

$$= |S(\lambda_1) - S(L)| + |S(\lambda_2) - S(\lambda_1)| + \dots + |S(U) - S(\lambda_n)|, \quad (5.10)$$

em que $\lambda_1, \lambda_2, \dots, \lambda_n$ são os sucessivos pontos onde $\frac{\partial S}{\partial \lambda}$ muda de sinal.

Assim, a estimativa do nível de significância é dada por

$$W = \Phi(-M) + V \exp\left(\frac{1}{2} M^2\right) / (8\pi)^{\frac{1}{2}}, \quad (5.11)$$

em que M é o máximo de $S(\lambda)$ no conjunto $L < \lambda < U$ e V é dado pela expressão (5.10). O autor sugere ainda que os valores de V dessa expressão sejam obtidos numa simulação preliminar e sua média utilizada na expressão (5.11).

Um teste ao nível α consistiria em rejeitar a hipótese nula se $W < \alpha$. A expressão (5.11) é um limite superior no nível de significância e pode ser maior que 1, já que a

aproximação usada na derivação tem como base o número esperado de “passagens pelo zero”.

De acordo com Siqueira (1992), o nível de significância pode ser facilmente obtido do gráfico de $S(\lambda)$ como função de λ , se esse gráfico tiver como característica uma curva suave, ou seja, não apresentar picos, caso em que a aproximação usada para obter o nível de significância não é razoável. Estes gráficos usualmente mostram dois ou três pontos em que $\frac{\partial S}{\partial \lambda}$ muda de sinal no intervalo $-3 \leq \lambda \leq 3$.

No teste de hipóteses bilateral, uma possível região crítica seria da forma:

$$S_2 = \sup_{L \leq \lambda \leq U} S(\lambda) > c$$

e analogamente pode ser mostrado que o limite superior no nível de significância é dado por

$$2 \left\{ \Phi(-M_2) + V \exp\left(-\frac{1}{2} M_2^2\right) / (8\pi)^{\frac{1}{2}} \right\} \quad (5.12)$$

onde M_2 é o máximo de $|S(\lambda)|$ para $L \leq \lambda \leq U$ e V é a expressão (5.10).

Davies (1987) apresentou derivações detalhadas para o caso onde, para λ fixo, a estatística de teste é da forma $S(\lambda) = Z_1^2 + Z_2^2 + \dots + Z_s^2$. O teste da razão de verossimilhanças é um exemplo de um processo com distribuição qui-quadrado.

Usando argumentos similares ao do caso com distribuição normal, Davies (1987) mostrou que o limite superior do nível de significância neste caso é dado por

$$P[\chi_s^2 > M] + VM^{\frac{1}{2}(s-1)} e^{-\frac{1}{2}M} 2^{-\frac{1}{2}s} / \Gamma\left(\frac{1}{2}s\right), \quad (5.13)$$

em que M é o máximo de $S(\lambda)$, $V = \int_L^U \left| \partial S^{\frac{1}{2}}(\lambda) / \partial \lambda \right| d\lambda$, Γ representa a função Gama e χ_s^2

é uma variável aleatória com distribuição qui-quadrado com s graus de liberdade.

A expressão V pode ser calculada por

$$V = \left| S^{\frac{1}{2}}(\lambda_1) - S^{\frac{1}{2}}(L) \right| + \left| S^{\frac{1}{2}}(\lambda_2) - S^{\frac{1}{2}}(\lambda_1) \right| + \dots + \left| S^{\frac{1}{2}}(U) - S^{\frac{1}{2}}(\lambda_n) \right|, \quad (5.14)$$

em que $\lambda_1, \lambda_2, \dots, \lambda_n$ são os sucessivos pontos onde a derivada de $S^{\frac{1}{2}}(\lambda)$ muda de sinal.

Para os testes definidos em (5.11) ou (5.12), $S(\lambda)$ é a estatística do teste de Wald, isto é, $S(\lambda) = \hat{\beta}(\lambda) / SE(\hat{\beta}(\lambda))$, onde $SE(\hat{\beta}(\lambda))$ é obtido a partir da matriz de informação 2×2 . Para o teste definido em (5.13), $S(\lambda)$ é a estatística do teste da razão de verossimilhanças dado por $2(\sup_{\Theta} \ell - \sup_{\Theta_0} \ell)$, onde ℓ é o logaritmo da função de verossimilhança, $\Theta = (A, B, \lambda)$ e $\Theta_0 = (A, \beta = 0, \lambda)$ são os espaços paramétricos. Para ambos os testes, (L, U) é tomado como o intervalo $(-3, 3)$.

Siqueira (1992) fez um estudo de simulação para avaliar o desempenho do teste de Davies. Foram utilizados 13 pontos ajustados para x : 0,01, 0,05, 0,10, 0,20, 0,30, 0,40, 0,50, 0,60, 0,70, 0,80, 0,90, 0,95 e 0,99 e variou-se p , o número de indivíduos para cada nível de x . Dados foram simulados de acordo com o modelo (5.1), com $\beta_0 = 1$ e $\beta_1 = 0$. As respostas binárias foram geradas através da sub-rotina RANUNI do SAS. Foram simulados experimentos balanceados com tamanhos amostrais diferentes para verificar o efeito do tamanho da amostra no teste de Davies. Os cálculos foram repetidos 500 vezes. O nível de significância verdadeiro foi calculado para a estatística de Wald e para o teste da razão de verossimilhanças para três casos: usando valores individuais de V , dados pelas expressões (5.10) ou (5.14), a média de V obtida por uma simulação preliminar e através de um teste aproximado.

Este teste aproximado é construído ignorando o fato de que sua estatística é o supremo de $S(\lambda)$ no conjunto de possíveis valores de λ . Desta forma, para a estatística de Wald e a estatística da razão de verossimilhanças, são usados os valores críticos usuais, da distribuição normal padrão e da distribuição qui-quadrado com um grau de liberdade.

A estatística do teste da razão de verossimilhança é dada por $2 \left[\sup_{\Theta} \ell - \sup_{\Theta_0} \ell \right]$,

em que $\Theta = (A, B, \lambda)$ e $\Theta_0 = (A, \beta_1 = 0, \lambda = \lambda_0)$ e λ_0 é qualquer valor fixo de λ . Como a diferença entre o número de parâmetros nestes dois espaços paramétricos é de 2, um teste aproximado natural para $\beta_1 = 0$ é baseada na distribuição qui-quadrado com dois graus de liberdade.

Os resultados das simulações mostraram que o nível de significância atingido tende a ser maior que o nível fixado. Observa-se efeito do tamanho da amostra de modo que o teste de Davies não é recomendado para amostras pequenas, mas seu desempenho

é substancialmente melhorado quando o número de indivíduos em cada nível de x é aumentado.

Em geral, os resultados do teste aproximado unilateral são piores que aqueles obtidos no teste de Davies, especialmente para casos em que a amostra é pequena. O mesmo acontece para o teste aproximado bilateral, tanto usando o valor crítico da distribuição normal padrão quanto na distribuição qui-quadrado com um grau de liberdade. Porém, se for utilizada o valor crítico da distribuição qui-quadrado com dois graus de liberdade, as discrepâncias não são tão notáveis, especialmente em amostras grandes. Além disso, o teste aproximado usando a distribuição qui-quadrado com dois graus de liberdade é claramente melhor que o teste aproximado adotando a distribuição qui-quadrado com um grau de liberdade.

Para o teste bilateral, os resultados obtidos usando a estatística de Wald são ligeiramente melhores do que os da razão de verossimilhanças, especialmente para amostras pequenas e médias. Ambas as estatísticas apresentam resultados iguais para amostras grandes.

Em resumo, o teste aproximado para $H_0: \beta_1 = 0$ apresentado tem desempenho razoável para amostras de tamanhos médios e grandes, mas não é recomendado quando a amostra é pequena.

Conclui-se então que o teste de Davies é uma possível opção especialmente quando é preciso tomar uma rápida decisão de verificação da significância da covariável. Como algumas deficiências foram detectadas, um teste mais eficiente precisaria ser desenvolvido para avaliar o efeito da covariável no modelo logístico de Box-Cox.

Pudemos analisar neste capítulo os testes de efeito da covariável transformada nos modelos logísticos de Box-Cox no que diz respeito aos parâmetros β_1 e λ . Também foi visto o teste para verificar a real necessidade da transformação da covariável através da comparação entre o modelo logístico padrão e o modelo logístico de Box-Cox.

No próximo capítulo, serão apresentados métodos de diagnóstico para avaliar a influência das observações nas estimativas de λ e a utilização destes métodos em um conjunto de dados com resposta binária.

Capítulo 6

Análise de diagnóstico no modelo de regressão logística de Box-Cox

6.1 Introdução

Transformações em covariáveis são aplicadas em análise de regressão por vários motivos e famílias de transformações paramétricas, como a de Box-Cox, são muito comuns. De acordo com Lee e Yick (1999), quando uma transformação paramétrica é usada, o estimador de máxima verossimilhança do parâmetro da transformação é frequentemente sensível a pequenas perturbações dos dados. A maioria dos métodos de diagnóstico nessa linha trata do uso de transformações na variável resposta ou em ambas, variável resposta e covariável. Além disso, o estudo restringe-se a modelos lineares.

Lee e Yick (1999) desenvolvem técnicas de diagnóstico quando são utilizadas transformações na covariável em modelos lineares generalizados.

Neste capítulo, serão apresentados dois métodos de diagnóstico propostas pelos autores e a sua aplicação a um conjunto de dados com resposta binária.

6.2 Métodos de diagnóstico

Serão apresentados nessa seção os métodos de diagnóstico: diferença de verossimilhanças e influência parcial para o modelo de regressão logística de Box-Cox.

Diferença de verossimilhanças

Seja $y = (y_1, \dots, y_n)^T$ o vetor de respostas com densidade:

$$f_{y_i}(y_i; \theta) = \exp\{[y_i \theta_i - b(\theta_i)] / a(\phi) + c(y_i, \phi)\}$$

com $\theta_i = k(\eta_i)$, em que η_i é o preditor linear e $a(\cdot)$, $b(\cdot)$, $c(\cdot)$ são funções conhecidas. Sem perda de generalidade, o parâmetro de dispersão ϕ é supostamente conhecido ou é substituído por um estimador $\hat{\phi}$ e escrito como $\hat{a} = a(\hat{\phi})$, que fornece uma densidade da família exponencial com parâmetro natural θ . O logaritmo da função de verossimilhança é então dada por:

$$\hat{a}^{-1} \sum_{i=1}^n [y_i k(\eta_i) - b\{k(\eta_i)\}].$$

Consideremos X uma matriz de covariáveis do modelo, $n \times (p + q)$, particionada na forma:

$$X = [\mathbf{x}, \mathbf{z}] = [x_1 \dots x_p, z_1 \dots z_q],$$

com $x_1 \dots x_p, z_1 \dots z_q$ correspondentes às colunas da matriz X .

Um melhor ajuste para o modelo linear generalizado proposto pode ser freqüentemente obtido transformando uma ou mais das covariáveis z presentes em $X = [\mathbf{x}, \mathbf{z}]$.

Seja o preditor linear do modelo transformado:

$$\eta = x\delta + G(z, \lambda)\xi,$$

em que $G(z, \lambda) = [g_1(z_{(1)}, \lambda_1), \dots, g_q(z_{(q)}, \lambda_q)]$ é uma matriz $n \times q$ e

$$g_j(z_{(j)}, \lambda_j) = \langle g_j(z_{(1j)}, \lambda_j), \dots, g_j(z_{(nj)}, \lambda_j) \rangle^T$$

representa uma família de transformação conhecida, duplamente diferenciável, indexada por λ_j ($j = 1, \dots, q$). e δ e ξ são vetores de parâmetros, p e q dimensionais. Temos aqui um interesse especial no vetor de parâmetros λ .

Sejam $L(\lambda, \delta, \xi)$ o logaritmo da função de verossimilhança, $\bar{\delta}(\lambda)$ e $\bar{\xi}(\lambda)$ funções que maximizam $L(\lambda, \delta, \xi)$, para λ fixo e $L(\lambda; \bar{\delta}(\lambda), \bar{\xi}(\lambda))$ o correspondente máximo obtido. Para avaliar a influência de casos individuais na estimativa de máxima verossimilhança $\hat{\lambda}$ de λ , a diferença entre $\hat{\lambda}$ e $\hat{\lambda}_{[i]}$, a estimativa de máxima verossimilhança sem o caso i , pode ser mensurada através da medida diferença de verossimilhanças dada por:

$$LD_i = 2[L(\hat{\lambda}) - L(\hat{\lambda}_{[i]})],$$

em que $L(\lambda) = L(\lambda; \bar{\delta}(\lambda), \bar{\xi}(\lambda))$.

Para o caso especial de regressão linear, Wei e Hickernell (1996) (citado por Lee e Yick (1999)) obtiveram aproximações para simplificar os cálculos dos valores de LD_i .

Um grande valor LD_i indica que provavelmente a estimativa $\hat{\lambda}$ é fortemente dependente do caso i .

Análise de influência parcial

Com a notação já introduzida, consideremos novamente o preditor linear antes da transformação da covariável z , dado por

$$\eta_0 = x\beta + z\gamma. \quad (6.1)$$

O modelo após a transformação é:

$$\eta = x\beta + G(z, \lambda)\xi \quad (6.2)$$

de modo que $z = G(z, \lambda_0)$ representaria a inexistência de transformação.

Um teste de hipótese $H_0: \lambda = \lambda_0$ pode ser baseado em $D_0 - D$, a redução da função desvio do modelo (6.1) para o modelo (6.2). É importante verificar se qualquer observação tem um impacto indevido neste teste. De modo geral, a função desvio de um particular modelo é definida como $D = -2 \ln \left[\frac{L}{L_0} \right]$, em que L é a função de verossimilhança desse modelo calculada nos estimadores de máxima verossimilhança e L_0 é a correspondente função para o modelo saturado. Denotando-se $D_{0[i]}$ e $D_{[i]}$ as funções desvio do modelo (6.1) e (6.2) respectivamente após excluir o caso i , Lee (1988) (citado por Lee e Yick (1999)) sugeriu uma medida de influência parcial para o impacto do caso i na transformação como:

$$d_i = (D_0 - D) - (D_{0[i]} - D_{[i]}),$$

que compara reduções no desvio devido à transformação com e sem a i -ésima observação. Novamente aqui, grandes valores de d_i implicam em grande alteração da redução no desvio com a retirada da i -ésima observação.

6.3 Aplicação

Neter, Kutner, Nachtsheim e Wasserman (1996) apresentam um conjunto de dados relativo à variável resposta compra de carro (sim ou não) em função de algumas características da família e que será analisado a seguir.

Uma montadora de veículos contrata uma empresa de pesquisa de mercado na condução de um estudo piloto para analisar a probabilidade de uma família comprar um carro novo durante o ano seguinte. Uma amostra casual de 33 famílias foi selecionada e foram coletados dados da renda familiar anual (X , em milhares de dólares). Doze meses depois, foi conduzida uma entrevista para determinar se, no decorrer do ano, houve compra de um veículo novo ($Y = 1$) ou não ($Y = 0$). Estes dados são apresentados na Tabela 6.1.

Tabela 6.1 – Dados relativos às 33 famílias.

Família (i)	Y_i	x_i
1	0	32
2	0	45
3	1	60
4	0	53
5	0	25
6	1	68
7	1	82
8	1	38
9	0	67
10	1	92
11	1	72
12	0	21
13	0	26
14	1	40
15	0	33
16	0	45
17	1	61
18	0	16
19	1	18
20	0	22
21	0	27
22	1	35
23	1	40
24	0	10
25	0	24
26	1	15
27	0	23
28	0	19
29	1	22
30	0	61
31	0	21
32	1	32
33	0	17

Com base nos dados apresentados na Tabela 6.1, serão feitas as análises de diagnóstico apresentadas na seção anterior. Será ajustado o modelo (3.10), $\log\left[\frac{p(x)}{1-p(x)}\right] = \beta_0 + \beta_1 x^{(\lambda)}$. Para estimar seus parâmetros, conforme descrito na Seção 3.3, serão fixados valores de λ no intervalo $[-2, 2]$: -2; -1,5; -1; -0,5; 0,5; 1; 1,5; 2 e para cada um deles, obteremos as estimativas de máxima verossimilhança de β_0 e β_1 . Calcula-se então o logaritmo da função de verossimilhança com base no valor de λ fixado e $\hat{\beta}_0$, $\hat{\beta}_1$, estimativas de máxima verossimilhança obtidas para cada um dos valores de λ . A estimativa de máxima verossimilhança de λ é o valor desse parâmetro que maximiza o logaritmo da função de verossimilhança.

Foi utilizado o programa estatístico R, para o ajuste do modelo com base nos dados. O modelo ajustado é $\log\left[\frac{\hat{p}(x)}{1-\hat{p}(x)}\right] = -1,209212 + 0,000974x^{(2)}$.

Após esse procedimento, para efeito de diagnóstico, retira-se cada uma das observações e repete-se o cálculo feito anteriormente. Com isso, é possível calcular cada uma das estimativas de máxima verossimilhança de λ sem o caso i , o que permitirá o cálculo da diferença de verossimilhanças,

$$DV_i = 2[L(\hat{\lambda}) - L(\hat{\lambda}_{[i]})].$$

Adicionalmente, obtivemos a medida $DV_{(i)} = 2[L(\hat{\lambda}) - L_{(i)}(\hat{\lambda}_{[i]})]$, onde $L_{(i)}(\hat{\lambda}_{[i]})$ é a função de verossimilhança com base nas $n - 1$ observações restantes, calculada no ponto $\hat{\lambda}_{[i]}$.

A Tabela 6.2 apresenta as estimativas de máxima verossimilhança de λ quando retiradas cada uma das observações e os valores de DV_i e $DV_{(i)}$.

Como pode ser observado na Tabela 6.2, os valores mais influentes são relativos às famílias 9, 19, 26 e 30, quando o cálculo é feito por $2[L(\hat{\lambda}) - L(\hat{\lambda}_{(i)})]$ e para o cálculo de $2[L(\hat{\lambda}) - L_{(i)}(\hat{\lambda}_{(i)})]$, os dados influentes são das famílias 9, 19, 26 e 29. Observa-se que são famílias que compraram automóvel e têm renda familiar baixa (caso das famílias

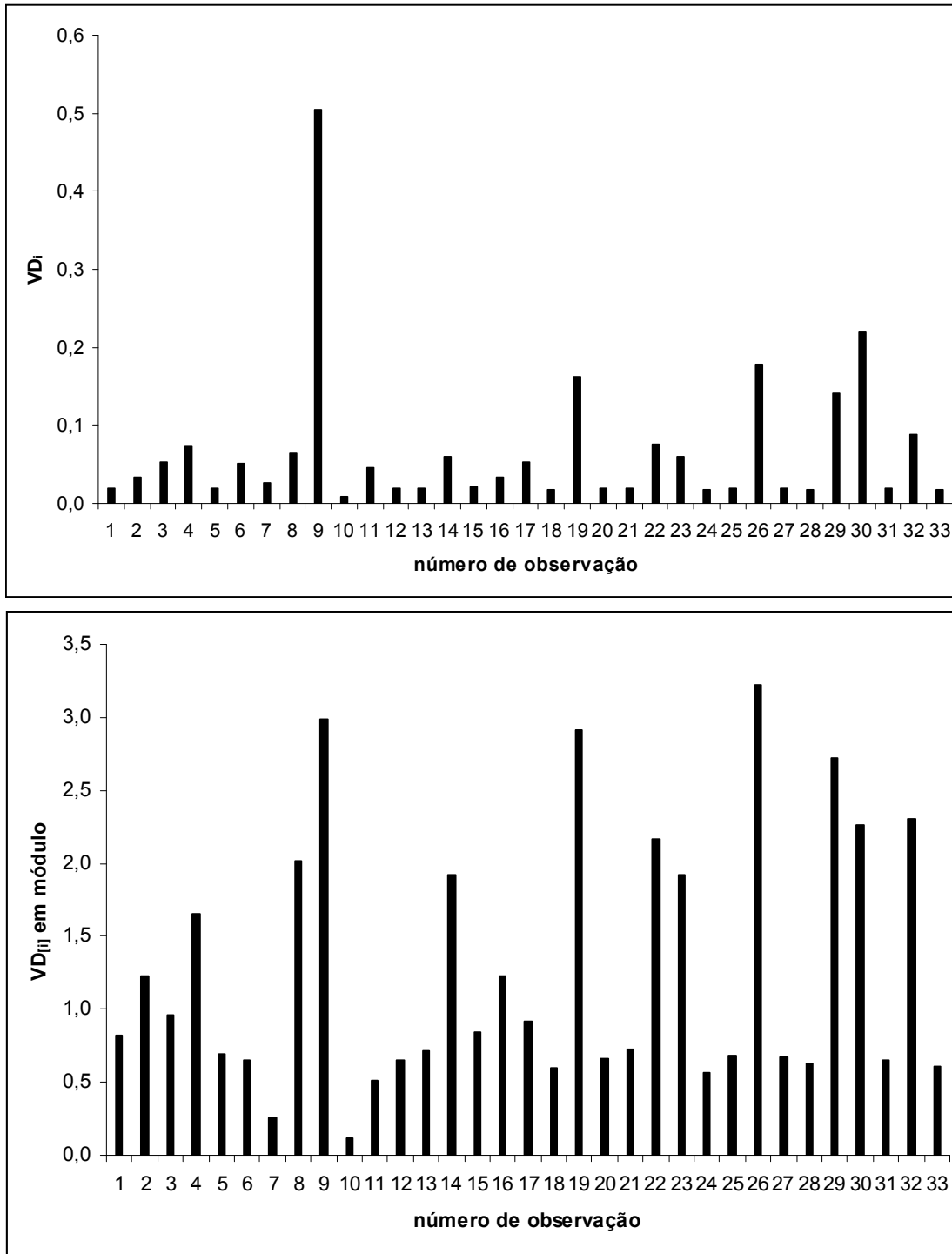
19, 26 e 29) ou famílias que não compraram um automóvel e têm renda familiar alta (famílias 9 e 30).

Tabela 6.2 – Diferenças de verossimilhanças para os dados da Tabela 6.1.

Observação retirada	$\hat{\lambda}_{[i]}$	$\hat{\lambda} - \hat{\lambda}_{(i)}$	$2[L(\hat{\lambda}) - L(\hat{\lambda}_{(i)})]$	$2[L(\hat{\lambda}) - L_{(i)}(\hat{\lambda}_{(i)})]$
nenhuma	2			
1	2	0	0,020093	-0,81857
2	1,5	0,5	0,033290	-1,22788
3	2	0	0,052070	-0,96537
4	1,5	0,5	0,074718	-1,65244
5	2	0	0,018934	-0,69749
6	2	0	0,050348	-0,65042
7	2	0	0,026162	-0,26141
8	2	0	0,065512	-2,01869
9	2	0	0,503970	-2,98286
10	2	0	0,009216	-0,11395
11	2	0	0,045782	-0,51673
12	2	0	0,018598	-0,64713
13	2	0	0,019043	-0,71204
14	2	0	0,060222	-1,92038
15	2	0	0,020397	-0,83994
16	1,5	0,5	0,033290	-1,22788
17	2	0	0,052290	-0,92292
18	2	0	0,018283	-0,59972
19	1	1	0,162789	-2,91422
20	2	0	0,018663	-0,65861
21	2	0	0,019166	-0,72744
22	2	0	0,075315	-2,16363
23	2	0	0,060222	-1,92038
24	2	0	0,018018	-0,56212
25	2	0	0,018838	-0,68376
26	0,5	1,5	0,178257	-3,22293
27	2	0	0,018737	-0,67081
28	2	0	0,018456	-0,62624
29	2	0	0,140307	-2,71912
30	2	0	0,220395	-2,26286
31	2	0	0,018598	-0,64713
32	2	0	0,087764	-2,30409
33	2	0	0,018339	-0,60795

A Figura 6.1 apresenta os gráficos de DV_i e o módulo de $DV_{[i]}$ para as 33 observações.

Figura 6.1 - Gráficos das diferenças de verossimilhança.



Para o cálculo das medidas de influência parcial foram usadas as funções desvio das saídas do programa estatístico R, quando ajustados os modelos (6.1) e (6.2) com base nos dados da Tabela 6.1.

Os cálculos das diferenças nas funções desvio e da medida d_i são dados na Tabela 6.3.

Tabela 6.3 – Medida de influência parcial com base nos dados da Tabela 6.1

Observação retirada	$D_{0[i]} - D_{[i]}$	$(D_0 - D) - (D_{0[i]} - D_{[i]})$
nenhuma		
1	0,13	0,08
2	0,04	1,17
3	0,27	-0,06
4	0,04	0,17
5	0,21	0,00
6	0,18	0,03
7	0,07	0,14
8	0,44	-0,23
9	0,37	-0,16
10	0,05	0,16
11	0,14	0,07
12	0,26	-0,05
13	0,19	0,02
14	0,44	-0,23
15	0,12	0,09
16	0,04	0,17
17	0,26	-0,05
18	0,31	-0,10
19	0,00	0,21
20	0,24	-0,03
21	0,19	0,02
22	0,41	-0,20
23	0,44	-0,23
24	0,37	-0,16
25	0,22	-0,01
26	0,05	0,16
27	0,23	-0,02
28	0,28	-0,07
29	0,12	0,09
30	0,13	0,08
31	0,26	-0,05
32	0,38	-0,17
33	0,30	-0,09

$D_0 - D = 0,21$

Observamos valores de d_i relativamente baixos em módulo, indicando inexistência de pontos influentes no teste que avalia a necessidade de transformação.

Lee e Yick (1999), ainda no contexto de transformações no modelo linear generalizado, apresentam algumas medidas de influência local, que não foram aqui calculadas e seriam sugestões para um trabalho futuro.

6.4 Considerações finais

Devido a facilidades computacionais e à ampla variedade de modelos existentes na literatura, parece ser uma tendência atual preferir utilizar um novo modelo, ao invés de se transformar os dados.

Neste trabalho apresentamos as duas abordagens, dando no entanto maior ênfase ao uso de transformações, em particular, à transformação de Box-Cox no modelo de regressão logística.

Acreditamos que o uso de transformações no modelo de regressão logística, apesar da riqueza de classes de modelos alternativos, ainda é uma boa opção. Destacamos a facilidade de interpretação dos parâmetros desse modelo, sua simplicidade e popularidade e ainda, a disponibilidade na maioria dos pacotes estatísticos.

Referências Bibliográficas

- Aranda-Ordaz, F. J. (1981). On two families of transformations to additivity for binary response data, *Biometrika*, 68, 2, 357-363.
- Berkson, J. (1944). Application of the logistic function to bio-assay, *Journal of the American Statistical Association*, 39, 357-365.
- Bickel, P. J. and Doksum, K. A. (1981). An Analysis of Transformations Revisited, *Journal of the American Statistical Association*, 76, 374, 296-311.
- Box, G. E. P. and Cox, D. R. (1964). An Analysis of Transformations, *Journal of the Royal Statistical Society, Ser. B*, 26, 211-252.
- Carrol, R. J. and Rupert, D. (1981). On Prediction and the Power Transformation Family, *Biometrika*, 68, 3, 609-615.
- Chambers, E. A. and Cox, D. R. (1967). Discrimination between alternative binary response models, *Biometrika*, 54, 573-578.
- Collet, D. (1991). *Modelling Binary Data*, First Edition. London: Chapman and Hall. 369p.
- Copenhaver, T. W. and Mielke, P. W. (1977). Quantit Analysis: A Quantal Assay Refinement, *Biometrics*, 33, 175-186.
- Cramér, H. and Leadbetter, M. R. (1967). *Stationary and Related Stochastic Processes*. New York: John Wiley. 348p.
- Davies, R. B. (1977). Hypothesis Testing when a Nuisance Parameter is Present Only Under the Alternative, *Biometrika*, 64, 2, 247-254.
- Davies, R. B. (1987). Hypothesis Testing when a Nuisance Parameter is Present Only Under the Alternative, *Biometrika*, 74, 1, 33-43.
- Finney, D. J. (1978). *Statistical Method in Biological Assay*. Third edition. London: Charles Griffin and Co Ltd. 508p.
- Guerrero, V. M. and Johnson, R. A. (1982). Use of the Box-Cox transformation with binary response models, *Biometrika*, 69, 2, 309-314.
- Hauck, W. (1990). Choice of Scale and Asymmetric Logistic Models, *Biom. J.*, 32, 1, 79-86.

- Kay, R. and Little, S. (1987). Transformations of the explanatory variables in the logistic regression model for binary data, *Biometrika*, 74, 3, 495-501.
- Lee, A. H. (1988). Assessing partial influence in generalized linear models, *Biometrics*, 44, 71-77.
- Lee, A. H. and Yick, J. S. (1999). Covariate Transformation Diagnostics for Generalized Linear Models, *Ann. Inst. Statist. Math.*, 51, No. 2, 383-398.
- Morgan, B. J. T. (1985). The Cubic Logistic Model for Quantal Assay Data, *Appl. Statist.*, 34, No. 2, 105-113.
- Neter, J., Kutner, M. H., Nachtsheim, C. J., Wasserman, W. (1996). *Applied Linear Statistical Models*. Fourth edition. Chicago: Irwin. 1408p.
- Page, E. (1977). Approximations to the cumulative normal function and its inverse for use on a pocket calculator. *Appl. Statist.*, 26, 75-76.
- Pregibon, D. (1980). Goodness of link tests for generalized linear models, *Applied Statistics*, 29, 15-24.
- Prentice, R. L. (1976). A Generalization of the Probit and Logit Methods for Dose Response Curves, *Biometrics*, 22, 761-768.
- Schneiderman, M. A. (1974). Safe dose? Problem of the statistician in the world of trans-science, *Journal of the Washington Academy of Sciences*, 64, 68-78.
- Silva, G. L. (1992). *Modelos Logísticos para Dados Binários*. Dissertação de Mestrado, IME-USP, 118p.
- Siqueira, A. L. (1992). *Box-Cox Transformation of the Covariate in Logistic Models*. Tese de doutorado, UCLA, 130p.
- Siqueira, A. L. and Taylor, J. M. G., (1999). Treatment Effects in a Logistic Model Involving the Box-Cox Transformation, *Journal of the American Statistical Association*, 94, No. 445, 240-246.
- Stukel, T. A. (1988). Generalized Logistic Models, *Journal of the American Statistical Association*, 83, No. 402, 426-431.
- Taylor, J. M. G. (1988). The Cost of Generalizing Logistic Regression, *Journal of the American Statistical Association*, 83, No. 404, 1078-1083.
- Tukey, J. W. (1957). On the Comparative Anatomy of Transformations, *Annals of Mathematical Statistics*, 28, 602-632.

Wey, B. C. and Hickernell, F. J. (1996). Regression transformation diagnostics for explanatory variables, *Statist. Sinica*, 6, 433-454.