ANDRÉ OLIVEIRA SOUZA

TESTES ESTATÍSTICOS EM REGRESSÃO LOGÍSTICA SOB A CONDIÇÃO DE SEPARABILIDADE

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2010

Ficha catalográfica preparada pela Seção de Catalogação e Classificação da Biblioteca Central da UFV

T
S729t
2010

Souza, André Oliveira, 1978-
Testes estatísticos em regressão logística sob a condição de separabilidade / André Oliveira Souza. – Viçosa, MG, 2010.
xi, 64f. : il. ; 29cm.

Inclui apêndices.
Orientador: Sebastião Martins Filho.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Referências bibliográficas: f. 37-38.

1. Estatística - Teste. 2. Estimativa de parâmetros.
3. Logística. I. Universidade Federal de Viçosa. II. Título.

CDD 22.ed. 519.5

ANDRÉ OLIVEIRA SOUZA

TESTES ESTATÍSTICOS EM REGRESSÃO LOGÍSTICA SOB A CONDIÇÃO DE SEPARABILIDADE

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

APROVADA: 25 de fevereiro de 2010

Prof.ª Rosangela Helena Loschi

Prof. José Ivo Ribeiro Junior

Prof. Enrico Antônio Colosimo
(Co-orientador)

Prof. Fabyano Fonseca e Silva
(Co-orientador)

Prof. Sebastião Martins Filho
(Orientador)

Aos meus pais Expedito Campos de Souza e Ana Balbina de Oliveira Souza, pelos esforços jamais negados, pelos exemplos sempre oferecidos e sem os quais esta conquista não seria possível.

À minha esposa Andrea Fernandes Teixeira pela compreensão de ausência como pai e marido.

A minha filha Ana Beatriz Teixeira Souza, por deixar tudo com mais sentido em minha vida.

Aos irmãos Adelson e Andréia.

Dedico

MENSAGEM

"Só existe uma coisa melhor do que fazer novos amigos, conservar os velhos"

Elmer G. Letterman

AGRADECIMENTOS

Deus, por ter dado saúde, disposição e sempre ter me iluminado dando forças para vencer e chegar até este momento.

À Universidade Federal de Viçosa, por intermédio do Programa de Pós Graduação de Estatística Aplicada e Biometria, pela oportunidade.

À Fundação de Amparo a Pesquisa do Estado de Minas Gerais – FAPEMIG pelo apoio ao desenvolvimento do projeto de pesquisa CAG - PPM-00255-08.

A todos os professores do Departamento de Estatística da Universidade Federal de Viçosa que contribuíram para que eu me tornasse uma pessoa melhor em minha profissão.

Ao secretário Altino pela eficiência sempre demonstrada e, sobretudo, pelo bom humor inesgotável em todos os momentos.

A secretária do DET-UFV Anita, pela presteza e simpatia.

Aos professores Enrico Antonio Colosimo e Sebastião Martins Filho pelo apoio durante o desenvolvimento deste trabalho, e pelos bons ensinamentos durante este tempo que trabalhamos juntos.

Aos colegas do mestrado UFV, em especial a todos do semestre 2008/01.

Ao professor Fabyano, pelo apoio, sabedoria enquanto professor e generosidade como ser humano.

A todos que de alguma forma contribuíram para a realização deste trabalho.

BIOGRAFIA

ANDRÉ OLIVEIRA SOUZA, filho de Ana Balbina Oliveira Souza e Expedito Campos de Souza, nasceu em 30 de setembro de 1978, em Senador Firmino – MG .

Em janeiro de 2000 graduou-se em licenciatura plena em Matemática pela Universidade Presidente Antonio Carlos (UNIPAC).

Em 2002 concluiu o curso de especialização em Matemática, pela Universidade Presidente Antonio Carlos (UNIPAC) com a monografia intitulada: Dificuldades no ensino-aprendizagem da matemática e propostas de solução.

Em março de 2008, iniciou o curso de Mestrado em Estatística Aplicada e Biometria, na Universidade Federal de Viçosa (UFV) tendo defendido a dissertação em 25 de fevereiro de 2010.

ÍNDICE

LISTA DE FIGURAS

LISTA DE TABELAS

RESUMO

SOUZA, André Oliveira, M.Sc., Universidade Federal de Viçosa, fevereiro de 2010. **Testes estatísticos em regressão logística sob a condição de separabilidade.** Orientador: Sebastião Martins Filho. Co-Orientadores: Enrico Antonio Colosimo e Fabyano Fonseca e Silva.

A regressão logística é o método estatístico usual de análise utilizado quando o objetivo é verificar a relação entre uma variável resposta dicotômica e variáveis explicativas de interesse. Usualmente, os parâmetros deste modelo são estimados pelo método de máxima verossimilhança genuína, e testes sobre estes parâmetros são construídos considerando as distribuições aproximadas dos estimadores. Isto significa que amostras grandes tornam-se necessárias para obter resultados mais confiáveis. Em estudos envolvendo dados binários, é frequente a presença de uma variável resposta cujo sucesso é pouco provável, ou seja, tem-se um evento raro, o que pode gerar uma amostra de dados esparsos. Nestes casos, diz-se que os dados podem estar sob a condição de separabilidade, e esta situação está frequentemente associada à presença de uma covariável categórica, podendo os estimadores de máxima verossimilhança, para pelo menos um parâmetro, não existir. Na situação de separabilidade recomenda-se utilizar o método de máxima verossimilhança penalizada proposto por Firth (1993). O objetivo principal deste trabalho foi verificar por meio de simulação Monte Carlo os poderes dos testes da razão de verossimilhanças (TRV) e de Wald obtido via máxima verossimilhança penalizada na condição de separabilidade. A metodologia apresentada neste trabalho foi aplicada a dois conjuntos de dados reais. A simulação Monte Carlo com uma variável explicativa no modelo possibilitou obter indicativos que o TRV tem maior poder que o teste de Wald.

ABSTRACT

SOUZA, André Oliveira, M.Sc., Universidade Federal de Viçosa, February, 2010. **Statistical Tests in logistic regression under separability condition.** Adviser: Sebastião Martins Filho. Co-Advisers: Enrico Antonio Colosimo and Fabyano Fonseca e Silva.

Logistic regression is the statistical method of analysis used when the objective is to verify the relationship between one dichotomic response variable and explicative variables of interest. Usually, the model parameters are estimated through the genuine maximum likelihood method, and tests about these parameters are built assuming approximated distributions for the estimators. This means that large samples become necessary to obtain trustable results. In studies involving binary data is common the occurrence of one response variable whose success has low probability, in other words, a rare event that can generate a sparse data sample. In such cases, the data are under separability condition, and this situation is frequently associated to the presence of one categorical co-variable, what means that the maximum likelihood estimators do not exist to one parameter at least. In the separability condition it is recommended to use the Penalized Maximum Likelihood method, proposed by Firth (1993). The main objective of this study was to verify the powers of the Likelihood Ratio Test (LRT) and Wald Test obtained through PML under separability condition by Monte Carlo simulation. The presented methodology has been applied to two real data sets. Monte Carlo simulation with one explicative variable in the model made possible to obtain indicatives that the LRT is most powerful than the Wald test.

INTRODUÇÃO

Em muitos estudos nas diversas áreas da ciência, a variável dependente ou variável resposta, apresenta apenas duas categorias, como exemplo o resultado de experimentos com germinação de sementes, nos quais tem-se como resposta o sim, se germinou, ou não, caso contrário. Tais respostas dicotômicas podem ser codificadas numericamente como 1 e 0, respectivamente correspondendo assim a um conjunto de dados binários.

Quando se tem o interesse na avaliação da influência de fatores sobre uma resposta dicotômica, a regressão logística é o método usualmente utilizado (Hosmer e Lemeshow, 1989). Geralmente os testes de hipóteses para os parâmetros do modelo logístico são fundamentados nas estatísticas de Wald e da razão de verossimilhanças, cujos poderes podem diferir em situações envolvendo diferentes configurações de dados amostrais.

Uma situação na qual uma comparação se faz necessária, devido a escassez de trabalhos na literatura especializada, é a da separabilidade esta ocorre quando, as respostas sim e não podem ser perfeitamente separadas por um fator ou por combinações lineares não-triviais de vários fatores. A probabilidade de ocorrência destas situações depende do tamanho da amostra e do número de fatores dicotômicos de interesse (Heinze e Schemper, 2002).

Inferências para os coeficientes do modelo logístico não podem ser fundamentadas na estatística de Wald, quando o método de estimação é o de máxima verossimilhança genuína, pois neste caso, tanto os estimadores quanto o erro padrão de pelo menos um dos coeficientes poderá ir para o infinito. Este fato implica em intervalo de confiança (IC) com amplitude infinita (Heinze e Schemper, 2002), tornando o teste de Wald não conclusivo. Portando, nesta situação apenas o teste da razão de verossimilhanças poderá ser utilizado.

Por outro lado, sob a configuração de separabilidade quando o método de estimação utilizado for o de máxima verossimilhança penalizada (Firth, 1993), inferências para os coeficientes do modelo podem ser fundamentadas na estatística de Wald.

Diante do assunto exposto o objetivo deste trabalho é investigar o poder do TRV, sob separabilidade, quando se utiliza o método de estimação de máxima verossimilhança genuína, e também, o poder do teste de Wald quando se trabalha com o método de máxima verossimilhança penalizada proposta por Firth (1993).

Os resultados desta investigação serão utilizados em dois conjuntos de dados reais sob configuração de separabilidade. Em um a resposta de interesse foi a ocorrência de

meningite durante os primeiros 30 dias após o paciente ser submetido a craniotomia (Colosimo *et al*., 1995), no outro foi avaliada a germinação de sementes de *Adenanthera pavonina* L.

Este trabalho está organizado da seguinte forma: No Capítulo 1 estão apresentados a motivação deste estudo, o modelo de regressão logística e inferência do modelo. No Capítulo 2 esta discutida a existência dos estimadores de máxima verossimilhança, a classificação dos dados logísticos, a estimação e a inferência obtida pelos métodos de máxima verossimilhança genuína e também o método de máxima verossimilhança penalizada. No Capítulo 3 encontra-se descrito todo o processo e estrutura da simulação de dados. No Capítulo 4 encontram-se aplicações dos testes fundamentados em resultados obtidos pela investigação realizada neste trabalho. E ao final são apresentadas as conclusões desta dissertação.

Capítulo 1 – Regressão logística

Neste capítulo é apresentada uma motivação do trabalho e uma breve revisão do modelo de regressão logística e inferências para a mesma.

1.1 – Motivação

Ao propor o modelo logístico para modelar dados provenientes de experimentos com geminação de sementes, no qual alguns efeitos dos fatores ou efeitos de combinações de fatores a geminação é nula, estimativas obtidas por máxima verossimilhança genuína para estimar o efeito de tais tratamentos são imprecisas e divergem para $\pm\infty$. Para este caso uma alternativa, proposta por Firth (1995), é a modificação do método de estimação no qual garante estimativas finitas e precisas para os coeficientes do modelo. A condição de separabilidade foi apresentada por Albert e Anderson (1984) em que os mesmos estabeleceram a fundamentação teórica para a análise deste fenômeno e, também Heinze e Schemper (2002) sugeriram algumas abordagens, para a classificação de dados logístico já discutidas por Albert e Anderson (1984). Neste trabalho serão modelados dois conjuntos de dados reais. O primeiro é conhecido da literatura, em que pacientes foram submetidos a craniotomia (Colosimo *et al.,* 1995). O segundo conjunto de dados é oriundo de um experimento com germinação de sementes de *Adenanthera pavonina* L realizado no laboratório de sementes florestais da Universidade Federal de Viçosa em 2009.

1.2 – Modelo de Regressão Logística

Um dos casos particulares dos modelos lineares generalizados (Dobson, 1990; Paula, 2004) são os modelos para variáveis que apresentam apenas duas categorias ou que de alguma forma foram dicotomizadas assumindo os valores 0 ou 1. São as chamadas variáveis *dummy* (ou indicadoras). Um dos mais importantes modelos é o de regressão logística, baseado na transformação *logit* para proporção.

Variáveis com duas categorias que podem ser classificadas em sucesso ou fracasso representando as possibilidades de respostas como, por exemplo, (1; 0), são caracterizadas pela distribuição de Bernoulli. Comumente é chamado de sucesso o resultado mais importante da resposta ou aquele que se pretende relacionar com outras variáveis de

interesse. A distribuição de Bernoulli para a variável aleatória binária Y com parâmetro π especifica as probabilidades como:

$$P(Y=1)=\pi \quad \text{e} \quad P(Y=0)=1-\pi$$

Por definição,

$$E(Y)=1\pi+0(1-\pi)=\pi$$

que é a proporção de respostas em que $Y=1$ e sendo,

$$Var(Y)=E(Y^2)-\left[E(Y)\right]^2=1^2\,\pi+0^2(1-\pi)-\pi^2\,.$$

$$=\pi(1-\pi)$$

A função de probabilidade de uma variável aleatória Bernoulli é,

$$f(Y,\pi)=\pi^y(1-\pi)^y$$

A regressão logística é conhecida desde os anos 50, entretanto, tornou se mais usual através de Cox (1970) e de Hosmer e Lemeshow (1989). Aspectos teóricos do modelo de regressão logística são amplamente discutidos na literatura, destacando-se Cox e Snell (1989), Hosmer e Lemeshow (1989), Agresti (1990), Kleinbaum (1994) entre outros.

1.2.1 – Regressão logística simples

Os métodos de regressão têm como objetivo descrever as relações entre a variável resposta (Y) e a variável explicativa (X). Na regressão logística, a variável resposta (Y) é dicotômica, isto é, atribui-se o valor 1 (um) para o evento de interesse sucesso e o valor (0) zero para o acontecimento complementar fracasso. Com probabilidade de sucesso

$\pi_i(x_i)=P[Y_i=1\mid X_i]$

em que X_i é a variável explicativa associada a i-ésima resposta Y_i.

Considera-se uma amostra de respostas binárias, em que $(Y_1,Y_2,Y_3,...,Y_n)$ são variáveis aleatórias independentes com distribuição Bernoulli, com probabilidade de sucesso π_i, isto é, $Y_i \sim Benoulli(\pi_i)$ e denota-se por $x_i^T=(1,x_i)$ a i-ésima linha da matriz **X** em que i=1,2,3,...n.

A probabilidade de sucesso do modelo logístico simples é definida como:

$$\pi_i=\pi_i(x_i)=P(Y_i=1\mid X_i=x_i)=\frac{\exp(\beta_0+\beta_1x_i)}{1+\exp(\beta_0+\beta_1x_i)}, \qquad (1.1)$$

e a probabilidade de fracasso,

$$1\text{-}\pi_i = P(Y_i = 0 \mid X_i = x_i) = \frac{1}{1+\exp(\beta_0 + \beta_1 x_i)}, \qquad (1.2)$$

em que β_0 *e* β_1 são os parâmetros desconhecidos. Em problemas de regressão o que se modela é o valor médio da variável resposta dado os valores das variáveis independentes. Esta quantidade é chamada de média condicional, denotada por $E(Y_i \mid X = x_i)$, em que Y_i é a variável resposta e x_i, os valores das variáveis independentes. Devido a natureza da variável resposta, a amplitude da média condicional varia no intervalo [0,1], ou seja, $0 \le E(Y_i \mid X = x_i) \le 1$ e usando a definição de variáveis aleatórias discreta, tem-se:

$$E(Y_i \mid X_i = x_i) = 1P(Y_i = 1 \mid X_i = x_i) + 0\,P(Y_i = 0 \mid X_i = x_i) = P[Y_i = 1 \mid X_i = x_i].$$

A variável resposta Y_i dado x_i é modelada por $Y_i = \pi_i + \varepsilon_i$. Como a quantidade ε_i pode assumir somente um de dois valores possíveis, isto é, $\varepsilon_i = 1 - \pi_i$ para $Y_i = 1$ ou $\varepsilon_i = -\pi_i$ para $Y_i = 0$, segue que ε_i tem distribuição com média zero e variância dada por $\pi_i(1-\pi_i)$ (Hosmer e Lemeshow, 1989), isto é, a distribuição condicional da variável resposta segue uma distribuição binomial com probabilidade dada pela média condicional π_i.

1.2.1.1 – Transformação logit

Para evitar o problema restritivo de que os valores de probabilidade sejam números no intervalo [0,1], a função logística pode ser linearizada pela transformação chamada *logit*.

A transformação *logit* que é central para estudo de regressão logística é definida como $g(x_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$, logo de (1.1) e (1.2) tem-se,

$$g(x_i) = \ln\left(\frac{\frac{\exp(\beta_0 + \beta_1 x_i)}{1+\exp(\beta_0 + \beta_1 x_i)}}{1 - \frac{\exp(\beta_0 + \beta_1 x_i)}{1+\exp(\beta_0 + \beta_1 x_i)}}\right) = l\,\mathrm{n}\left(\frac{\frac{\exp(\beta_0 + \beta_1 x_i)}{1+\exp(\beta_0 + \beta_1 x_i)}}{\frac{1}{1+\exp(\beta_0 + \beta_1 x_i)}}\right),$$

$$g(x_i) = \ln(\exp(\beta_0 + \beta_1 x_i)) = \beta_0 + \beta_1 x_i. \qquad (1.3)$$

A função $g(x_i)$ apresenta as seguintes propriedades: é linear em seus parâmetros, contínua, varia no intervalo $(-\infty,+\infty)$ com correspondentes no intervalo [0,1] para π_i. No contexto de modelos lineares generalizados, uma função monótona e derivável que relaciona a média ao preditor linear é denominada de função de ligação, assim $g(x_i)=\ln\left(\frac{\pi_i}{1-\pi_i}\right)$ é a função de ligação canônica para o modelo binomial.

1.2.1.2 – Estimação dos parâmetros

Supondo que (x_i, y_i) seja uma amostra independente com n pares de observações, y_i representa o valor da variável dicotômica e x_i o valor da variável independente da i-ésima observação em que i=1, 2,3,...,n. Para o ajuste do modelo de regressão logística simples, segundo a equação (1.1), é necessário estimar os parâmetros desconhecidos; β_0 e β_1. O método mais usado para estimar esses parâmetros considerando uma regressão linear clássica é o de mínimos quadrados. Neste método, a escolha de β_0 e β_1 é dada pelos valores que minimizam a soma de quadrados dos desvios para os valores observados (y_i) em relação ao valor predito ($\hat{y}_i$) baseado no modelo, neste caso, a matriz de projeção H da solução de mínimos quadrados é:

$$H = X(X^T X)^{-1} X^T,$$

em que X a matriz de dados, no entanto, no modelo de regressão logística, a variância $Var(\varepsilon_i)=\pi_i(1-\pi_i)$ não é constante, sendo utilizada a definição de mínimos quadrados ponderados, definindo a matriz de projeção para o modelo logístico como:

$$H = Q^{1/2} X(X^T Q X)^{-1} X^T Q^{1/2},$$

em que, $Q = diag\left[\pi_i(1-\pi_i)\right]$, i=1,...,n.

Usualmente o método de máxima verossimilhança é utilizado para estimar os parâmetros no caso de modelo de regressão logística. Como as observações são independentes, a função de distribuição de probabilidade conjunta de $y_1, y_2, ..., y_n$ será:

$$\prod_{i=1}^{n} f(y_i,\pi_i) = \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i},\ y_i \in [0,1]. \quad (1.4)$$

Então a função de verossimilhança é dada por:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \boldsymbol{\pi}_i^{\,y_i} (1-\boldsymbol{\pi}_i)^{1-y_i}, \boldsymbol{\beta} \in \mathbb{R}^2 . \tag{1.5}$$

O método de máxima verossimilhança consiste em estimar $\boldsymbol{\beta}$ considerando o valor deste parâmetro que maximiza $L(\boldsymbol{\beta})$. Aplicando o logaritmo em $L(\boldsymbol{\beta})$, a expressão é definida como:

$$\begin{aligned} l(\beta) &= \ln\left[L(\beta)\right] = \ln\left[\prod_{i=1}^{n} \pi_i^{\,y_i} (1-\pi_i)^{1-y_i}\right] \\ &= \sum_{i=1}^{n} [\mathrm{y_i} \ln(\pi_\mathrm{i}) + (1\text{-}\mathrm{y_i}) \ln(1\text{-}\pi_\mathrm{i})] \\ &= \sum_{i=1}^{n} [\mathrm{y_i} \ln(\pi_\mathrm{i}) + \ln(1\text{-}\pi_\mathrm{i}) \text{-} \mathrm{y_i} \ln(1\text{-}\pi_\mathrm{i})] \\ &= \sum_{i=1}^{n} [\mathrm{y_i} \ln\left(\frac{\pi_\mathrm{i}}{1\text{-}\pi_i}\right) + \ln(1\text{-}\pi_\mathrm{i})]. \end{aligned} \tag{1.6}$$

Substituindo em (1.6) as equações (1.2) e (1.3), tem-se:

$$\begin{aligned} l(\beta) &= \sum_{i=1}^{n} \left[\mathrm{y_i}(\beta_0 + \beta_1 \mathrm{x_i}) + \ln\left(\frac{1}{1+\exp(\beta_0 + \beta_1 x_i)}\right)\right] \\ &= \sum_{i=1}^{n} \left[\mathrm{y_i}(\beta_0 + \beta_1 \mathrm{x_i}) + \ln(1) \text{-} \ln(1+\exp(\beta_0 + \beta_1 x_i)\right] \\ &= \sum_{i=1}^{n} \left[\mathrm{y_i}(\beta_0 + \beta_1 \mathrm{x_i}) + \ln(1) \text{-} \ln(1+\exp(\beta_0 + \beta_1 x_i)\right] \\ &= \sum_{i=1}^{n} \left[\mathrm{y_i}(\beta_0 + \beta_1 \mathrm{x_i}) \text{-} \ln(1+\exp(\beta_0 + \beta_1 x_i)\right]. \end{aligned} \tag{1.7}$$

Para encontrar o valor de β que maximiza $l(\beta)$, deriva se $l(\beta)$ em relação a cada parâmetro (β_0, β_1), obtendo-se duas equações.

$$\frac{\partial l(\beta)}{\partial \beta_0} = \sum_{i=1}^{n}\left[\mathrm{y_i} - \frac{1}{1+\exp(\beta_0 + \beta_1 x_i)} \exp(\beta_0 + \beta_1 x_i) \right]$$

$$\frac{\partial l(\beta)}{\partial \beta_1} = \sum_{i=1}^{n}\left[\mathrm{y_i} x_i - \frac{1}{1+\exp(\beta_0 + \beta_1 x_i)} \exp(\beta_0 + \beta_1 x_i) x_i \right],$$

que, igualando a zero geram o sistema de equações:

$$\sum_{i=1}^{n}(y_i - \pi_i) = 0 \qquad (1.8)$$

$$\sum_{i=1}^{n}(y_i - \pi_i)x_i = 0 \qquad (1.9)$$

em que i=1,2,3,....,n e $\pi_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1+\exp(\beta_0 + \beta_1 x_i)}$.

Como as equações (1.8) e (1.9) são não lineares em β_0 *e* β_1, são necessários métodos iterativos para resolução, e estes estão implementados em vários softwares estatísticos.

1.2.2 – Regressão Logística múltipla

Hosmer e Lemeshow (1989) generalizaram o modelo de regressão para o caso de uma ou mais variáveis independentes.

Seja um conjunto de p variáveis independentes, denotado por $x_i^T = (x_{i0}, x_{i2}, x_{i3}, ..., x_{pi})$, o vetor da i-ésima linha da matriz (X) das variáveis explicativas, em que cada elemento da matriz corresponde ao ij-ésimo componente (x_{ij}), em que i=1, 2, 3,..., n e j =1, 2, 3,..., p, com $x_{i0} = 1$. Denota-se por $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, ..., \beta_p)^T$, o vetor de parâmetros desconhecidos e β_j é o j-ésimo parâmetro associado à variável explicativa x_j. No modelo de regressão logística múltipla a probabilidade de sucesso é dada por:

$$\pi_i = \pi(x_i) = P(Y_i = 1 \mid X_i = x_i) = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_{i2} + ... + \beta_p x_{ip})}{1+\exp(\beta_0 + \beta_1 x_{i1} + \beta_{i2} + ... + \beta_p x_{ip})}, \qquad (1.10)$$

$$\pi(x_i) = P(Y_i = 1 \mid X_i = x_i) = \frac{\exp(x_i^T \beta)}{1+\exp(x_i^T \beta)}$$

E a probabilidade de fracasso é dada por,

$$1-\pi_i = 1-\pi(x_i) = P(Y_i = 0 \mid X_i = x_i) = \frac{1}{1+\exp(\beta_0 + \beta_1 x_{i1} + \beta_{i2} + ... + \beta_p x_{ip})}, \quad (1.11)$$

$$1-\pi(x_i) = P(Y_i = 0 \mid X_i = x_i) = \frac{1}{1+\exp(x_i^T \beta)}$$

O "*logit*" para o modelo de regressão linear múltipla é dado pela equação:

$$g(x_i) = \ln\left[\frac{\pi_i}{1-\pi_i}\right] = x_i^T \beta = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} \quad (1.12)$$

Assim o logaritmo da função verossimilhança pode ser escrito:

$$l(\beta) = \sum_{i=1}^{n} \left[y_i x_i^T \beta - \ln(1+\exp(x_i^T \beta)) \right]. \quad (1.13)$$

1.2.2.1 – Estimação dos parâmetros

Para estimar os parâmetros da regressão logística múltipla por máxima verossimilhança encontra-se o valor de β que maximiza $l(\beta)$, o que exige um processo iterativo e que faz necessário derivar $l(\beta)$ em relação a cada parâmetro;

$$\frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} \left[y_i x_{ij} - \frac{\exp(x_i^T \beta)}{1+\exp(x_i^T \beta)} x_{ij} \right]$$

$$= \sum_{i=1}^{n} \left[y_i - \pi_i \right] x_{ij}, \quad (1.14)$$

dessa forma, o vetor score;

$$U(\beta) = X^T y - X^T \pi^T = X^T (y - \pi^T), \quad (1.15)$$

em que $\pi^T = (\pi_1, ..., \pi_n)$.

A matriz de informação de Fischer é dada por:

$$I(\beta) = E\left(-\frac{\partial^2 l(\beta)}{\partial \beta \, \partial \beta^T} \right) = X^T Q X \quad (1.16)$$

em que, $Q = diag\left[\pi_i(1-\pi_i)\right]$, i=1,...,n e X a matriz de dados, e sua inversa $\left[I(\beta)^{-1} \right]$, é a matriz de variância e covariância das estimativas de máxima verossimilhança dos parâmetros.

A solução para as equações (1.14) é obtida por método iterativo de Newton Raphson. O conjunto de equações iterativas é dado por:

$$\beta^{(t+1)} = \beta^{(t)} + \left[I\left(\beta^{(t)}\right) \right]^{-1} U\left(\beta^{(t)}\right); \; t = 1, 2, 3.... \qquad (1.17)$$
$$= \beta^{(t)} + \left[X^T Q^{(t)} X \right]^{-1} X^T \left(y - \pi^{(t)} \right)$$

em que β^t e β^{t+1} são vetores de parâmetros estimados nos passos t e $t+1$, respectivamente.

Para o valor inicial, é usualmente tomado, os coeficientes iguais a zero. Estes valores iniciais são distribuídos no primeiro membro da equação (1.17), que dará o resultado para a primeira iteração, $\beta^{(1)}$. Os valores então são novamente distribuídos no primeiro membro da equação (1.17), $U(\beta)$ e $I(\beta)$ são recalculados, encontrando $\beta^{(2)}$. Esse processo é repetido, até que a máxima mudança em cada parâmetro estimado do próximo passo seja menor que um critério. Se o valor absoluto do corrente parâmetro estimado $\beta^{(t)}$ é menor ou igual a 0,01, o critério mais usual para convergência é $\left|\beta^{(t+1)} - \beta^t\right| < 0,0001$. Se o parâmetro estimado for maior que 0,01, assume se o seguinte critério $\dfrac{\left|\beta^{(t+1)} - \beta^t\right|}{\beta^t} < 0,001$, conforme (Allison 1999).

1.3 – Estatística *deviance*

O processo de ajuste de um modelo consiste em propor ao mesmo um pequeno número de parâmetros, de tal forma que resuma toda informação da amostra.

Dado um conjunto de n observações, um modelo de até n parâmetros pode ser ajustado, sendo denominado modelo saturado, sendo que este indica toda variação ao componente sistemático e reproduz exatamente os dados. Por outro lado, o modelo mais simples tem somente um parâmetro, β_0, sendo denominado modelo nulo, e indicando toda variação ao componente aleatório. Na prática, o modelo nulo é em geral muito simples e o modelo saturado não é informativo, uma vez que não resume os dados, somente os reproduz. Entretanto, o modelo saturado serve como base para medir a discrepância de um modelo intermediário com p parâmetros em que $p < n$.

Existem muitas estatísticas para medir esta discrepância, das quais a mais utilizada está baseada na função de verossimilhança, proposta por Nelder e Wedderburn (1972),

com o nome *deviance.* Os autores comparam o valor da função de verossimilhança, para o modelo proposto com $p+1$ parâmetros $\left(L(\hat{\beta}_0, \hat{\beta}_1, ...\hat{\beta}_p)\right)$ ao seu valor no modelo saturado $\left(\left(L(y_1, y_2, ...y_n)\right)\right)$. Para esta comparação é conveniente utilizar menos duas vezes o logaritmo do quociente destes máximos. Assim, a *deviance* é definida como:

$$G = -2\ln\left[\frac{L(\hat{\beta}_0, \hat{\beta}_1, ...\hat{\beta}_p)}{L(y_1, y_2, ...y_n)}\right], \qquad (1.18)$$

equação na qual verifica-se a utilização de um teste da razão de verossimilhanças generalizado.

No modelo de regressão logística, considerando o modelo com as proporções estimadas $\hat{\pi}_i$, temos que a *deviance* pode ser escrita como:

$$\begin{aligned} G &= -2\sum_{i=1}^{n}\left[y_i - \ln(\hat{\pi}_i) + (1-y_i)\ln(1-\hat{\pi}_i) - y_i\ln(y_i) + (1-y_i)\ln(1-y_i)\right] \\ &= -2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{\hat{\pi}_i}{y_i}\right) + (1-y_i)\ln\left(\frac{1-\hat{\pi}_i}{1-y_i}\right)\right] \\ &= 2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{y_i}{\hat{\pi}_i}\right) + (1-y_i)\ln\left(\frac{1-y_i}{1-\hat{\pi}_i}\right)\right] \qquad (1.19) \end{aligned}$$

A *deviance* é sempre positiva e quanto menor seu valor, melhor é o ajuste do modelo.

1.4 – Testes Estatísticos

Geralmente não é possível encontrar distribuições exatas para os estimadores, assim sendo trabalha-se com resultados assintóticos considerando-se que o modelo escolhido irá satisfazer as condições de regularidades.

Cox e Hinkley (1986) demonstram que, em problemas regulares, a função escore $U(\beta) = \frac{\partial l(\beta)}{\partial \beta}$ tem valor esperado igual a zero e a estrutura de covariância é igual à matriz de informação de Fischer $I(\beta) = E\left(-\frac{\partial^2 l(\beta)}{\partial \beta\, \partial \beta^T}\right)$. Assim a distribuição assintótica dos $\hat{\beta}$ é dada por:

$$\widehat{\beta} \sim N_p(\beta, I(\beta)^{-1}) \qquad (1.20)$$

Os métodos de inferência são baseados na teoria de máxima verossimilhança. Conforme esta teoria existem três estatísticas para testar hipóteses relacionadas aos parâmetros (Razão de verossimilhança, de Wald e Escore), que são deduzidas de distribuições assintóticas de funções adequadas dos parâmetros (Demétrios, 2002).

As duas primeiras estatísticas estão definidas abaixo:

1. Estatística da razão de verossimilhanças:

O teste da razão de verossimilhanças é obtido por meio da comparação entre o modelo sob, Ho: $\beta = \beta_0$, e o irrestrito. A estatística deste teste, sob Ho, tem aproximadamente uma distribuição de qui-quadrado com número de graus de liberdade igual à diferença do número de parâmetros dos modelos que estão sendo comparados.

$$G = -2\ln\left[\frac{L(\hat{\beta}_0)}{L(\hat{\beta})}\right], \qquad (1.21)$$

2. A estatística de Wald:

O teste de Wald é baseado na distribuição assintótica de $\hat{\beta}$ e é uma generalização do teste t de Student (Wald, 1943)**.** Sob a hipótese Ho: $\beta = \beta_0$ a estatística do teste é dada por:

$$W = \left(\widehat{\beta} - \beta_0\right)^T I(\widehat{\beta})\left(\widehat{\beta} - \beta_0\right) \qquad (1.22)$$

em que $I(\widehat{\beta})$ é a matriz de informação de Fischer avaliada em $\widehat{\beta}$, em que sob Ho, W tem aproximadamente uma distribuição de qui-quadrado com graus de liberdade igual ao número parâmetros testados.

Capítulo 2 – Existência de estimadores de máxima verossimilhança modelos de regressão logística

Neste capítulo é apresentado os conceitos algébricos e empíricos de separação completa, separação quase-completa e superposição (*overlap*) utilizada para classificar dados logísticos. O método de estimação de máxima verossimilhança penalizada, proposto por Firth (1993), é apresentado para os parâmetros do modelo logísticos. Também são discutidos os testes de Wald e o da razão de verossimilhanças para inferência dos parâmetros em cada um dos métodos de estimação.

2.1 – Classificações de um conjunto de dados logísticos

Segundo Albert e Anderson (1984) as configurações dos dados logísticos podem ser classificadas em três categorias mutuamente exclusivas e exaustivas: separação completa; separação quase completa e superposição (casos comuns, *overlap*). A separabilidade ocorre quando as respostas sim e não podem ser perfeitamente separadas por uma covariável de interesse ou por combinações lineares não-triviais de covariáveis.

A seguir apresenta-se formalmente esta classificação. Para isto, serão considerados as configurações possíveis dos n valores amostrais no espaço de observação $\mathbb{R}^p$ e a partir destes valores defini-se cada uma das categorias citadas.

2.1.1 – Separação Completa

Ocorre separação completa quando, baseada na informação de uma covariável ou combinação de covariáveis, pode-se predizer corretamente o valor de uma variável de interesse. Isto implica na existência de um vetor $\beta \in \mathbb{R}^{p+1}$ pelo qual todos os valores amostrais podem ser perfeitamente classificados entre Y=1 ou Y=0, tal que todo i$\in E_j$, j=0,1, tem-se

$$X_i^{'}\beta > 0,\ i \in E_0,$$
$$X_i^{'}\beta < 0,\ i \in E_1,$$

em que E_j é o conjunto de linhas identificadas da matriz X com valores de Y= j. A Figura 2.1 (a) ilustra esta categoria de separação para $\mathbb{R}^2$.

2.1.2 – Separação Quase Completa

Ocorre separação quase-completa quando, baseado na informação de uma covariável ou combinação de covariáveis, pode-se predizer perfeitamente os valores de pelo menos um grupo da variável de interesse, ou seja, Y=0 ou Y=1. A separação quase-completa implica na existência de vetor $\beta \in \mathbb{R}^{p+1}$ tal que, para todo i$\in E_j$, j=0,1

$$X_i^{'}\beta \geq 0,\ i \in E_0,$$
$$X_i^{'}\beta \leq 0,\ i \in E_1,$$

com igualdade para, pelo menos, um, valor de i. A Figura 2.1 (b) ilustra esta categoria de separação para $\mathbb{R}^2$.

2.1.3 – Superposição (*overla*p)

Se os dados não estão nas duas categorias anteriores, necessariamente, eles estão na categoria de superposição. A Figura 2.1(c) ilustra esta categoria para $\mathbb{R}^2$.

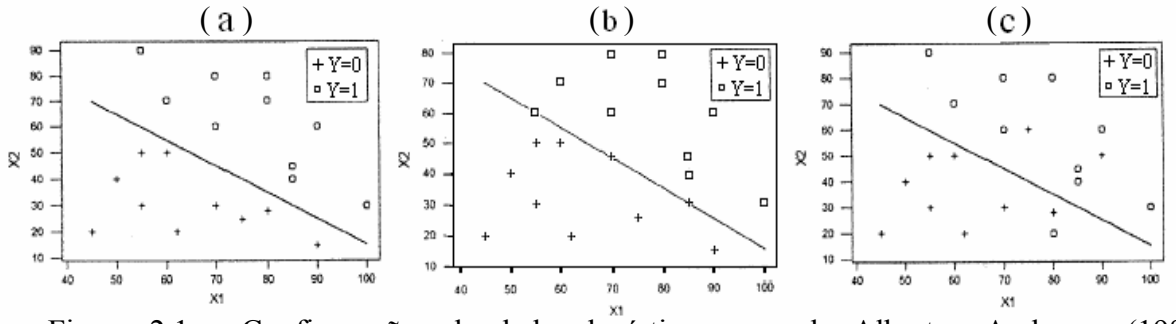


Figura 2.1 – Configurações de dados logísticos segundo Albert e Anderson (1984), separação completa (a), quase-completa (b) e "*overlap*" (c).

Segundo Albert e Anderson (1984) a detecção da separação entre grupos pode ser abordada de duas maneiras distintas, algébrica ou empírica. Em expansão do trabalho de Albert e Anderson (1984), Santner e Duffy (1986) apresentaram um modelo de Programação Linear que classifica os dados como (i) completamente separados, (ii) quase separados (iii) sobrepostos. Clarkson e Jenrick (1991) também apresentaram procedimentos computacionais sofisticados para detectar a separação dos dados, no entanto, na prática há duas alternativas simples para identificar a separação: Caso a covariável seja continua, monitorar a variância dos coeficientes estimados da regressão

(Heinze e Schemper, 2002), se observar variâncias grandes para algum parâmetro estimado, há um indicativo de separabilidade. Outra alternativa, caso a covariável seja categórica, é fazer uma tabela de contingência, cruzando a variável resposta com cada uma das covariáveis categóricas e verificar se existem caselas com valores observados iguais a zero (Nacle, 2004). O valor zero em apenas uma, e somente uma, casela indica separação quase-completa, dois zeros em caselas discordantes indicam separação completa.

2.2 – Estimadores de máxima verossimilhança

A estimação dos parâmetros no caso do modelo de regressão logística, geralmente é realizada utilizando o método de máxima verossimilhança. No entanto, Albert e Anderson (1984) provaram que quando um conjunto de dados está nas categorias de separação completa ou quase-completa, a função de verossimilhança genuína (Figura 2.2 b) do modelo logístico é monótona e, portanto, por este método obtêm-se estimativas infinitas. Desta forma, torna-se importante encontrar um procedimento eficiente para a estimação destes parâmetros.

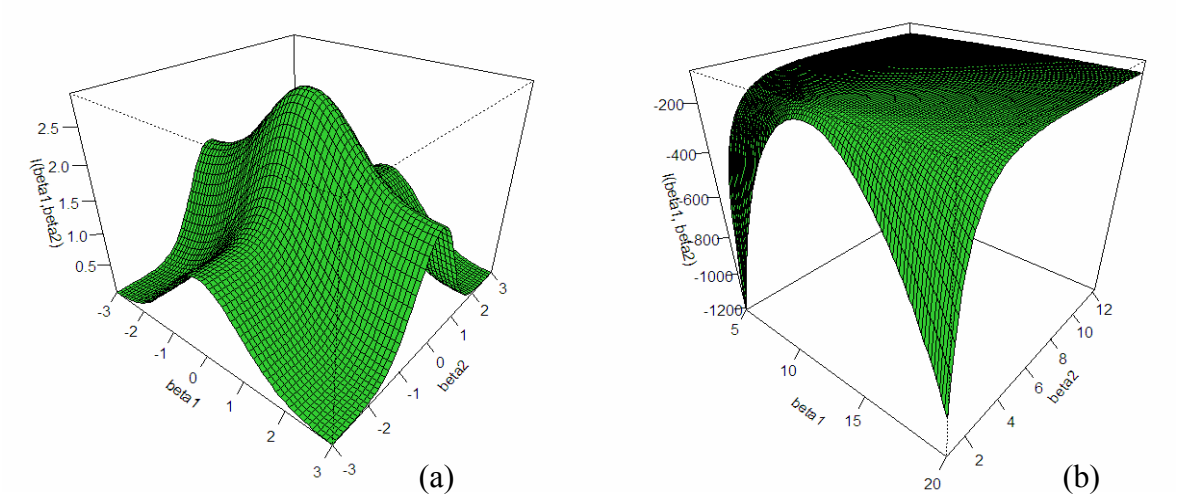


Figura 2.2 – Ilustração de uma função de verossimilhança, com estimativas finitas (a) e infinitas (b).

Heinze e Schemper (2002) propuseram as seguintes soluções para tratar uma situação em que se observa separação completa ou quase-completa: Omissão da covariável do modelo, utilização de uma função de ligação diferente da *logit* para o modelo de regressão logística, manipulação de dados, regressão logística exata, modificação da função escore proposta por Firth (1993), sendo a última recomendada por estes autores.

O método de máxima verossimilhança penalizada proposto por Firth (1993) fornece uma solução simples, válida e fácil de ser implementada em problemas de separabilidade (Zorn, 2005). Este método não envolve manipulação arbitrária de dados nem modificações complicadas de modelos padrão. Ele também não altera a interpretação dos modelos e esta disponível em pacotes estatísticos. Ainda, segundo Zorn (2005), talvez a melhor vantagem seja que este procedimento é, assintoticamente equivalente ao método de máxima verossimilhança genuína no caso de amostras grandes e, superior no caso de pequenas ou médias amostras, onde a separabilidade é mais provável de ocorrer.

2.2.1 – O método de máxima verossimilhança penalizada

O método de máxima verossimilhança penalizada foi originalmente desenvolvido por Firth (1993). A finalidade deste método é reduzir o viés de primeira ordem das estimativas de máxima verossimilhança genuína, produzindo estimativas finitas para os parâmetros do modelo. A modificação proposta por Firth (1993) foi introduzir um pequeno viés na função escore. Segundo este autor, se o parâmetro alvo é o parâmetro canônico de uma família exponencial, o método simplesmente penaliza a verossimilhança pela distribuição a *priori* invariante de Jeffreys (Jeffreys, 1946), que corresponde a informação de Fischer. Para outras parametrizações do modelo da família exponencial ou não exponencial uma escolha para correção do viés está disponível usando informações observadas e esperadas, mas este método resulta numa perda de eficiência de segunda ordem (Firth, 1993). Especificamente em modelos de regressão logística é desejável a penalização para produzir estimativas finitas para os parâmetros da regressão logística na presença de separabilidade.

Quando as estimativas são obtidas por máxima verossimilhança genuína, as soluções são encontradas usando a função escore,

$$U_j(\beta) = \frac{\partial l(\beta)}{\beta_j}, \quad j = 1, 2, ..., p+1, \quad \text{tal que } U_j(\beta) = 0, \tag{1.23}$$

no entanto na presença de separabilidade, Firth (1993) sugere a estimação baseada na função escore modificada, dada por:

$$U_j(\beta)^* = U_j(\beta) + \frac{1}{2} traço\left[I(\beta)^{-1} \left\{ \frac{\partial I(\beta)}{\partial \beta_j} \right\} \right], \; j = 1, 2, ..., p+1.$$

A função escore modificada $U_j(\beta)^*$ esta relacionada com a função logarítmica da verossimilhança penalizada $l(\beta)^* = l(\beta) + \frac{1}{2}\ln|I(\beta)|$ e com a função de verossimilhança penalizada $L(\beta)^* = L(\beta)|I(\beta)|^{\frac{1}{2}}$, onde a penalização $|I(\beta)|^{\frac{1}{2}}$ tem efeito assintoticamente desprezível (Zorn, 2005).

Aplicando o método geral de Firth (1993) para o modelo de regressão logística a função escore (1.14) é substituída pela função escore modificada $U_j(\beta)^* = \sum_{i=1}^{n}\left\{y_i - \pi_i + h_i\left(\frac{1}{2} - \pi_i\right)\right\}x_{ij}$, $j = 1, 2, ..., p+1$ onde os $h_i{'}s$ são os elementos da diagonal da matriz $\hat{H} = W^{\frac{1}{2}}X(X^TWX)^{-1}X^TW^{\frac{1}{2}}$ e $W = diag\{\pi_i(1-\pi_i)\}$, i=1,...,n. As estimativas podem ser obtidas iterativamente pelo método usual de convergência (Collett, 1994) em $U_j(\beta)^* = 0$ e, $\beta^{(s+1)} = \beta^s + I^{-1}(\beta^{(s)})U(\beta^{(s)})^*$ onde, o sobrescrito se refere à s-ésima iteração.

O método de penalização proposto por Firth (1993), encontra-se implementado em alguns programas computacionais. Como exemplos podem ser citadas as bibliotecas **logistf e brglm** (Kosmidis e Firth, 2008), todas estas implementações estão no software R (R Development Core Team, 2009). Outra alternativa é a macro (fl) do software SAS®, que atualmente foi implementada no procedimento **PROC LOGISTIC** com a opção **FIRTH** do software SAS® 9.2 (SAS, 2009).

2.2.2 – Testes estatísticos sob separabilidade

O teste de Wald é um dos mais utilizados para fazer inferências sobre os parâmetros do modelo logístico. Entretanto, Hauck e Donner (1977) investigando o problema de utilizar o teste de Wald, considerando o modelo logístico binomial com um único parâmetro, tendo em vista resultados de simulações sob Ho, observaram que o mesmo, para determinados tamanhos de amostras, apresenta um comportamento atípico. Estes resultados dizem respeito as grandes diferenças entre os valores estimados e o valor paramétrico, neste caso zero, e também a tendência da estatística do teste em assumir zero, implicando em baixo poder do teste. Ainda em relação a este baixo poder do teste de Wald,

Agresti (2002) relata que o teste da razão de verossimilhanças (TRV) é mais confiável e também mais realista para pequenas amostras.

Quando os dados estiverem na configuração de separabilidade, segundo Heinze e Schemper (2002), o teste de Wald resultará em intervalos de confiança com amplitude infinita, que é consequência da obtenção de estimativas imprecisas para os parâmetros sujeitos a esta condição. Portanto, sob a configuração de separabilidade e quando o método de estimação é o de máxima verossimilhança genuína não se recomenda a utilização da estatística de Wald para fazer inferências.

O teste da razão de verossimilhanças mesmo sendo preferível por vários autores tais como Hauck e Donner (1977) e Agresti (2002), quando a configuração de dados está sob separabilidade e o método de estimação é o de máxima verossimilhança genuína o comportamento do poder desta estatística teste não foi avaliado na literatura e será investigado neste trabalho.

Por outro lado inferências quando o método de estimação é o de máxima verossimilhança penalizada podem ser feitas pelo teste de Wald, pois, tal método de estimação produz estimativas finitas e mais precisas Firth (1993). No entanto , desconhece-se o comportamento do poder desta estatística teste. No Capítulo 3 será avaliado, por meio de simulação Monte Carlo, o poder deste teste, assim como compará-lo com o da estatística TRV.

Capítulo 3 – Avaliação de testes estatísticos em regressão logística sob a condição de separabilidade

Neste capítulo é apresentado os cenários para a simulação de dados utilizados para a comparação dos testes de hipóteses de interesse, o modelo proposto para análise dos dados simulados e resultados comparativo do poder dos testes em questão. Tendo em vista a revisão descrita no Capítulo 2 sobre os testes e suas possíveis aplicações às situações envolvendo separabilidade, confeccionou-se a Tabela 3.1 com o intuito de resumir os testes possíveis de serem comparados na simulação.

Tabela 3.1 – Testes estatísticos sob separabilidade

Testes	Método de estimação	
	MV penalizada	MV genuína
TRV	———————	Possível
Wald	Possível	Não aplicável

Diante das duas alternativas possíveis apresentadas na Tabela 3.1, torna-se interessante sob o ponto de vista estatístico comparar o comportamento do teste de Wald considerando o método MV penalizada com a do TRV considerando o método da MV genuína. Para tanto, foi proposta uma simulação Monte Carlo com o objetivo de avaliar o poder destes testes.

3.1 – Modelo utilizado na simulação dos dados binários

Considerou-se o seguinte modelo no processo de simulação que teve como objetivo avaliar o poder do teste de Wald e o teste da razão de verossimilhanças (TRV).

$$\pi_i = P(Y_i = 1 \mid X_i = x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}, \text{ i = 1, 2,..., } \eta, \text{ em que:} \qquad (3.1)$$

Y_i é a variável binária (Y_i=0 ou Y_i=1);

$\beta_0 \, e \, \beta_1$ são os parâmetros do modelo logístico em questão;

x_i é a variável explicativa binária (x_i=-1 ou x_i =1).

Os diferentes cenários assumidos no estudo de simulação foram definidos pelas combinações mostradas na Tabela 3.2 .

Tabela 3.2 – Valores de β_0, β_1 e η utilizados na simulação

β_0	β_1	$\eta/2$
-5	[-5,5], com variação 0,1	5, 15, 25, 50, 100, 200
-4	[-5,5], com variação 0,1	5, 15, 25, 50, 100, 200
-3	[-5,5], com variação 0,1	5, 15, 25, 50, 100, 200
-2	[-5,5], com variação 0,1	5, 15, 25, 50, 100, 200
-1	[-5,5], com variação 0,1	5, 15, 25, 50, 100, 200
0	[-5,5], com variação 0,1	5, 15, 25, 50, 100, 200
1	[-5,5], com variação 0,1	5, 15, 25, 50, 100, 200
2	[-5,5], com variação 0,1	5, 15, 25, 50, 100, 200

η/2 = número de observações geradas para cada grupo da variável x_i, ou seja, -1 e 1, sendo η o tamanho amostral para cada simulação.

A codificação adotada para x_i (1 e -1) teve como objetivo fazer com que os valores de π_i fossem determinados pelos valores dos dois parâmetros (β_0 e β_1) simultaneamente, de forma que a condição de separabilidade fosse determinada pela combinação de valores assumidos por estes dois parâmetros.

Com a codificação utilizada, observa-se as probabilidades

$$\pi_{i1} = P(Y_i = 1 \mid x_i = -1) = \frac{e^{\beta_0 + \beta_1(-1)}}{e^{\beta_0 + \beta_1(-1)}} = \frac{e^{\beta_0 - \beta_1}}{1 + e^{\beta_0 - \beta_1}}, \tag{3.2}$$

$$\pi_{i2} = P(Y_i = 1 \mid x_i = 1) = \frac{e^{\beta_0 + \beta_1(1)}}{e^{\beta_0 + \beta_1(1)}} = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}. \tag{3.3}$$

Para ilustrar a relevância da codificação usada (x_i = -1 ou x_i = 1), ao fixar $\beta_0 = -3$ e variar β_1, tem se as curvas de probabilidade apresentadas na Figura 3.1.

De acordo com a Figura 3.1, ao optar por $\beta_1 = -2$, e tendo em vista o valor fixo $\beta_0 = -3$, observa-se que $\pi_{i1} = 0{,}2689$ e $\pi_{i2} = 0{,}00669$, (dadas respectivamente pelas probabilidades de sucesso quando x_i = -1 e x_i = 1). Como π_{i2} é muito pequena, espera-se que para x_i = 1 a maioria absoluta dos valores observados de Y_i sejam zero (fracasso), uma vez que esta variável é gerada por meio das probabilidades em questão.

Este processo de geração de valores de Y_i mediante π_{i1} e π_{i2} consiste simplesmente em gerar N valores de uma distribuição de Bernoulli (π_{i1}) e outros N valores de uma outra distribuição Bernoulli (π_{i2}).

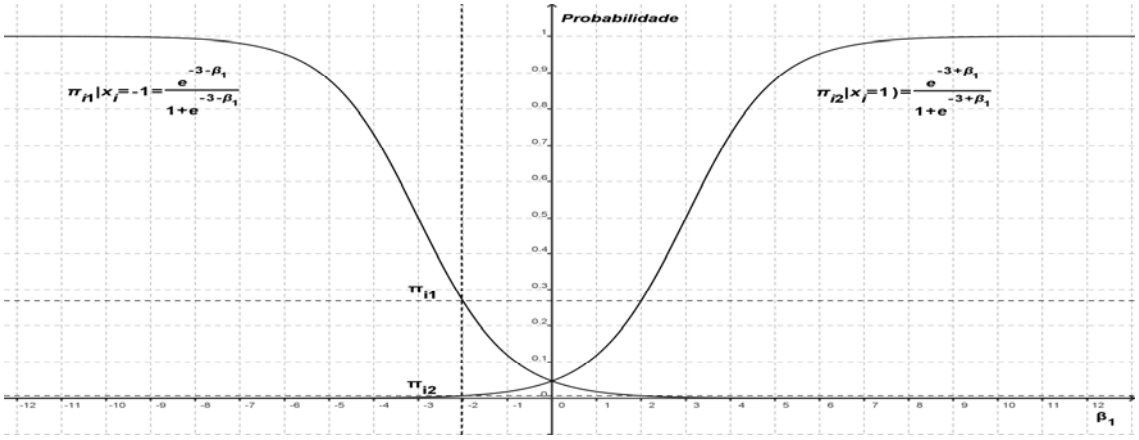


Figura 3.1 – Probabilidade de sucesso obtidas ao se variar β_1 e fixar $\beta_0 = -3$.

O processo de simulação descrito anteriormente foi repetido n = 2000 vezes para cada cenário, isto é, para cada combinação de valores de β_0, β_1 e η. Deste total de n repetições foram calculadas as proporções de conjuntos de dados que se classificavam de acordo com as três possíveis configurações: separabilidade completa, quase completa e casos comuns (*overlap)* (item 2.1 do Capítulo 2). Tais proporções são apresentadas em tabelas do apêndice A com o intuito de auxiliar na avaliação do poder dos testes estudados.

Na Figura 3.2 são apresentadas as curvas de probabilidade que determinam todos os possíveis valores para π_{i1} e π_{i2} provenientes de todas as combinações entre os valores considerados para β_0 e β_1.

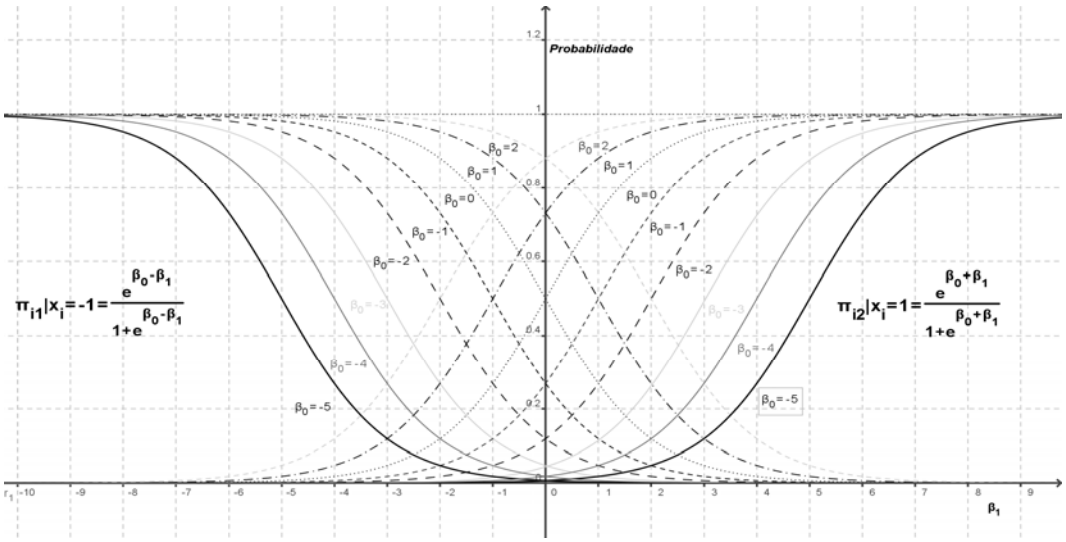


Figura 3.2 – Probabilidade de sucesso obtidas ao se variar β_0 e β_1.

3.2 – Análises dos dados simulados e critérios de comparação

Os dados simulados no item anterior foram analisados por meio do modelo logístico (Hosmer e Lemeshow, 1989):

$$Y_i = E(Y_i \mid x_i) + e_i = \pi_i + e_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} + e_i \text{, em que:} \qquad (3.4)$$

π_i é a $P(Y_i = 1 \mid X_i = x_i)$ conforme expressão (3.1);

e_i é o termo de erro aleatório, em que $E(e_i) = 0$ e $\mathrm{V}(e_i) = \pi_i(1 - \pi_i)$.

O modelo apresentado em (3.4) foi ajustado aos dados gerados pelo processo de simulação descrito na seção 3.1 considerando os dois métodos apresentados na Tabela 3.1: máxima verossimilhança genuína e máxima verossimilhança penalizada. Para tanto, foram utilizados, respectivamente, *glm* e *brglm* do software R (R Development Core Team, 2009).

Dentre todos os cenários definidos por combinações de valores de β_0 e β_1, um em especial, caracterizado por β_1=0, representa a condição em que os dados foram simulados sob a hipótese de nulidade, ou seja: $H_0 : \beta_1 = 0$ vs $H_a : \beta_1 \neq 0$. A relevância desta condição está fundamentada no fato da mesma permitir a avaliação do poder dos dois testes propostos na Tabela 3.1. Isto é porque, sob a configuração de separabilidade não se sabe ao certo o poder do teste de razão de verossimilhanças, quando se utiliza o método da máxima verossimilhança genuína e nem o poder do teste de Wald quando se utiliza o método de máxima verossimilhança penalizada.

Em uma análise de simulação de dados, uma forma prática e eficiente de se comparar o poder entre diferentes testes é por meio de uma análise gráfica, a qual consiste em plotar a proporção de rejeição de Ho em função dos valores considerados para o parâmetro testado nesta mesma hipótese. A proporção em questão é calculada pela razão entre o número de repetições da simulação na qual o valor da estatística do teste foi maior que um valor tabelado, após a especificação de um dado nível de significância, e o número total de repetições usado na simulação.

Uma ilustração é apresentada na Figura 3.3, na qual se observa, por exemplo, que para β_1= -1,5 a curva de poder do teste A fornece uma proporção de rejeição de 0,33, enquanto que a curva do teste B fornece uma proporção de 0,01. Como o valor de β_1 não corresponde ao valor considerado em Ho (Ho: β_1=0), conclui-se que o teste com maior

proporção de rejeição de Ho será o mais poderoso, neste caso tal teste é o A. Verifica-se no gráfico em questão, que este resultado se repete para todos os valores de β_1.

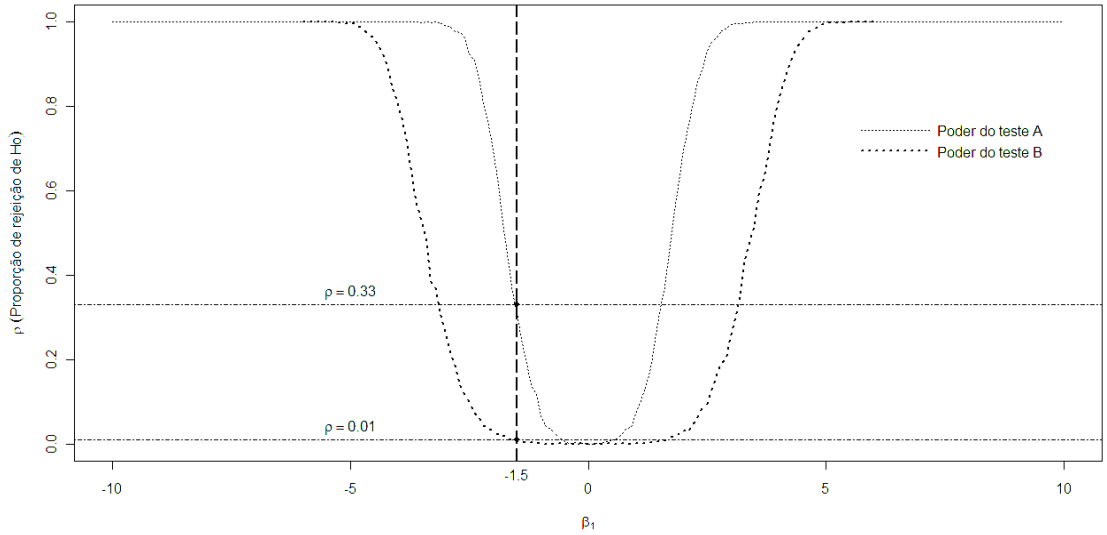


Figura 3.3 – Ilustração das curvas de poder para os testes A e B.

No presente estudo, curvas similares às apresentadas na Figura 3.3, foram confeccionadas para o TRV e de Wald para diferentes situações, as quais foram representadas pelas variações de β_0 e η, que por sua vez representam diferentes proporções de separabilidade nos dados gerados. Para tanto, foi utilizado um nível de significância de 5%, de forma que para rejeitar Ho, os valores das estatísticas dos testes em questão foram comparados com o valor 3,84, tendo em vista a distribuição aproximada $\chi^2_{\nu=1}$.

Outra forma de estudar um teste de hipótese é por meio da avaliação de seu comportamento assintótico sob Ho, pois ao aumentar o tamanho da amostra, espera-se que a proporção de rejeição desta hipótese apresente uma convergência para o nível de significância adotado. Dessa forma, uma análise gráfica deste processo permite comparar diferentes testes por meio da visualização da velocidade com que estes convergem para α%, assim o teste que atinge tal valor com um tamanho de amostra menor é aquele que apresenta melhor comportamento assintótico.

Uma ilustração é apresentada na Figura 3.4, na qual observa-se que sob Ho ambos os testes (C e D) comparados a um nível de 5% de significância apresentam a referida convergência, porém esta é atingida mais rapidamente pelo C, implicando em uma melhor performance deste teste sob o ponto de vista assintótico.

No presente trabalho gráficos similares ao apresentado na Figura 3.4 foram confeccionados com o intuito de avaliar o comportamento assintótico do TRV e de Wald

considerando diferentes valores de β_0, os quais representam as diferentes proporções de separabilidade nos dados gerados.

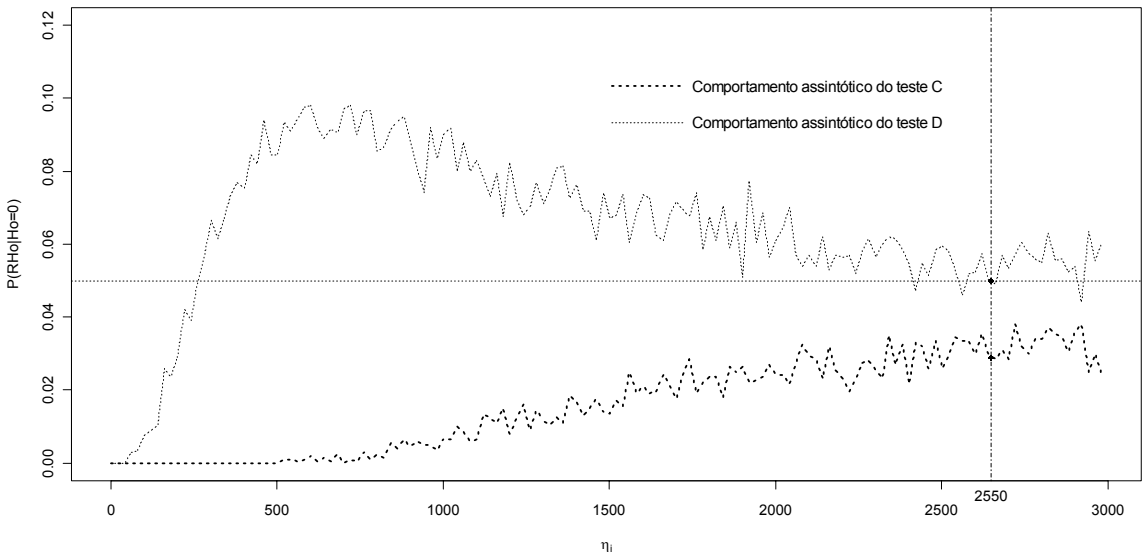


Figura 3.4 – Comportamento assintótico dos testes C e D.

3.3 – Resultados e discussão

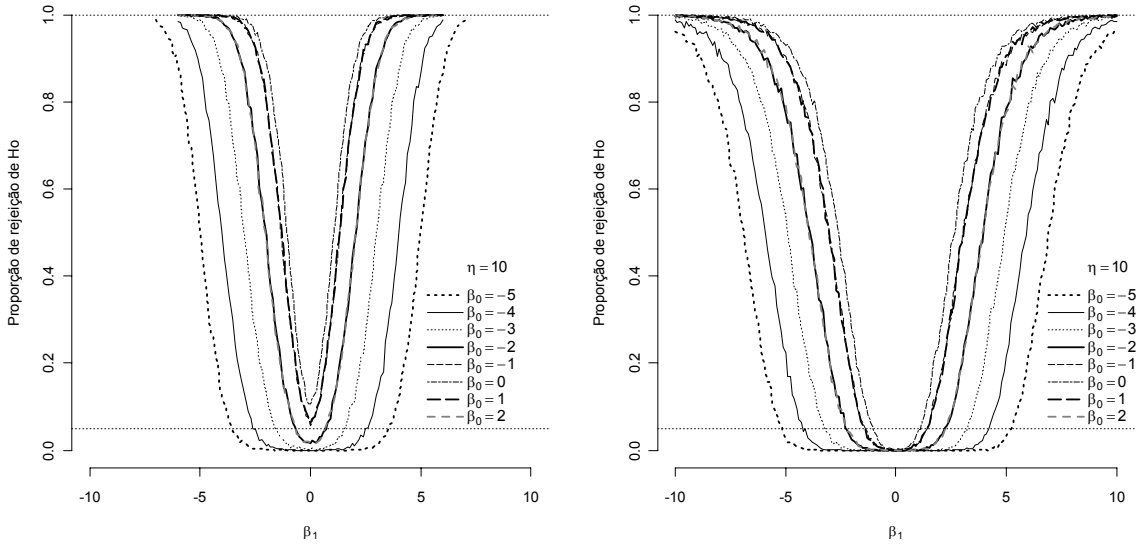
As Tabelas A1, A2, A3, A4, A5, A6, A7 e A8, todas apresentadas no Apêndice A, mostram as proporções de cada categoria de conjunto de dados (separação quase completa, separação completa e casos comuns ou *overlap*) simulados considerando os cenários definidos pelas combinações dos valores assumidos para η, β_0 e β_1.

Observa-se que a proporção de conjuntos de dados na configuração de separabilidade depende principalmente do tamanho da amostra ($2N = \eta$), uma vez que quanto maior este valor, maior é a quantidade esperada de sucessos em situações desfavoráveis para a ocorrência dos mesmos. Este fato descaracteriza a condição de separabilidade, ou seja, a ausência de sucesso. Estas situações desfavoráveis são verificadas para baixos valores de π_{i1} ou para baixos valores de π_{i2}, uma vez que estes desfavorecem respectivamente a ocorrência de sucesso para os valores de x_i=-1 e x_i=1. Em resumo, se η é grande, independentemente dos valores de $\beta_0, \beta_1, \pi_{i1} e \pi_{i2}$, maior a proporção de casos comuns (*overlap*). Como exemplo, nota-se que para $\beta_0 > -3$, conforme Tabelas A4, A5, A6, A7 e A8, e $N > 100$, tem-se uma quase totalidade de *overlap*, ou seja, ausência de separabilidade, situação na qual Agresti (2002) relata que a superioridade do TRV sobre de Wald já é conhecida.

De forma geral, as Tabelas A1 a A8 mencionadas anteriormente têm como objetivo auxiliar a interpretação dos gráficos representativos das curvas de poder de cada teste. Uma vez que o objetivo é avaliar o poder dos testes em conjuntos de dados na configuração de separabilidade, a indicação da proporção de dados nesta configuração encontra-se nas Tabelas em questão.

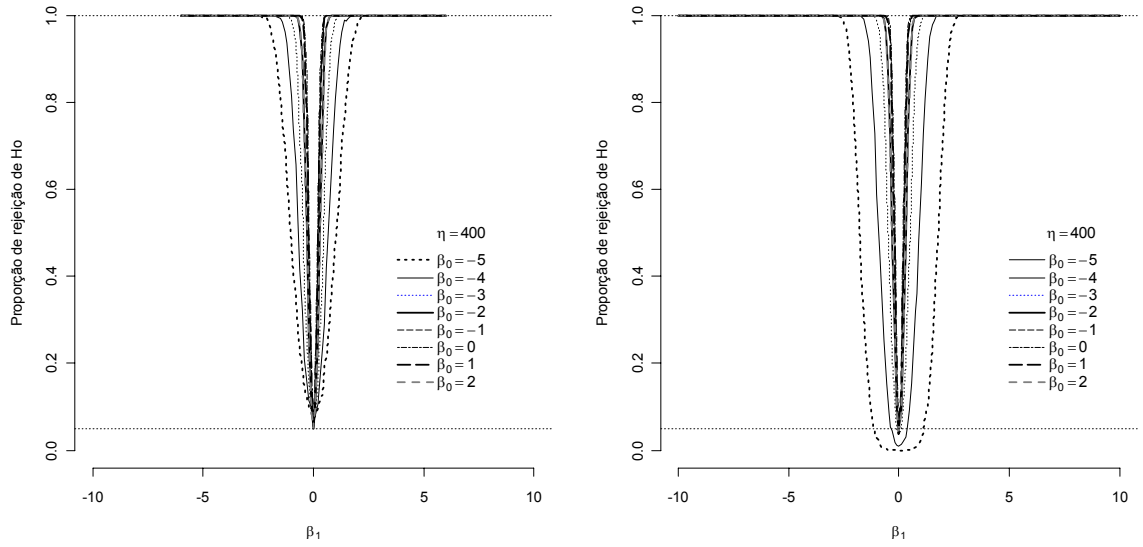
Nas Figuras 3.5 e 3.6 são apresentadas as curvas de poder dos dois testes obtidas respectivamente para o menor (η=10) e maior (η=400) tamanhos de amostra. Nota-se nestas Figuras que para η=400 (Figura 3.6) as curvas para ambos os testes são mais fechadas que aquelas observadas para η=10 (Figura 3.5), evidenciando que aumentado o tamanho da amostra os testes ganham poder. Observa-se em todas figuras que as curvas mais abertas, em que β_0 são menores, são aquelas para os cenários com maiores proporções de separabilidade, mostradas nas Tabelas do Apêndice A.

Tendo em vista que o principal objetivo do presente trabalho é avaliar os testes sob condição de separabilidade, há indícios que para η=10 (Figura 3.5) o TRV apresenta curvas mais fechadas em relação ao teste de Wald. Isto é um indicativo que o mesmo é mais poderoso.



(a) (b)

Figura 3.5 – Função poder empírica dos testes da razão de verossimilhanças (TRV) (a) e de Wald (b) para amostras de tamanho η=10.



(a) (b)

Figura 3.6 – Função poder empírica dos testes da razão de verossimilhanças (TRV) (a) e de Wald (b) para amostras de tamanho η=400.

Uma vez que baixos valores de β_0 também propiciam a condição de separabilidade (Figura 3.2), as Figuras 3.7 e 3.8 mostram as curvas de poder de ambos os testes para o menor valor de β_0 considerado ($\beta_0 = -5$), respectivamente para o menor e maior tamanho amostral.

Nota-se nas Figuras 3.7 e 3.8 as quais são partes das Figuras 3.5 e 3.6, que realmente o TRV é mais poderoso que o teste de Wald sob separabilidade, pois em ambos os gráficos esta condição está presente, porém com maior incidência para η=10 (Figura 3.7) conforme Tabela A1 mostrada no apêndice. As Figuras A13 a A16 apresentadas no apêndice mostram o poder dos testes para cenários intermediários.

O tamanho empírico, nível descritivo, dos testes em questão é mostrado nas Figuras 3.9 (a) e (b), e pode ser verificado, que os dois testes convergem para o nível descritivo quando η cresce.

Nota-se que nas Figuras 3.9 (a) e (b) a proporção de rejeição de Ho utilizando-se a estatística TRV e da estatística de Wald convergem para o nível descritivo estabelecido em todos os cenários, porém, a estatística de Wald converge mais lentamente. Esta convergência pode ser observada mais facilmente para β_0= -5 que é o cenário no qual apresenta maior proporção de separabilidade como pode ser observado nas Tabelas A1 a A8.

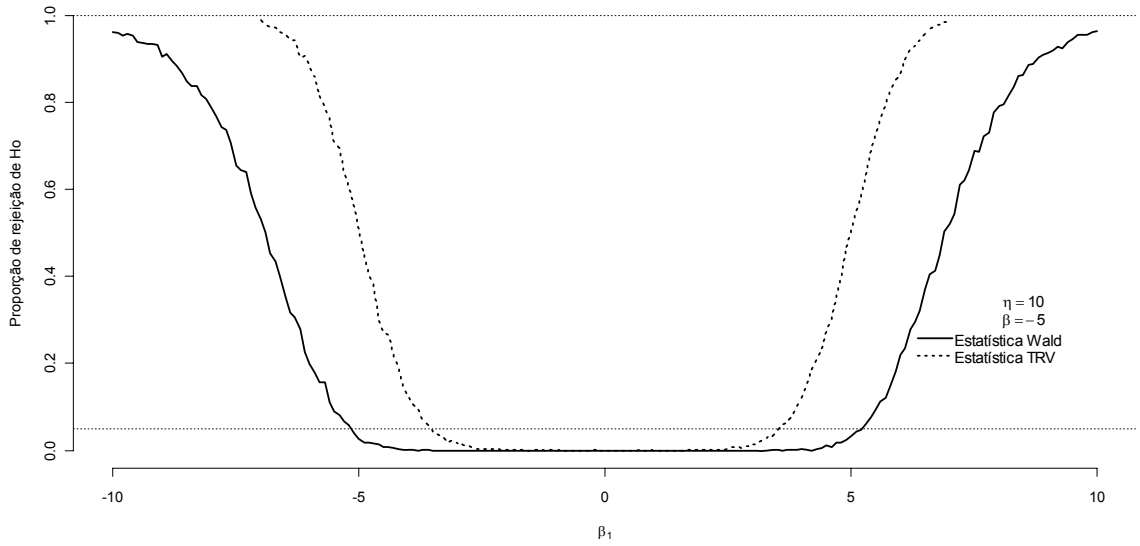


Figura 3.7 – Poder do TRV e Wald para amostras de tamanho $\eta = 10$ e $\beta_0 = -5$.

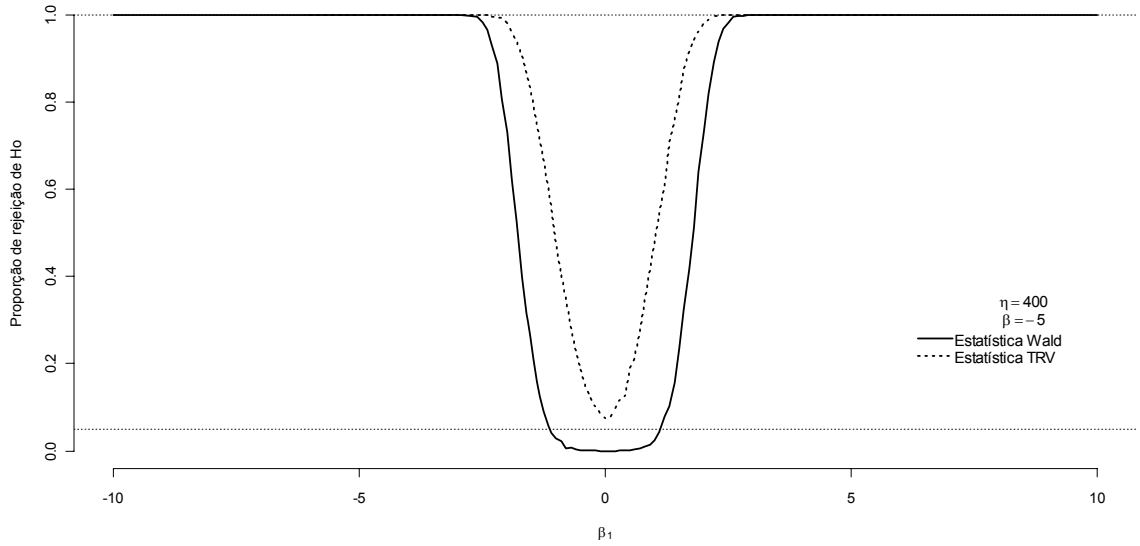
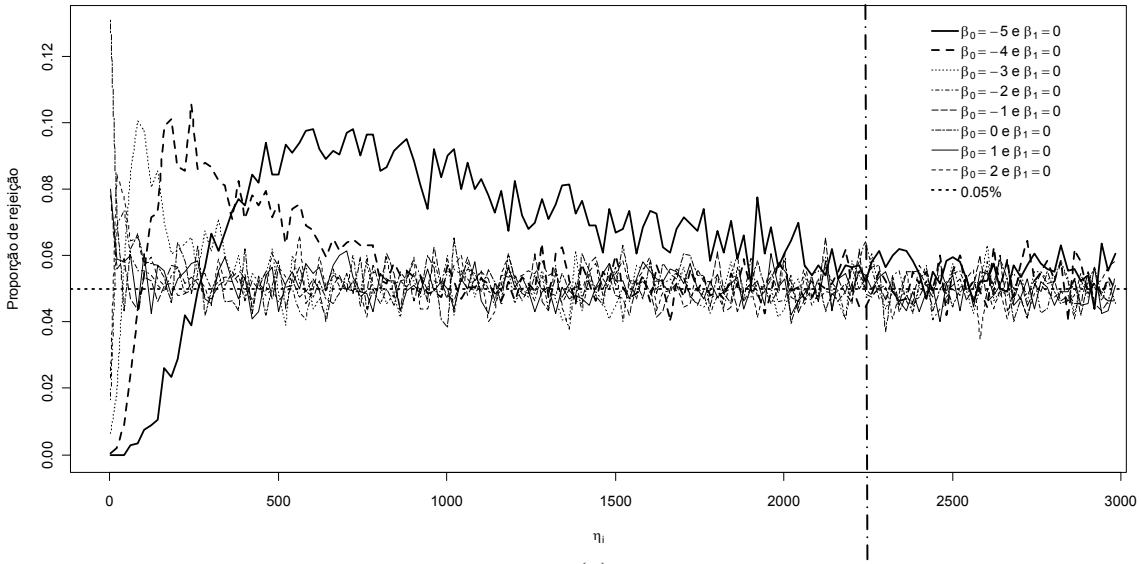
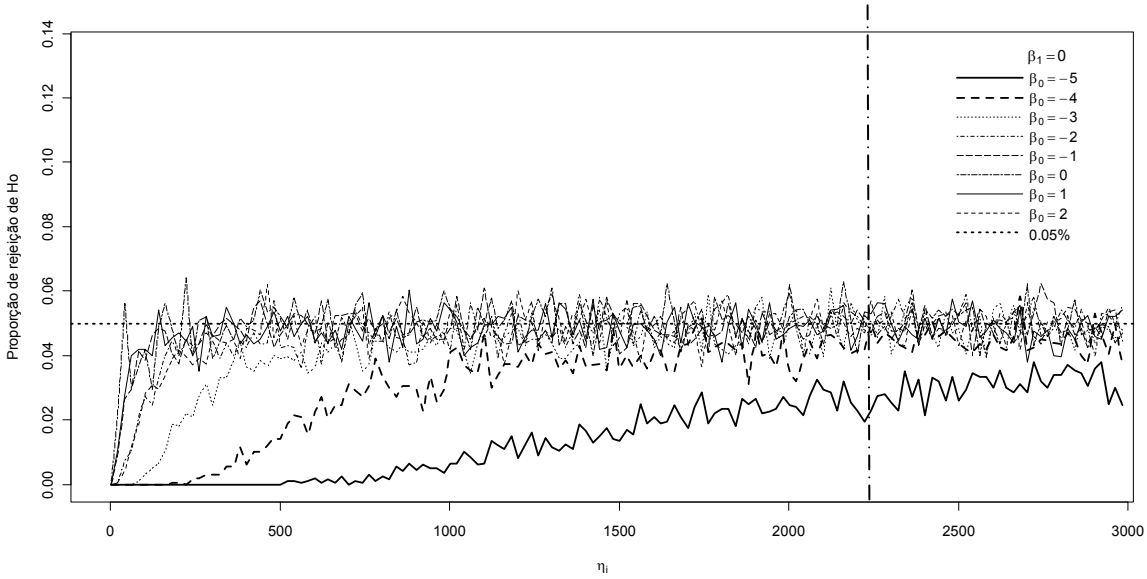


Figura 3.8 – Poder do TRV e Wald para amostras de tamanho $\eta = 400$ e $\beta_0 = -5$.

Observa-se nas mesmas Figuras 3.9 (a) e (b) que para $\eta \geq 2250$ e para alguns cenários, que tal convergência ainda não foi atingida para a estatística de Wald enquanto, este nível já foi atingido para a estatística do TRV a partir de $\eta \geq 1900$, evidenciando uma convergência mais rápida para o TRV, principalmente sob a condição de separabilidade.



(a)



(b)

Figura 3.9 – Probabilidade do erro tipo I com variações de β_0 e tamanhos de amostras, para as estatísticas TRV (a) e de Wald (b).

Capítulo 4 – Aplicação

Neste capítulo são apresentadas aplicações da metodologia a dois conjuntos de dados reais. Um destes bancos de dados é conhecido na literatura (Colosimo *et al*., 1995), o outro conjunto de dados é oriundo de um experimento com germinação de sementes de *Adenanthera pavonina* L.

4.1 – Pacientes submetidos a craniotomia

No conjunto de dados apresentado por Colosimo *et al*. (1995), considerou-se um grupo de 102 pacientes submetidos à craniotomia em que a resposta de interesse Y_i foi a ocorrência ou não de meningite nos primeiros 30 dias após a cirurgia. Duas covariáveis são consideradas na explicação desta variável resposta Y_i: X1 gravidade do caso (0 = baixa e 1 = alta) e, X2 tempo de duração da cirurgia (horas) com N indivíduos por categoria.

Tabela 4.1 – Conjunto de dados dos pacientes submetidos à craniotomia

$N1=Y_i$	N	X1	X2	$N1=Y_i$	N	X1	X2
0	1	0	2,5	0	1	0	2,17
0	1	1	1,33	0	1	1	6,5
0	2	1	6	0	3	1	1
0	1	0	4,5	0	4	1	4
0	3	0	1,5	0	8	1	3
0	4	0	1,33	0	8	0	4
0	3	0	5	0	1	0	4,75
0	1	1	0,75	0	13	0	3
0	8	0	2	0	1	1	8
0	3	0	3,5	0	1	0	5,5
0	1	1	3,25	0	1	0	2,67
0	4	0	1,83	0	1	0	2,25
0	1	1	7	0	2	0	7
0	1	0	1,67	0	1	0	3,67
0	1	0	8	0	1	0	2,33
0	1	1	3,5	0	1	0	6,5
0	1	0	3,17	0	3	0	1
0	1	1	5,5	0	3	0	6
0	6	1	2	1	2	1	1,5
0	1	0	1,25	1	1	1	10

A Tabela 4.2 apresenta a distribuição conjunta dos pacientes submetidos à craniotomia segundo a gravidade do caso e a ocorrência de meningite.

Tabela 4.2 – Distribuição dos pacientes segundo a gravidade do caso e a presença de meningite

Gravidade (X1)	Ocorrência de meningite (y_i)		Total
	Sim	Não	
Baixa	0	68	68
Alta	2	32	34
Total	2	100	102

Na Tabela 4.2 observa se uma casela vazia de acordo com Nacle (2004), cujo conjunto de dados está na configuração de separação quase completa. Ainda segundo Heinze e Schemper, (2002) outra alternativa para verificar tal configuração é monitorar a variância para um dado parâmetro. Esta alternativa é apresentada na Tabela 4.3.

Tabela 4.3 – Estimativas de máxima verossimilhança genuína para os coeficientes do modelo de regressão logística para os dados de craniotomia

Software	Coeficientes	Estimativas	Erro Padrão
R – glm	β_0	-23,58	4530,96
	β_1	19,05	4530,96
	β_2	0,40	0,28

Diante das alternativas apresentadas na Tabela 3.1, o TRV é o teste mais indicado para fazer inferências quando for utilizado o método de máxima verossimilhança genuína. Os resultados deste teste estão apresentados na Tabela 4.4.

Os modelos nulo, apenas com X1, apenas com X2 e com X1 + X2 apresentaram, respectivamente, os seguintes valores de *deviance* residual: 16,92, 12,44, 14,08 e 10,44. Com base nestes valores, avaliou-se os efeitos mostrados na Tabela 4.4.

Tabela 4.4 – Teste da razão de verossimilhanças (TRV) para as estimativas de máxima verossimilhança genuína

Efeito	Hipótese nula	TRV	valor – p
Intercepto	-	-	-
X1 | X2	$\beta_1=0$	(14,08 - 10,44) =3,64	0,0560
X2 | X1	$\beta_2=0$	(12,44-10,44) = 2,00	0,1572
X1+X2	$\beta_1= \beta_2=0$	(16,92-10,44) = 6,48	0,0394

$X_i|X_j$: Efeito de X_i dado que X_j está no modelo

Na Tabela 4.4 pode ser verificado os efeitos das covariáveis X1 e X2, cujos resultados mostram não haver evidências de efeitos significativo de X1 e X2 isoladamente ao nível de 5% de probabilidade. No entanto, pode ser verificado ainda, que quando estas variáveis estão conjuntamente no modelo, há evidencias de efeito significativo ao nível de 5% de probabilidade.

No entanto Heinze e Schemper, (2002) recomendam a modificação da função escore proposta por Firth (1993), cujo resultado está apresentado na Tabela 4.5. Pode ser verificado que o erro padrão observado para β_0 e β_1 foi corrigido de 4530,96 (Tabela 4.3) para 1,71 e 1,53 (Tabela 4.5).

Tabela 4.5 – Estimativas de máxima verossimilhança penalizada para os coeficientes do modelo de regressão logística para os dados de craniotomia

Software	Coeficientes	Estimativas	Erro Padrão
R – brglm	β_0	-6,14	1,71
	β_1	2,17	1,53
	β_2	0,36	0,24

Conforme a Tabela 3.1, outra alternativa para inferência, no caso de se trabalhar com o método de máxima verossimilhança penalizada, é o teste de Wald. Os resultados deste teste estão apresentados na Tabela 4.6.

Tabela 4.6 – Testes individuais de Wald para as estimativas de máxima verossimilhança penalizada

Efeitos	Hipótese nula	Wald	$Wald^2$	valor – p
Intercepto	-	-	-	-
X1 | X2	$\beta_1=0$	1,41	2,00	0,1600
X2 | X1	$\beta_2=0$	1,48	2,20	0,1300
X1+X2	$\beta_1=\beta_2=0$	2,19	4,80	0,0930

$X_i|X_j$: Efeito de X_i dado que X_j está no modelo

Na Tabela 4.6 estão apresentados o teste de significância das covariáveis X1, X2 e X1+X2 cujos resultados mostram evidência que nenhum dos efeitos testados são significativos ao nível de 5% de probabilidade.

Levando-se em consideração os resultados da simulação discutida no Capítulo 3, em que o TRV apresentou ter maior poder que o teste de Wald, há uma indicação de que a inferência apresentada na Tabela 4.4 é mais confiável que a apresentada na Tabela 4.6.

4.2 – Germinação de sementes de *Adenanthera pavonina* L.

No teste de germinação de sementes de *Adenanthera pavonina* L, foi avaliado o número de sementes germinadas em quatro repetições de 25 sementes. A avaliação foi realizada até o décimo quinto dia após a semeadura, sendo consideradas germinadas as sementes que apresentaram protusão de raiz primária superior a 2 mm. Para a superação da dormência tegumentar utilizou-se os seguintes métodos pré-germinativos: desponte com alicate no lado oposto ao hilo, atrito em lixa de madeira (n°36) no lado oposto ao hilo, imersão em ácido sulfúrico concentrado por 20 minutos, e testemunha.

Ensaios para estudos do efeito da temperatura e dos métodos pré-germinativos na germinação das sementes foram realizados nas seguintes isotermas: 20, 25 e 30ºC, sob luz contínua, em incubadoras do tipo B.O.D., com sementes distribuídas em placas de Petri sobre duas folhas de papel germitest saturado com água destilada. Na Tabela 4.7 estão apresentados os resultados do experimento e na qual a temperatura está indicada por X1 e os métodos pré-germinativos por X2.

Tabela 4.7 – Conjunto de dados *Adenanthera pavonina* L

X1	X2	Germinou	Não Germinou	X1	X2	Germinou	Não Germinou
20	Alicate	21	4	25	Ácido	15	10
20	Alicate	20	5	25	Ácido	21	4
20	Alicate	23	2	25	Ácido	12	13
20	Alicate	24	1	25	Acido	23	2
20	Lixa	24	1	25	Testemunha	0	25
20	Lixa	23	2	25	Testemunha	0	25
20	Lixa	21	4	25	Testemunha	0	25
20	Lixa	20	5	25	Testemunha	0	25
20	Ácido	21	4	30	Alicate	10	15
20	Ácido	24	1	30	Alicate	12	13
20	Ácido	23	2	30	Alicate	11	14
20	Ácido	23	2	30	Alicate	9	16
20	Testemunha	0	25	30	Lixa	11	14
20	Testemunha	0	25	30	Lixa	21	4
20	Testemunha	0	25	30	Lixa	19	6
20	Testemunha	0	25	30	Lixa	16	9
25	Alicate	20	5	30	Ácido	13	12
25	Alicate	19	6	30	Ácido	14	11
25	Alicate	20	5	30	Ácido	19	6
25	Alicate	18	7	30	Ácido	19	6
25	Lixa	21	4	30	Testemunha	0	25
25	Lixa	22	3	30	Testemunha	0	25
25	Lixa	22	3	30	Testemunha	0	25
25	Lixa	22	3	30	Testemunha	0	25

A Tabela 4.8 apresenta o número de sementes germinadas por tratamento utilizados no experimento.

Tabela 4.8 – Número de sementes germinadas de *Adenanthera pavonina* L por tratamento

X1	X2	Germinação		Total
		Sim	Não	
20	Alicate	88	12	100
	Lixa	88	12	100
	Acido	91	9	100
	Testemunha	0	100	100
25	Alicate	77	23	100
	Lixa	87	13	100
	Acido	71	29	100
	Testemunha	0	100	100
30	Alicate	42	58	100
	Lixa	67	33	100
	Acido	65	35	100
	Testemunha	0	100	100
Total		676	524	1200

Na Tabela 4.8 ao combinarmos X2 dentro dos níveis de X1 observa-se uma casela vazia indicando que este conjunto de dados está na configuração de separação quase completa, ainda segundo Heinze e Schemper, (2002) outra alternativa para verificar tal configuração é monitorar a variância para um dado parâmetro. Esta alternativa é apresentada na Tabela 4.9.

Tabela 4.9 – Estimativas de máxima verossimilhança genuína para os coeficientes do modelo de regressão logística para os dados de germinação de *Adenanthera pavonina* L

Software	Covariáveis	Estimativas	Erro Padrão
R – glm	Intercepto	4,96	0,95
	X2: Método Alicate	2,14	1,38
	Lixa	0,13	1,41
	Testemunha	-26,93	12768,45
	X1 : Temperatura	-0,15	0,04
	X1*X2 (Temp.*Alicarte)	-0,10	0,05
	(Temp.*Lixa)	0,01	0,05
	(Temp.*Testemunha)	0,15	504,06

Diante das alternativas apresentadas na Tabela 3.1, o TRV é o teste mais indicado para fazer inferências quando for utilizado o método de máxima verossimilhança genuína.

Foi testado primeiramente a interação entre as variáveis X1 e X2. Os modelos sem e com a interação apresentaram, respectivamente, os seguintes valores de *deviance* residual: 56,66 e 51,85. Com base nestes valores, avaliou-se os efeitos mostrados na Tabela 4.10, onde pode ser verificado que a interação foi não significativa a 5% de probabilidade.

Seguindo a sugestão de Colosimo e Giolo (2006) mesmo a interação sendo não significativa foi realizada testes individuais para os paramentos desta interação, os quais mostraram ser também não significativos, desta forma optou-se por um modelo sem a interação.

Tabela 4.10 – Teste da razão de verossimilhanças (TRV) para verificar o efeito da interação entre X1 e X2

Efeito	Hipótese nula	TRV	valor–p
Intercepto	-	-	-
X1+X2			
X1+X2+X1*X2	$\beta_{interação}= 0$	(56,66-51,85)= 4,81	0,1863

Para avaliar os efeitos de X1, X2 e X1 + X2, mostrados na Tabela 4.11, foram obtidas as *deviances* residuais do modelo nulo, somente com X1, somente com X2 e com X1 + X2: 783,10, 738,72, 137,86, e 56,66 respectivamente. Pode ser verificado na presente Tabela que todos os efeitos testados foram significativos.

Tabela 4.11 – Teste da razão de verossimilhanças (TRV) para verificar o efeito de X1, X2 e X1+X2

Efeito	Hipótese nula	TRV	valor–p
Intercepto	-	-	-
X1 | X2	$\beta_{temp}= 0$	(137,86 - 56,66) = 81,20	0,0000
X2 | X1	$\beta_{método}= 0$	(738,72 - 56,66) = 682,06	0,0000
X1+X2	$\beta_{método}= \beta_{temp}=0$	(783,10 - 56,66) =726,44	0,0000

$X_i|X_j$: Efeito de X_i dado que X_j está no modelo

No entanto Heinze e Schemper, (2002) recomendam a modificação da função escore proposta por Firth (1993) cujo resultado está apresentado na Tabela 4.12. Pode ser verificado que o erro padrão observado para X2:Testemunha e X1*X2: (Temp.*Testemunha) foi corrigido respectivamente de 12768,45 e 504,06 (Tabela 4.9) para 6,31 e 0,25 (Tabela 4.12).

Tabela 4.12 – Estimativas de máxima verossimilhança penalizada para os coeficientes do modelo de regressão logística para os dados de germinação de *Adenanthera pavonina* L

Software	Covariáveis	Estimativas	Erro Padrão
R-brglm	Intercepto	4,90	0,95
	X2: Método Alicate	2,11	1,37
	Lixa	0,11	1,40
	Testemunha	-10,61	6,31
	X1 : Temperatura	-0,15	0,04
	X1*X2 (Temp.*Alicarte)	-0,09	0,05
	(Temp.*Lixa)	0,01	0,05
	(Temp.*Testemunha)	0,15	0,25

Conforme a Tabela 3.1, outra alternativa para fazer inferência no caso de se trabalhar com o método de máxima verossimilhança penalizada, é o teste de Wald.

Da mesma maneira como foi feito para o TRV, primeiramente foi testado a interação entre as variáveis X1 e X2, cujo efeito está apresentado na Tabela 4.13, em que se nota interação foi não significativa a 5% de probabilidade.

Tabela 4.13 – Teste de Wald para verificar o efeito da interação entre X1 e X2

Efeito	Hipótese nula	Wald	$Wald^2$	valor–p
Intercepto	-	-	-	-
X1+X2				
X1+X2+X1*X2	$\beta_{interação}= 0$	2,28	5,20	0,1600

Para avaliar os efeitos de X1, X2 e X1 + X2 mostrados na Tabela 4.14, foi aplicado o teste Wald para comparar modelos. Pode ser verificado na presente Tabela que todos os efeitos testados foram significativos.

Tabela 4.14 – Teste de Wald para as estimativas de máxima verossimilhança penalizada

Efeitos	Hipótese nula	Wald	$Wald^2$	valor – p
Intercepto	-	-	-	-
X1	$\beta_{temp}= 0$	8,46	71.57	0,0000
X2	$\beta_{método}= 0$	6,50	42,25	0,0000
X1+X2	$\beta_{método}= \beta_{temp}=0$	10,24	104,86	0,0000

Fazendo uma comparação dos valores das Tabelas 4.11 e 4.14 pode-se verificar que os resultados são similares.

Conclusões

Neste trabalho foi discutido o poder do teste da razão de verossimilhanças e de Wald sob a condição de separabilidade. A simulação Monte Carlo com uma variável explicativa no modelo possibilitou obter indícios que o TRV tem maior poder que o teste Wald sob esta condição. No entanto, em trabalhos aplicados, geralmente existe mais de uma variável explicativa, ficando como sugestão em trabalhos futuros realizar simulações com mais de uma variável explicativa no modelo, com intuito de obter um indicativo mais geral sobre o comportamento de tais testes sob a condição de separabilidade. Com esta estrutura de dados simulados no presente trabalho foi possível observar alguns padrões: sob a condição de separabilidade o TRV mostrou ter um poder maior que o teste de Wald independente de β_0 e β_1; aumentado o tamanho da amostra os testes ganham poder; quando a configuração de dados encontra-se na classificação de *overlap* os dois testes tem poder semelhantes, mas, com vantagem para o TRV para todos os cenários independente de β_0 e β_1; a ocorrência de separabilidade está diretamente ligada ao tamanho amostral, pois, para amostras pequenas há muitos zeros, ou seja é uma amostra esparsa e com maior probabilidade de classificação em separabilidade; o comportamento assintótico do TRV converge mais rapidamente para o nível descritivo que o teste de Wald.

REFERENCIAS BIBLIOGRÁFICAS

[1] Agresti, A**. An introduction to Categorical Data Analysis.** New York: John Wiley, 1990. 290p.

[2] Agresti, A. **Categorical Data Analysis**. John Wiley & Sons, Inc. Hoboken, New Jersey, 2002.710p.

[3] Albert A.; Anderson, J.A. On the existence of maximum likelihood estimates in logistic regression models. **Biometrika**, v.71, n.1, p.1-10, 1984.

[4] Allison, P.D. **Logistic regression using the SAS System, theory and application**. SAS Institute, 1999. 304p.

[5] Clarkson, D.B.; Jenrick, R.I. Computing extended maximum likelihood estimates for linear parameter models. **Journal of the Royal Statistical Society**. Series B **(Methodological)**, v.53, **n.2,** p.417-426, 1991.

[6] Collett D. **Modelling Survival Data in Medical Research**. Chapman and Hall, London, 1994.

[7] Cox, D. R. **The analysis of Binary Data**. Methuen, London. 1970

[8] Cox, D.R.; Hinkley, D.V.; **Theoretical statistics.** Londron: Chapman & Hall, 1986. 174 p.

[9] Cox, D.R.; Snell, E. J. **Analysis of Binary Data.** London: Chapman & Hall, 1989. 236p.

[10] Colosimo, E.A.; Giolo, S.R. **Análise de sobrevivência aplicada**. ABE - Projeto Fisher. São Paulo: Edgar Blücher, 2006.

[11] Colosimo, E.A., Franco, G.C., Couto, B.M. The Logistic Regression Model and Rare Events. **Estadística**, 47, 148, 149:1-16. 1995.

[12] Demétrio, C. G. B. **Modelos Lineares Generalizados em Experimentação Agronômica**. Piracicaba, 2002, 121p.

[13] Dobson, A. J. **An introduction to generalized liner models.** London: Chapman & Hall, 1990. 225 p.

[14] Firth, D. Bias reduction of maximum likelihood estimates. **Biometrika**, v.80, n.1, p.27-38. 1993.

[15] Hauck. W.W.; Donner**,** A. Wald's Test as Applied to Hypotheses in Logit Analysis. **Journal of the American Statistical Association**, 1977. Vol. 72, 851-853.

[16] Heinze, G.; Schemper, M. A solution to the problem of separation in logistic regression. **Statistics in Medicine,** v.21, p.2409-2419. 2002.

[17] Hosmer, D.W.; Lemeshow, S. **Applied logistic regression**. New York: Jonhn Wiley, 1989. 307p.

[18] Jeffreys, H. An Invariant Form for the Prior Probability in Estimation Problems. **Proceedings of the Royal Society of London**. Series A, Mathematical and Physical Sciences , v.186, n. 1007, p. 453–461. 1946.

[19] Kleinbaum, D.G.; **Logistic regression: a self-learning text.** New York: Springer-Verlac, 1994. 278p.

[20] Kosmidis, I. and D. Firth (2008). Bias reduction in exponential family non-linear models. **Technical Report** 8-5, CRiSM working paper series, University of Warwick.<http://www2.warwick.ac.uk/fac/sci/statistics/crism/research/2008/paper08-05/08-05wv2.pdf>

[21] Nacle, D.P. **Estimadores de Máxima Verossimilhança em Modelos de Regressão Logística na Situação de Separação Quase-Completa.** Belo Horizonte: Dissertação de Mestrado (Departamento de Estatística) – UFMG. 2004

[22] Nelder, J.A.; Wedderburn, R.W.M. Generalized linear models. **Journal of the Royal Statistical Society**, London, v.135, p370-384, 1972.

[23] Paula, G.A. **Modelos de regressão com apoio computacional**. IME-USP, São Paulo, 2004, 253p.

[24] R Development Core Team 2009. **R: a language and environment for statistical computing**.Vienna, Austria: R Foundation for Statistical Computing. Disponível em: <http://www.R-project.org>. Acesso em: Nov. 2009.

[25] Santner, T.J; Duffy, D.E. A note on A. Albert and J.A. Anderson's conditions for the existence of maximum likelihood estimates in logistic regression models. **Biometrika**, v.73, n.3, p.755-758, 1986.

[26] SAS Institute Inc. *SAS 9.2.* **Help and Documentation**, Cary, NC: SAS® Institute Inc., 2009.

[27] Wald, A. Tests of Statistical Hypotheses concerning Several Parameters when the Number of Observations is Large. Trans. Amer. Math. Soc., 54, 426-482. 1943

[28] Zorn, C. A. Solution to Separation in Binary Response Models. **Political Analysis,** v.13, p.157-170, 2005.

APÊNDICE A

Tabela A1 – Proporção de casos de regressão simulados para beta0=-5 e diferentes tamanhos de amostra e valores beta1

2N=10	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,2154	0,0976	0,0664	0,0702	0,0622	0,0594	0,0730	0,0926	0,2192
Separação Completa	0,7834	0,9016	0,9328	0,9294	0,9368	0,9394	0,9258	0,9060	0,7798
Casos Comuns	0,0012	0,0008	0,0008	0,0004	0,0010	0,0012	0,0012	0,0014	0,0010
2N=30	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,5094	0,2568	0,1888	0,1706	0,1796	0,1774	0,1956	0,2588	0,5130
Separação Completa	0,4838	0,7342	0,8014	0,8194	0,8126	0,8140	0,7946	0,7330	0,4796
Casos Comuns	0,0068	0,0090	0,0098	0,0100	0,0078	0,0086	0,0098	0,0082	0,0074
2N=50	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,6930	0,3818	0,3034	0,2644	0,2648	0,2622	0,2942	0,3864	0,6916
Separação Completa	0,2906	0,5958	0,6754	0,7152	0,7118	0,7116	0,6794	0,5914	0,2954
Casos Comuns	0,0164	0,0224	0,0212	0,0204	0,0234	0,0262	0,0264	0,0222	0,0130
2N=100	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,8712	0,5848	0,4504	0,4134	0,4028	0,4114	0,4436	0,5766	0,8850
Separação Completa	0,0862	0,3456	0,4666	0,4962	0,5150	0,5052	0,4778	0,3538	0,0732
Casos Comuns	0,0426	0,0696	0,0830	0,0904	0,0822	0,0834	0,0786	0,0696	0,0418
2N=200	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,9082	0,6916	0,5546	0,4986	0,4986	0,5036	0,5562	0,6862	0,8982
Separação Completa	0,0048	0,1250	0,2282	0,2566	0,2532	0,2552	0,2270	0,1292	0,0092
Casos Comuns	0,0870	0,1834	0,2172	0,2448	0,2482	0,2412	0,2168	0,1846	0,0926
2N=400	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,8306	0,5994	0,4588	0,3912	0,3796	0,3864	0,4538	0,6044	0,8306
Separação Completa	0,0000	0,0182	0,0446	0,0690	0,0608	0,0668	0,0458	0,0142	0,0000
Casos Comuns	0,1694	0,3824	0,4966	0,5398	0,5596	0,5468	0,5004	0,3814	0,1694

Variando b0 e os mesmos valores de b1 e N na Função1 obtêm-se as Tabelas A2, A3, A4, A5, A6, A7 e A8.

Tabela A2 – Proporção de casos de regressão simulados para beta0=-4 e diferentes tamanhos de amostra e valores beta1

2N=10	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,4722	0,2286	0,1826	0,1650	0,1524	0,1568	0,1804	0,2396	0,4664
Separação Completa	0,5202	0,7632	0,8098	0,8268	0,8372	0,8360	0,8106	0,7548	0,5286
Casos Comuns	0,0076	0,0082	0,0076	0,0082	0,0104	0,0072	0,0090	0,0056	0,0050
2N=30	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,8288	0,5124	0,3908	0,3584	0,3624	0,3608	0,4028	0,5002	0,8212
Separação Completa	0,1424	0,4346	0,5612	0,5804	0,5794	0,5822	0,5346	0,4452	0,1458
Casos Comuns	0,0288	0,0530	0,0480	0,0612	0,0582	0,0570	0,0626	0,0546	0,0330
2N=50	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,9004	0,6390	0,5160	0,4608	0,4758	0,4660	0,5100	0,6352	0,9066
Separação Completa	0,0376	0,2536	0,3528	0,4078	0,3936	0,3968	0,3666	0,2602	0,0388
Casos Comuns	0,0620	0,1074	0,1312	0,1314	0,1306	0,1372	0,1234	0,1046	0,0546
2N=100	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,8884	0,6736	0,5348	0,4852	0,4948	0,4900	0,5472	0,6764	0,8854
Separação Completa	0,0010	0,0598	0,1322	0,1614	0,1582	0,1570	0,1346	0,0604	0,0010
Casos Comuns	0,1106	0,2666	0,3330	0,3534	0,3470	0,3530	0,3182	0,2632	0,1136
2N=200	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,7860	0,5206	0,3472	0,2756	0,2684	0,2710	0,3410	0,5080	0,7718
Separação Completa	0,0000	0,0018	0,0156	0,0286	0,0276	0,0280	0,0192	0,0028	0,0000
Casos Comuns	0,2140	0,4776	0,6372	0,6958	0,7040	0,7010	0,6398	0,4892	0,2282
2N=400	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,6088	0,2676	0,1120	0,0512	0,0540	0,0542	0,1176	0,2596	0,6076
Separação Completa	0,0000	0,0000	0,0002	0,0006	0,0008	0,0006	0,0000	0,0000	0,0000
Casos Comuns	0,3912	0,7324	0,8878	0,9482	0,9452	0,9452	0,8824	0,7404	0,3924

Tabela A3 – Proporção de casos de regressão simulados para beta0=-3 e diferentes tamanhos de amostra e valores beta1

2N=10	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,7682	0,4708	0,3614	0,3404	0,3450	0,3430	0,3728	0,4806	0,7768
Separação Completa	0,2066	0,4888	0,5936	0,6100	0,6082	0,6086	0,5800	0,4774	0,1982
Casos Comuns	0,0252	0,0404	0,0450	0,0496	0,0468	0,0484	0,0472	0,0420	0,0250
2N=30	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,8994	0,6882	0,5608	0,5038	0,4948	0,5000	0,5616	0,6864	0,8982
Separação Completa	0,0070	0,1170	0,1942	0,2274	0,2404	0,2392	0,1884	0,1132	0,0082
Casos Comuns	0,0936	0,1948	0,2450	0,2688	0,2648	0,2608	0,2500	0,2004	0,0936
2N=50	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,8572	0,6232	0,4838	0,4168	0,4156	0,4162	0,4780	0,6222	0,8414
Separação Completa	0,0000	0,0250	0,0596	0,0908	0,0808	0,0890	0,0626	0,0262	0,0012
Casos Comuns	0,1428	0,3518	0,4566	0,4924	0,5036	0,4948	0,4594	0,3516	0,1574
2N=100	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,7072	0,4074	0,2474	0,1728	0,1596	0,1700	0,2390	0,3958	0,7132
Separação Completa	0,0000	0,0008	0,0052	0,0080	0,0074	0,0088	0,0034	0,0006	0,0000
Casos Comuns	0,2928	0,5918	0,7474	0,8192	0,8330	0,8212	0,7576	0,6036	0,2868
2N=200	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,503	0,168	0,053	0,018	0,011	0,015	0,049	0,158	0,524
Separação Completa	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Casos Comuns	0,497	0,832	0,947	0,982	0,988	0,985	0,951	0,842	0,476
2N=400	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,2592	0,0278	0,0018	0,0004	0,0000	0,0002	0,0024	0,0248	0,2658
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Casos Comuns	0,7408	0,9722	0,9982	0,9996	1,0000	0,9998	0,9976	0,9752	0,7342

Tabela A4 – Proporção de casos de regressão simulados para beta0=-2 e diferentes tamanhos de amostra e valores beta1

2N=10	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,8906	0,6700	0,5560	0,4994	0,5010	0,4958	0,5516	0,6666	0,8892
Separação Completa	0,0304	0,1574	0,2440	0,2888	0,2820	0,2802	0,2454	0,1604	0,0314
Casos Comuns	0,0790	0,1726	0,2000	0,2118	0,2170	0,2240	0,2030	0,1730	0,0794
2N=30	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,7630	0,4724	0,3224	0,2654	0,2520	0,2552	0,3226	0,4776	0,7578
Separação Completa	0,0000	0,0054	0,0152	0,0232	0,0210	0,0184	0,0170	0,0048	0,0000
Casos Comuns	0,2370	0,5222	0,6624	0,7114	0,7270	0,7264	0,6604	0,5176	0,2422
2N=50	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,6364	0,2990	0,1526	0,0774	0,0752	0,0848	0,1420	0,2870	0,6306
Separação Completa	0,0000	0,0002	0,0020	0,0018	0,0012	0,0028	0,0012	0,0000	0,0000
Casos Comuns	0,3636	0,7008	0,8454	0,9208	0,9236	0,9124	0,8568	0,7130	0,3694
2N=100	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,4022	0,0840	0,0192	0,0034	0,0024	0,0040	0,0176	0,0872	0,4100
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0000	0,0000	0,0000
Casos Comuns	0,5978	0,9160	0,9808	0,9966	0,9976	0,9958	0,9824	0,9128	0,5900
2N=200	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,1666	0,0092	0,0006	0,0000	0,0000	0,0000	0,0004	0,0072	0,1680
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Casos Comuns	0,8334	0,9908	0,9994	1,0000	1,0000	1,0000	0,9996	0,9928	0,8320
2N=400	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,0266	0,0002	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0292
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Casos Comuns	0,9734	0,9998	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9708

Tabela A5 – Proporção de casos de regressão simulados para beta0=-1 e diferentes tamanhos de amostra e valores beta1

2N=10	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,8296	0,5410	0,4000	0,3398	0,3274	0,3428	0,3904	0,5502	0,8332
Separação Completa	0,0016	0,0148	0,0394	0,0396	0,0424	0,0414	0,0368	0,0168	0,0006
Casos Comuns	0,1688	0,4442	0,5606	0,6206	0,6302	0,6158	0,5728	0,4330	0,1662
2N=30	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,4870	0,1458	0,0490	0,0190	0,0224	0,0186	0,0500	0,1432	0,4800
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0002	0,0002	0,0000	0,0000	0,0000
Casos Comuns	0,5130	0,8542	0,9510	0,9810	0,9774	0,9812	0,9500	0,8568	0,5200
2N=50	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,3000	0,0412	0,0064	0,0012	0,0010	0,0010	0,0046	0,0376	0,2860
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Casos Comuns	0,7000	0,9588	0,9936	0,9988	0,9990	0,9990	0,9954	0,9624	0,7140
2N=100	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,0908	0,0012	0,0000	0,0000	0,0000	0,0000	0,0002	0,0018	0,0850
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Casos Comuns	0,9092	0,9988	1,0000	1,0000	1,0000	1,0000	0,9998	0,9982	0,9150
2N=200	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,0058	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0072
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Casos Comuns	0,9942	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9928
2N=400	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Casos Comuns	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Tabela A6 – Proporção de casos de regressão simulados para beta0=0 e diferentes tamanhos de amostra e valores beta1

2N=10	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,7786	0,3666	0,2094	0,1274	0,1172	0,1274	0,1872	0,3788	0,7734
Separação Completa	0,0000	0,0002	0,0022	0,0020	0,0014	0,0018	0,0014	0,0002	0,0000
Casos Comuns	0,2214	0,6332	0,7884	0,8706	0,8814	0,8708	0,8114	0,6210	0,2266
2N=30	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,2770	0,0168	0,0016	0,0000	0,0000	0,0000	0,0024	0,0184	0,2658
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Casos Comuns	0,7230	0,9832	0,9984	1,0000	1,0000	1,0000	0,9976	0,9816	0,7342
2N=50	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,0820	0,0012	0,0000	0,0000	0,0000	0,0000	0,0000	0,0008	0,0826
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Casos Comuns	0,9180	0,9988	1,0000	1,0000	1,0000	1,0000	1,0000	0,9992	0,9174
2N=100	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,0024	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0028
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Casos Comuns	0,9976	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9972
2N=200	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Casos Comuns	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
2N=400	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Casos Comuns	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Tabela A7 – Proporção de casos de regressão simulados para beta0=1 e diferentes tamanhos de amostra e valores beta1

2N=10	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,8152	0,5340	0,3878	0,3406	0,3294	0,3258	0,4028	0,5424	0,8332
Separação Completa	0,0010	0,0174	0,0344	0,0410	0,0470	0,0428	0,0344	0,0176	0,0004
Casos Comuns	0,1838	0,4486	0,5778	0,6184	0,6236	0,6314	0,5628	0,4400	0,1664
2N=30	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,4682	0,1550	0,0494	0,0174	0,0168	0,0202	0,0536	0,1450	0,4892
Separação Completa	0,0000	0,0000	0,0002	0,0000	0,0002	0,0000	0,0000	0,0000	0,0000
Casos Comuns	0,5318	0,8450	0,9504	0,9826	0,9830	0,9798	0,9464	0,8550	0,5108
2N=50	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,2898	0,0440	0,0066	0,0008	0,0006	0,0002	0,0078	0,0394	0,2968
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Casos Comuns	0,7102	0,9560	0,9934	0,9992	0,9994	0,9998	0,9922	0,9606	0,7032
2N=100	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,0824	0,0010	0,0000	0,0000	0,0000	0,0000	0,0000	0,0016	0,0852
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Casos Comuns	0,9176	0,9990	1,0000	1,0000	1,0000	1,0000	1,0000	0,9984	0,9148
2N=200	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,0058	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0056
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Casos Comuns	0,9942	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9944
2N=400	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,0002	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Casos Comuns	0,9998	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Tabela A8 – Proporção de casos de regressão simulados para beta0=2 e diferentes tamanhos de amostra e valores beta1

2N=10	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,8958	0,6656	0,5356	0,5102	0,5018	0,5046	0,5486	0,6654	0,8928
Separação Completa	0,0304	0,1570	0,2532	0,2692	0,2802	0,2806	0,2494	0,1612	0,0276
Casos Comuns	0,0738	0,1774	0,2112	0,2206	0,2180	0,2148	0,2020	0,1734	0,0796
2N=30	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,7514	0,4822	0,3350	0,2598	0,2498	0,2586	0,3244	0,4876	0,7652
Separação Completa	0,0000	0,0034	0,0168	0,0210	0,0238	0,0250	0,0118	0,0032	0,0000
Casos Comuns	0,2486	0,5144	0,6482	0,7192	0,7264	0,7164	0,6638	0,5092	0,2348
2N=50	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,6448	0,3052	0,1430	0,0766	0,0800	0,0792	0,1450	0,2924	0,6370
Separação Completa	0,0000	0,0002	0,0004	0,0010	0,0012	0,0016	0,0008	0,0004	0,0000
Casos Comuns	0,3552	0,6946	0,8566	0,9224	0,9188	0,9192	0,8542	0,7072	0,3630
2N=100	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,3972	0,0846	0,0178	0,0034	0,0028	0,0044	0,0182	0,0830	0,4072
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Casos Comuns	0,6028	0,9154	0,9822	0,9966	0,9972	0,9956	0,9818	0,9170	0,5928
2N=200	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,1622	0,0088	0,0000	0,0000	0,0000	0,0000	0,0004	0,0072	0,1682
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Casos Comuns	0,8378	0,9912	1,0000	1,0000	1,0000	1,0000	0,9996	0,9928	0,8318
2N=400	beta1=-2	beta1=-1	beta1=-0,5	beta1=-0,1	beta1=0	beta1=0,1	beta1=0,5	beta1=1	beta1=2
Separação Quase Completa	0,0210	0,0002	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0296
Separação Completa	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Casos Comuns	0,9790	0,9998	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9704

APÊNDICE B

Códigos de programação no software R e gráficos de resultados

```
##############################################################################
                                    Função1
Função para contagem da proporção de cada configuração de dados logísticos gerados pela
simulação com n=2000 (Tabela A1)
##############################################################################
```

```
##############################################################################
Foram gerados dados sob separabilidade e superposição (casos comuns, overlap)
simultaneamente em função de n, N, b0 e b1, onde η=2N é o tamanho da amostra e n é o
número de repetições do experimento, b0 e b1 são os parâmetros do modelo logístico. A
Função 1 retorna a proporção de cada configuração de dados logísticos simulados em
função dos parâmetros n, N, b0, b1
##############################################################################
```

```
options(digits=10)
simula0<-function(n,b1,b0,N)
{
q1<-q2<-q3<-0
for(i in 1:n)
  {
x=c(rep(1,N),rep(-1,N))
px0<-cbind((exp(b0+b1*x))/(1+exp(b0+b1*x)))
simula0<-rbinom(2*N,1,rep(c(px0[1,],px0[(N+1),]),each=N))
simula=simula0[1:N]
simula1=simula0[(N+1):(2*N)]
sim<-c(sum(simula),sum(simula1))
não<-c(N-sum(simula),N-sum(simula1))
v<-c(1,-1)
tab<-cbind(v,sim,não)
if((tab[1,2]!=0&tab[2,2]!=0&tab[1,3]!=0&tab[2,3]!=0)){
q3<-q3+1
}else if(sum(tab %in% 0)==1){
q1<-q1+1
}else{
q2<-q2+1
}
}
return(rbind(Separação_Quase_Completa=q1/n,Separação_Completa=q2/n,Casos_Comuns
=q3/n))}
```

```
##############################################################################
Tabela de proporção para N=5 com b1 variando, b1=(-2,-1,-0,5, -0.1,0,0.1,0.5,1,2) e b0=-5
##############################################################################

b0<-(-5)

a1<-simula0(5000,-2,b0,5)
a2<-simula0(5000,-1,b0,5)
a22<-simula0(5000,-0.5,b0,5)
a222<-simula0(5000,-0.1,b0,5)
a3<-simula0(5000,0,b0,5)
a333<-simula0(5000,0.1,b0,5)
a24<-simula0(5000,0.5,b0,5)
a4<-simula0(5000,1,b0,5)
a5<-simula0(5000,2,b0,5)

dad1<-cbind(a1,a2,a22,a222,a3,a333,a24,a4,a5)
colnames(dad1)<-c("beta1=-2","beta1=-1","beta1=-0.5","beta1=-
0.1","beta1=0","beta1=0.1","beta1=0.5","beta1=1","beta1=2")
dad1

##############################################################################
Para outros tamanhos de amostras  mostrados na Tabela A1 basta variar os valores de N=
(15, 25, 50, 100,200) em a1, a2, a22, a222, a3, a333, a24, a4, a5
##############################################################################

##############################################################################
Para gerar as Tabelas A2, A3, A4, A5, A6, A7 e A8 basta variar o valor de b0 e repetir o
processo de geração da Tabela A1
##############################################################################
```

VARIAÇÃO DO TAMANHO AMOSTRAL DADO QUE Ho=0 (b1=0), PARA AVALIAR O COMPORTAMENTO ASINTÓTICO (TAMANHO DESCRITIVO OU EMPÍRICO) DO TESTE DA RAZÃO DE VEROSSIMILHANÇAS. Figura 3.9 (a)

```
##############################################################################
                                    Função2
Função retorna o vetor de TRVs genuíno de um total de n simulações
##############################################################################

##############################################################################
São gerados dados sob configuração de separabilidade e superposição (casos comuns,
overlap) simultaneamente, em que η = 2N é o tamanho da amostra n é o número de
repetições do experimento e b0, b1 são os parâmetros do modelo logístico. Função 2
retorna proporção de rejeição de Ho quando são simulados dados sob Ho=0 (b1=0) com
variação de N e b0
##############################################################################
```

```
simula1<-function(n,b1,b0,N)
{
d<-numeric()
for(i in 1:n)
  {
x=c(rep(1,N),rep(-1,N))
px0<-cbind((exp(b0+b1*x))/(1+exp(b0+b1*x)))
simula0<-rbinom(2*N,1,rep(c(px0[1,],px0[(N+1),]),each=N))
simula=simula0[1:N]
simula1=simula0[(N+1):(2*N)]
sim<-c(sum(simula),sum(simula1))
não<-c(N-sum(simula),N-sum(simula1))
v<-c(1,-1)
tab<-cbind(v,sim,não)
ajust2<-glm(tab[,c(2,3)]~v,family=binomial)
d[i]<-ajust2$null.deviance-ajust2$deviance
}
return(d)
}
```

##

Função 3

Função para avaliar o comportamento assintótico (Tamanho descritivo ou empírico) do TRV. Variação do tamanho amostral η=2N entre [4,3000] sob Ho=0 (b1=0) e variação de b1= (-5, -4, -2, -3, -2, -1, 0, 1,2) com b0=-5. A Função 3 retorna a proporção de TRVs maiores que 3.84 (valor da estatística qui-quadrado com 1 grau de liberdade a 5% de probabilidade), ou seja, retorna a proporção de rejeição de Ho.

##

```
b0<-(-5)
n<-2000
N<-seq(2,1500,10)
{
p1<-NULL
for(i in 1:length(N))
p1[i]<-sum(simula1(n,0,b0,N[i])>=3.84)/n
}
```

##

Para obter os vetores p2, p3, p4, p5, p6, p7 e p8 é necessário variar apenas os valores b0=(-4,-3,-2,-1,0,1,2) na Função3

##

PODER DO TESTE DA RAZÃO DE VEROSSIMILHANÇAS

```
##############################################################################
```

Função 4

São gerados dados sob separabilidade e superposição (casos comuns, *overlap*) simultaneamente

```
##############################################################################
```

```
##############################################################################
```

Em que η=2N é o tamanho da amostra n número de repetições de regressões logísticas, b0 e b1 são os parâmetros do modelo logístico. Função 4 retorna o vetor de TRVs genuíno.

```
##############################################################################
```

```
simula2<-function(n,b1,b0,N)
{
d<-numeric()
for(i in 1:n)
  {
x=c(rep(1,N),rep(-1,N))
px0<-cbind((exp(b0+b1*x))/(1+exp(b0+b1*x)))
simula0<-rbinom(2*N,1,rep(c(px0[1,],px0[(N+1),]),each=N))
simula=simula0[1:N]
simula1=simula0[(N+1):(2*N)]
sim<-c(sum(simula),sum(simula1))
não<-c(N-sum(simula),N-sum(simula1))
v<-c(1,-1)
tab<-cbind(v,sim,não)
ajust2<-glm(tab[,c(2,3)]~v,family=binomial)
d[i]<-ajust2$null.deviance-ajust2$deviance
}
return(d)
}
```

```
##############################################################################
```

Função 5

Variação de b1=[-7,7], N=5 e b0=-5, onde a Função5 retorna a proporção de rejeição de Ho.

```
##############################################################################
```

```
b0<-(-5)
N<-5
n<-2000
b1<-seq(-7,7,by=0.1)
{
p1<-NULL
for(i in 1:length(b1))
p1[i]<-sum(simula2(n,b1[i],b0,N)>=3.84)/n
}
```

##
Para obtenção dos vetores p2, p3, p4, p5, p6, p7 e p8 basta variar b0= (-4, -3, -2, -1, 0, 1,2) na Função5
##
Gráfico poder do TRV para tamanho de amostra η=2N=10
##

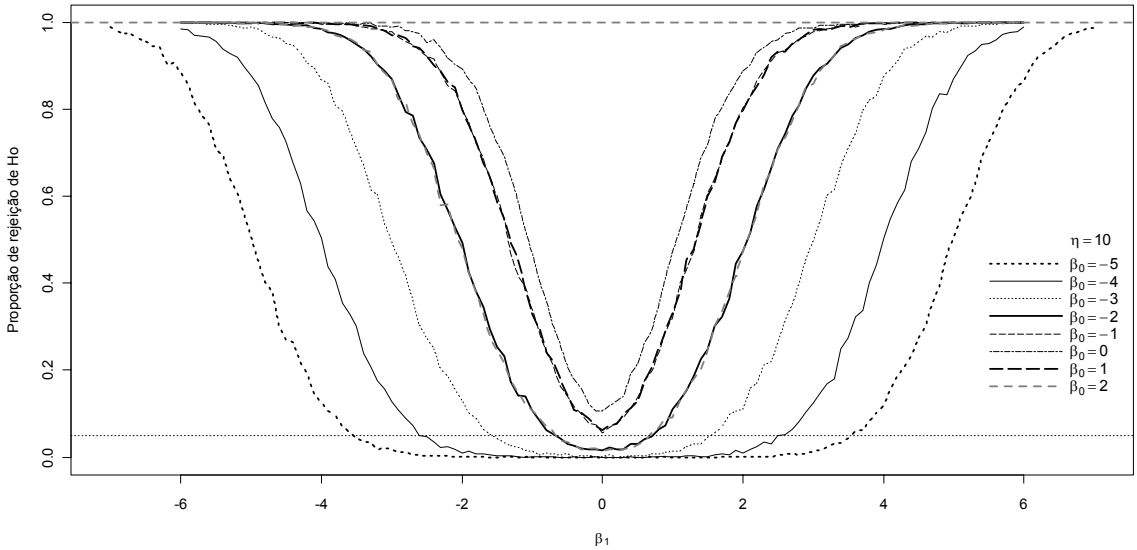


Figura A1 – Poder do teste da estatística TRV para o tamanho amostral η=10 e diferentes valores dos parâmetros $\beta_0 e \beta_1$.

##
Para construção das Figuras A2, A3, A4, A5, A6, basta variar N= (15, 25, 50, 100,200) para cada b0= (-4, -3, -2, -1, 0, 1,2) na Função5
##

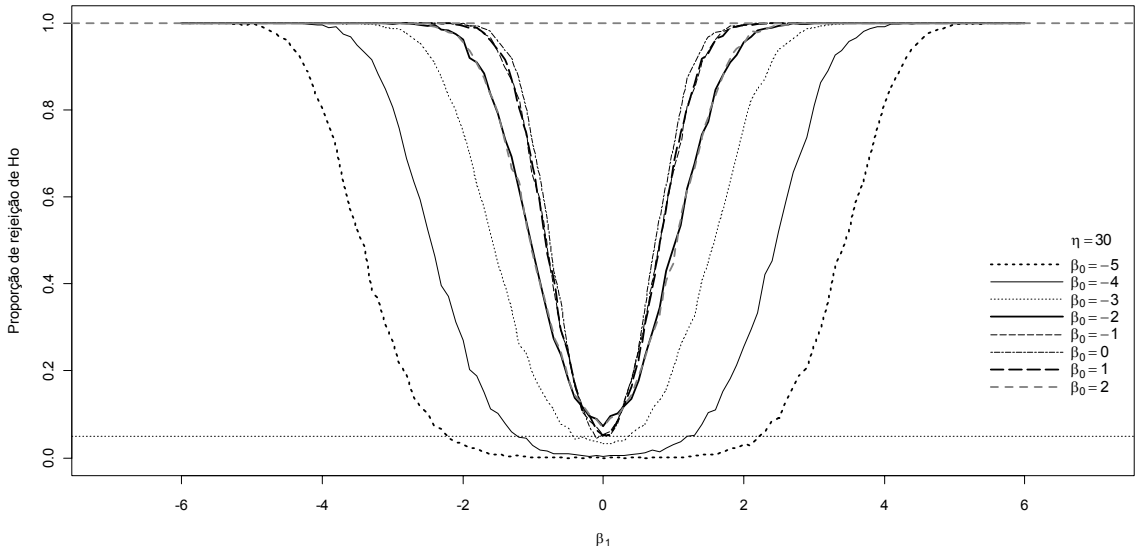


Figura A2 – Poder do teste da estatística TRV para o tamanho amostral η=30 e diferentes valores dos parâmetros $\beta_0 e \beta_1$.

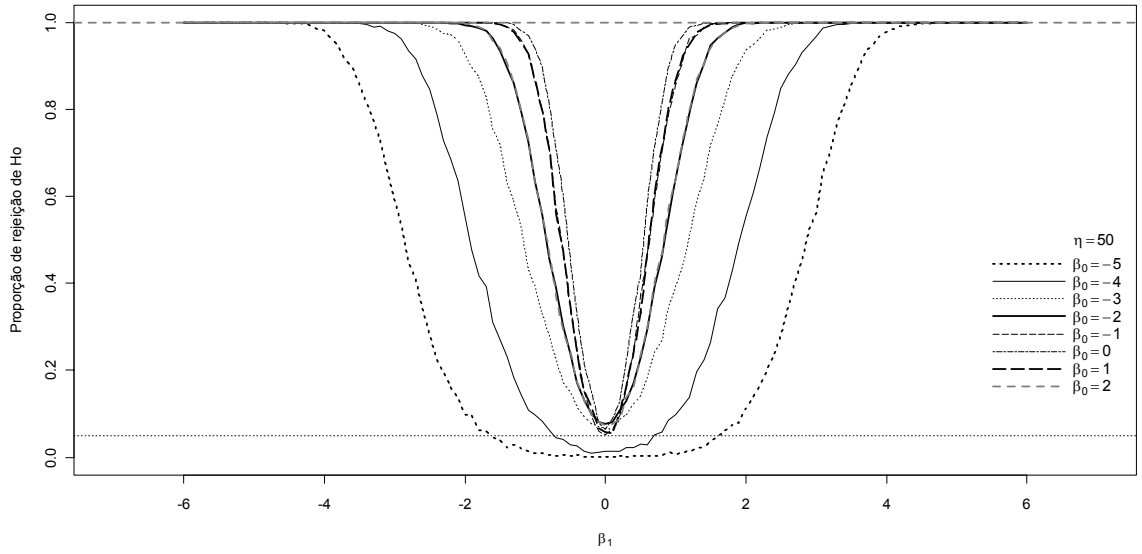


Figura A3 – Poder do teste da estatística TRV para o tamanho amostral η=50 e diferentes valores dos parâmetros $\beta_0 e \beta_1$.

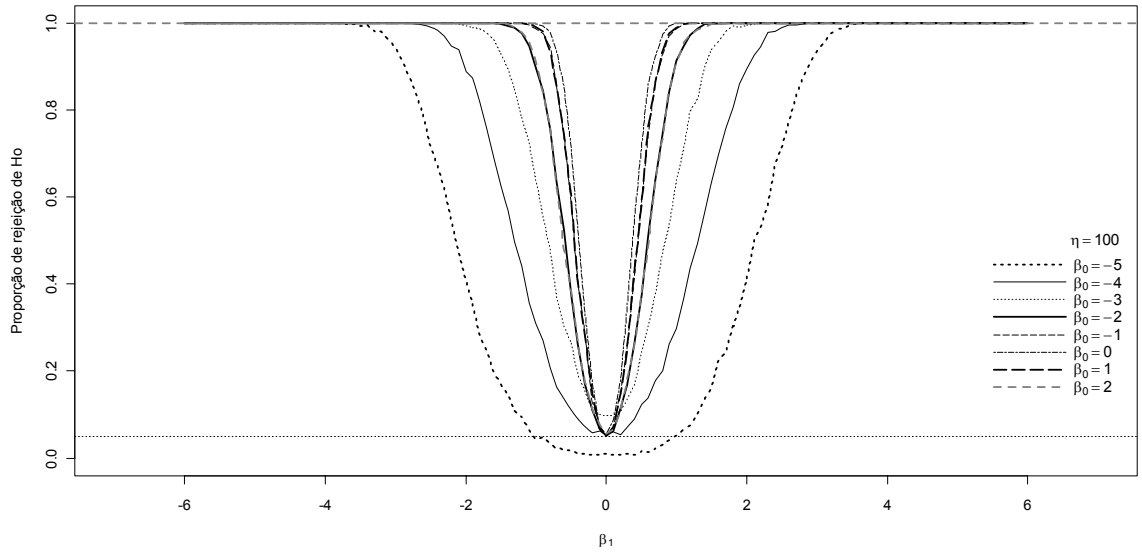


Figura A4 – Poder do teste da estatística TRV para o tamanho amostral η=100 e diferentes valores dos parâmetros $\beta_0 e \beta_1$.

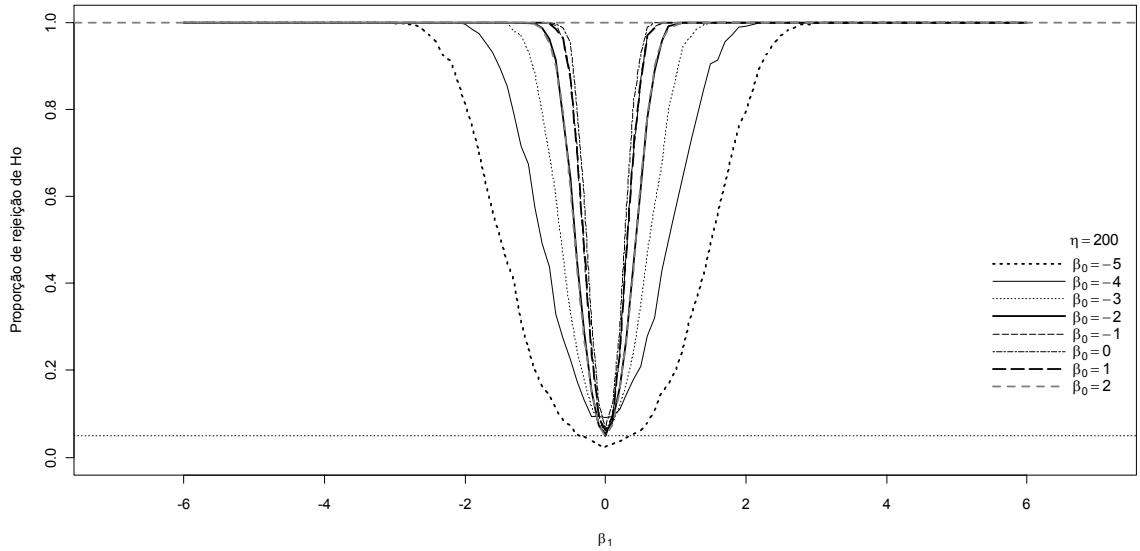


Figura A5 – Poder do teste da estatística TRV para o tamanho amostral n=200 e diferentes valores dos parâmetros $\beta_0 \, e \, \beta_1$.

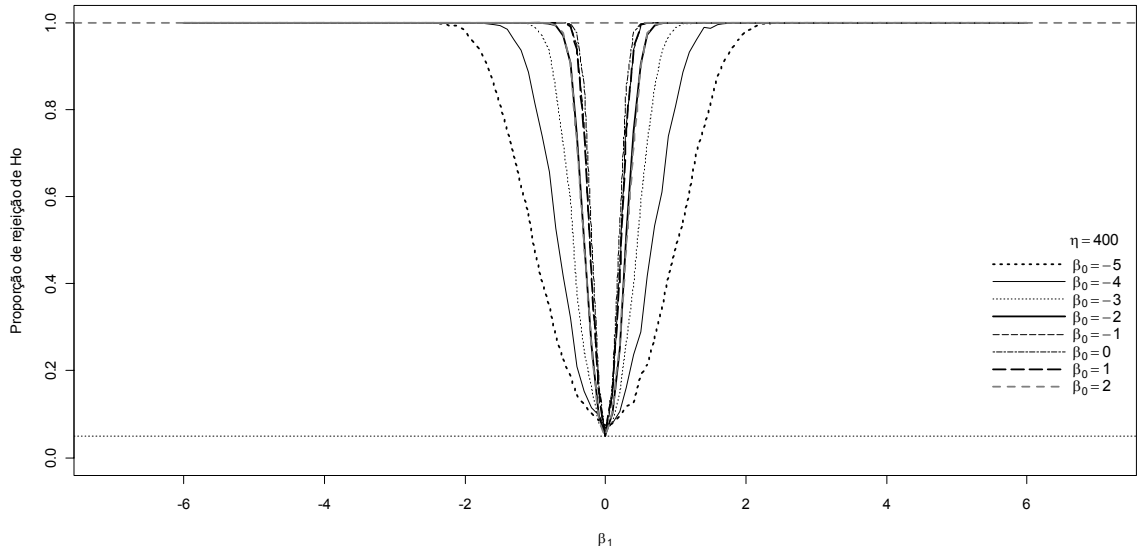


Figura A6 – Poder do teste da estatística TRV para o tamanho amostral η=400 e diferentes valores dos parâmetros $\beta_0 \, e \, \beta_1$.

```
##############################################################################
```
VARIAÇÃO DO TAMANHO AMOSTRAL SOB H_0 PARA AVALIAR O COMPORTAMENTO ASINTÓTICO (TAMANHO DESCRITIVO) DO TESTE DE WALD PENALIZADO. Figura 3.9 (b)
```
##############################################################################
```
Função6

Função para avaliação do comportamento assintótico (tamanho descritivo) do teste de wald penalizado. São gerados dados sob configuração de separabilidade e superposição (casos comuns, *overlap*) simultaneamente em que η=2N tamanho da amostra e n é o número de repetição do experimento, b0 e b1 são os parâmetros do modelo logístico. Função 6 retorna o vetor da estatística de Wald penalizada elevada ao quadrado.
```
##############################################################################

require(brglm)
simula3<-function(n,b1,b0,N)
{
d<-numeric()
for(i in 1:n)
  {
x=c(rep(1,N),rep(-1,N))
px0<-cbind((exp(b0+b1*x))/(1+exp(b0+b1*x)))
simula0<-rbinom(2*N,1,rep(c(px0[1,],px0[(N+1),]),each=N))
simula=simula0[1:N]
simula1=simula0[(N+1):(2*N)]
sim<-c(sum(simula),sum(simula1))
não<-c(N-sum(simula),N-sum(simula1))
v<-c(1,-1)
tab<-cbind(v,sim,não)
ajust2<-brglm(tab[,c(2,3)]~v,family=binomial,method="brglm.fit")
f<-solve(ajust2$FisherInfo)
d[i]<-ajust2$coefficients[2]^2/f[2,2]
}
return(d)
}
##############################################################################
```
Função 7

Variação do tamanho amostral sob Ho (b1=0), b1= (-5, -4, -3, -2, -1, 0, 1,2) e variando 2N entre [4,3000] com b0=-5
```
##############################################################################
n<-2000
b0<-(-5)
N<-seq(2,1500,10)
{
p1<-NULL
for(i in 1:length(N))
p1[i]<-sum(simula3(n,0,b0,N[i])>=3.84)/n
}
##############################################################################
```
Para obter os vetores p2, p3, p4, p5, p6, p7, p8 é necessário variar b0 na Função 7
```
##############################################################################
```

PODER DO TESTE WALD PENALIZADO

```
##############################################################################
```

Função8

Função para avaliação do teste de wald penalizado. São gerados dados sob configuração de separabilidade e superposição (casos comuns, *overlap*) simultaneamente, em que η=2N tamanho da amostra e n é repetição do experimento b0 e b1 parâmetros do modelo logístico A Função 8 retorna o vetor da estatística de Wald penalizada elevada ao quadrado.

```
##############################################################################
require(brglm)
simula4<-function(n,b1,b0,N)
{
d<-numeric()
for(i in 1:n)
  {
x=c(rep(1,N),rep(-1,N))
px0<-cbind((exp(b0+b1*x))/(1+exp(b0+b1*x)))
simula0<-rbinom(2*N,1,rep(c(px0[1,],px0[(N+1),]),each=N))
simula=simula0[1:N]
simula1=simula0[(N+1):(2*N)]
sim<-c(sum(simula),sum(simula1))
não<-c(N-sum(simula),N-sum(simula1))
v<-c(1,-1)
tab<-cbind(v,sim,não)
ajust2<-brglm(tab[,c(2,3)]~v,family=binomial,method="brglm.fit")
f<-solve(ajust2$FisherInfo)
d[i]<-ajust2$coefficients[2]^2/f[2,2]
}
return(d)
}

##############################################################################
```

Função 9

Gráfico poder do teste Wald penalizado fixando η=2N=10 e b0=-5

```
##############################################################################

n<-2000
b0<-(-5)
b1<-seq(-10,10,by=0.1)
{
p1<-NULL
for(i in 1:length(b1))
p1[i]<-sum(simula4(n,b1[i],b0,5)>=3.84)/n
}

##############################################################################
```

Para obter os vetores de p2, p3, p4, p5, p6, p7 e p8 é necessário variar b0 na Função 9

```
##############################################################################
```

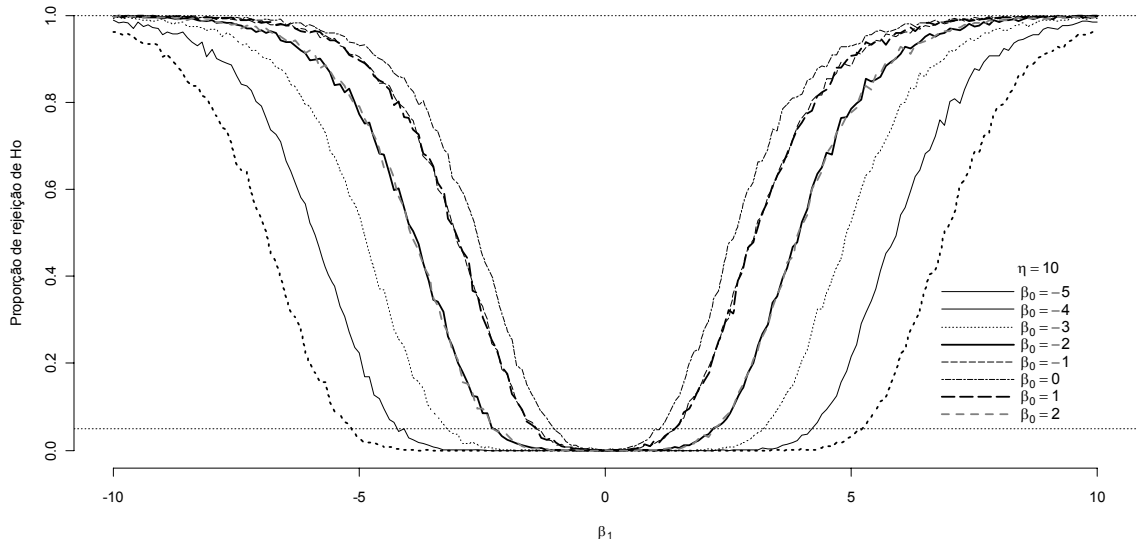


Figura A7 – Poder do teste da estatística de Wald penalizada para o tamanho amostral η=10 e diferentes valores dos parâmetros $\beta_0 \, e \, \beta_1$.

##
Para construção das Figuras A8, A9, A10, A11, A12 é necessário variar η=2N na Função 9
##

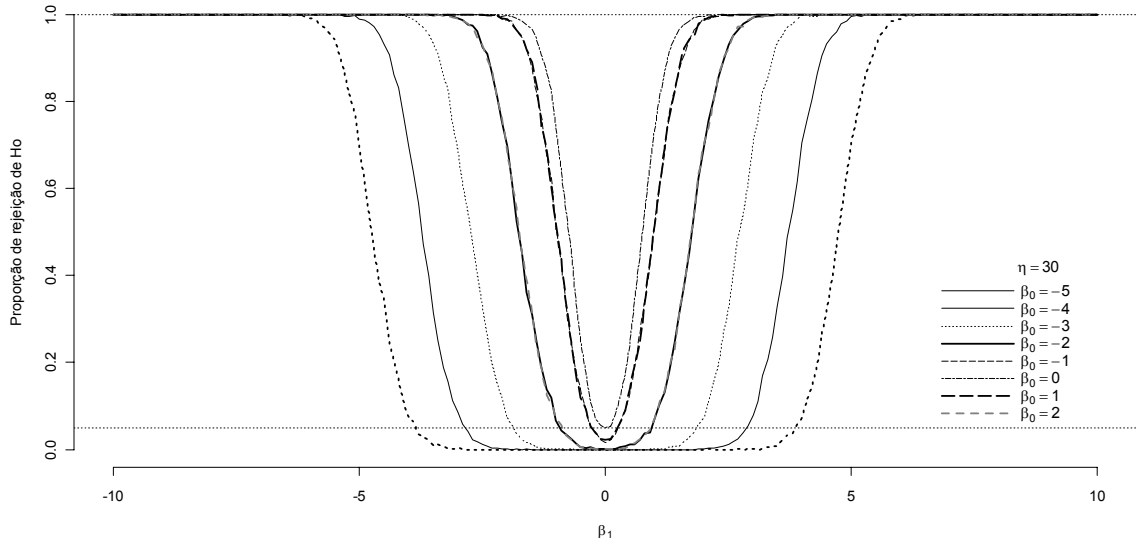


Figura A8 – Poder do teste da estatística de Wald penalizada para o tamanho amostral η=30 e diferentes valores dos parâmetros $\beta_0 \, e \, \beta_1$.

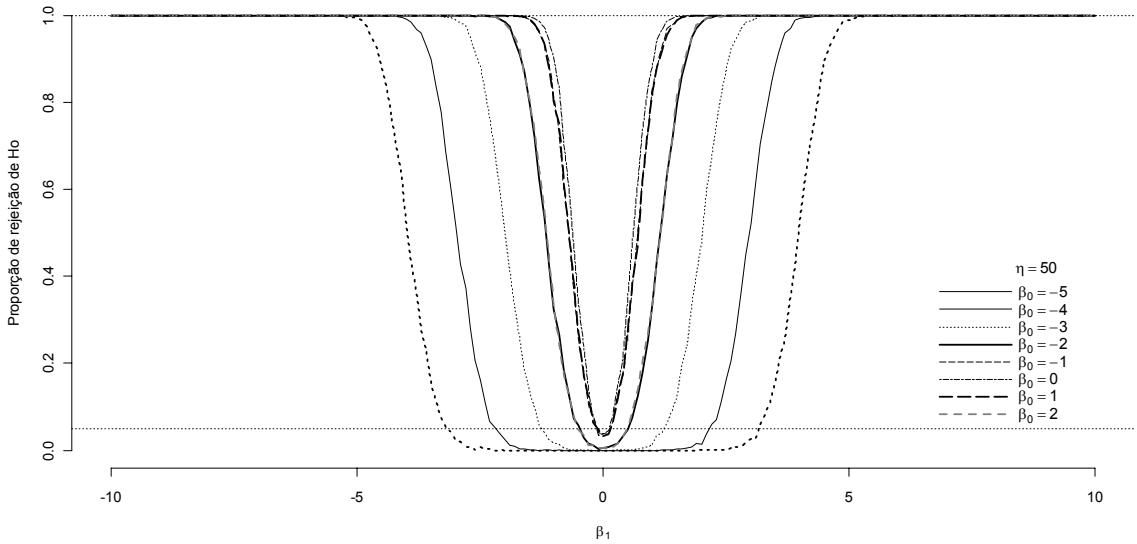


Figura A9 – Poder do teste da estatística de Wald penalizada para o tamanho amostral η=50 e diferentes valores dos parâmetros $\beta_0 \, e \, \beta_1$.

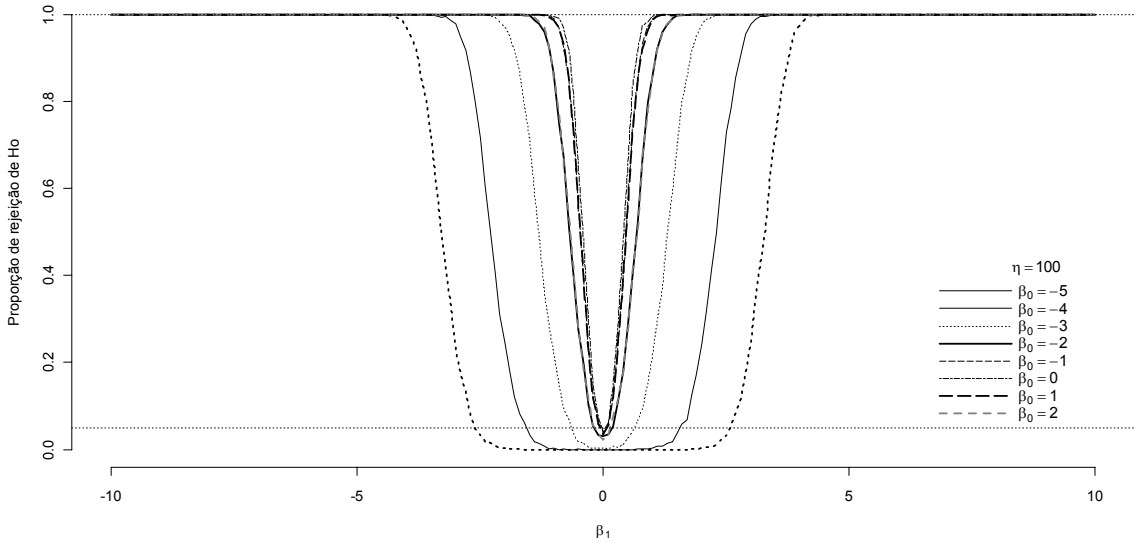


Figura A10 – Poder do teste da estatística de Wald penalizada para o tamanho amostral η=100 e diferentes valores dos parâmetros $\beta_0 \, e \, \beta_1$.

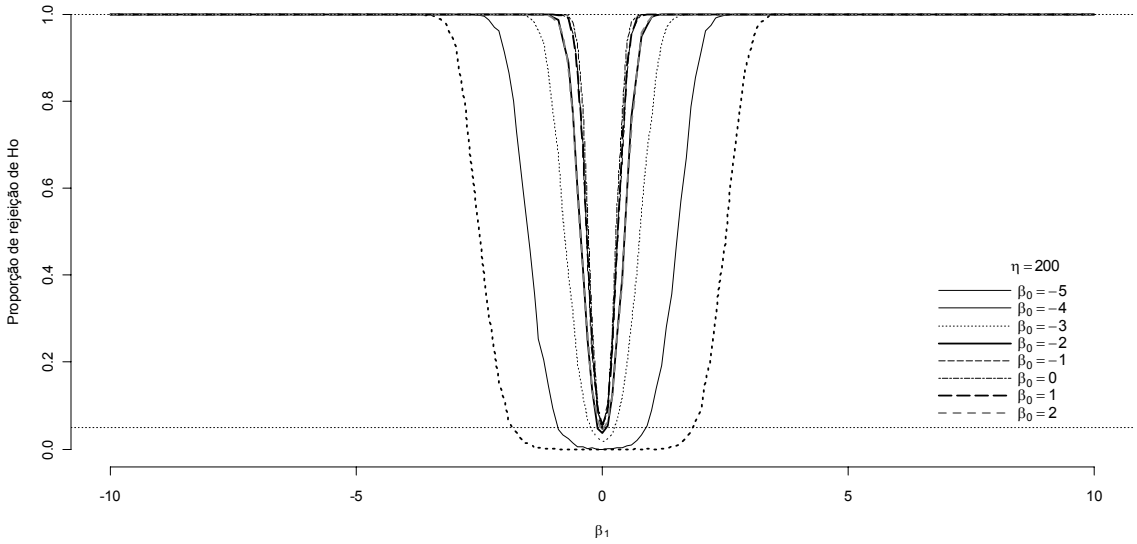


Figura A11 – Poder do teste da estatística de Wald penalizada para o tamanho amostral η=200 e diferentes valores dos parâmetros $\beta_0 \, e \, \beta_1$.

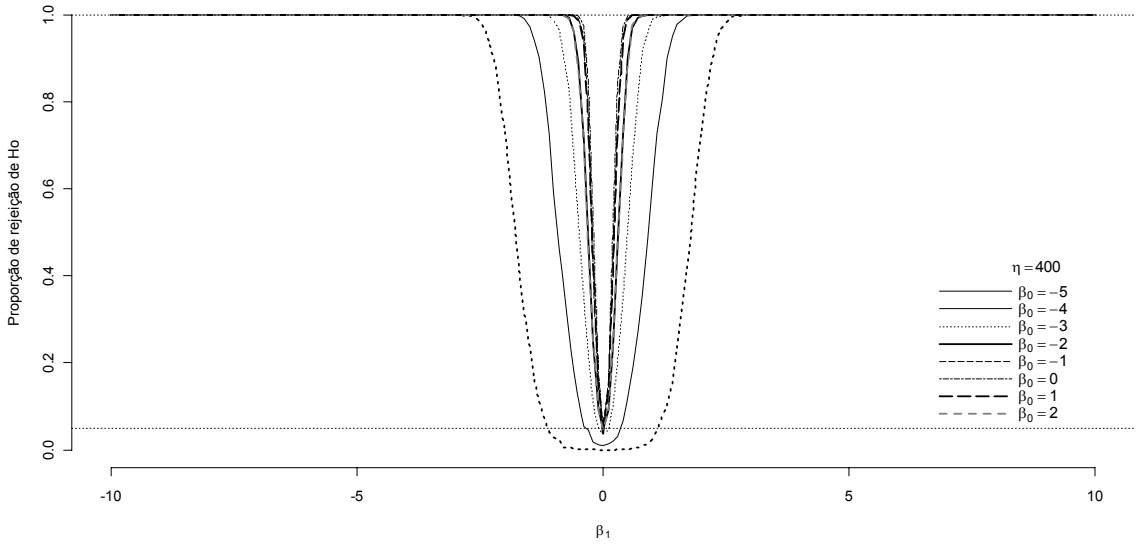
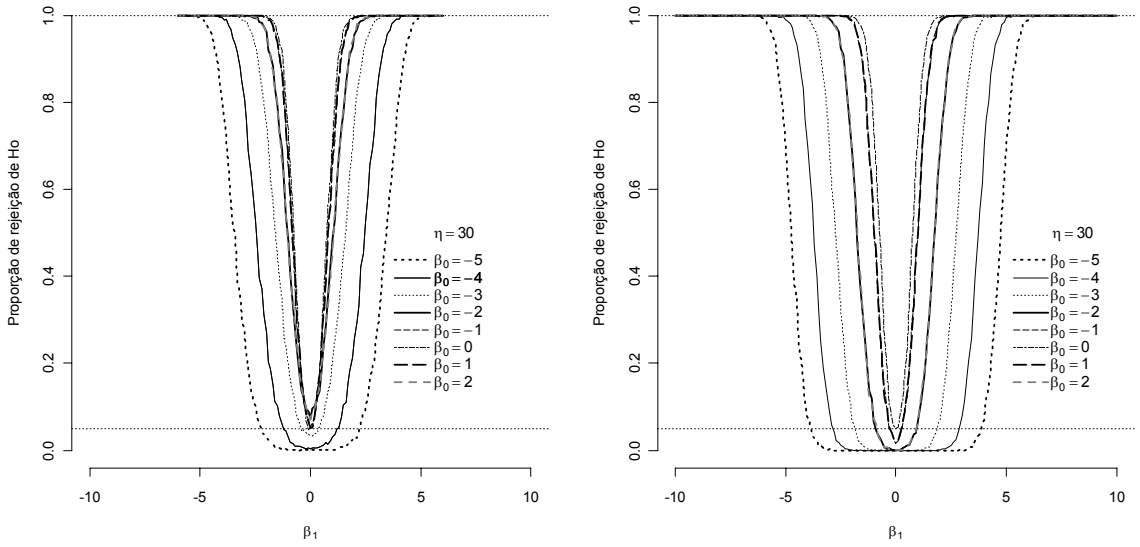


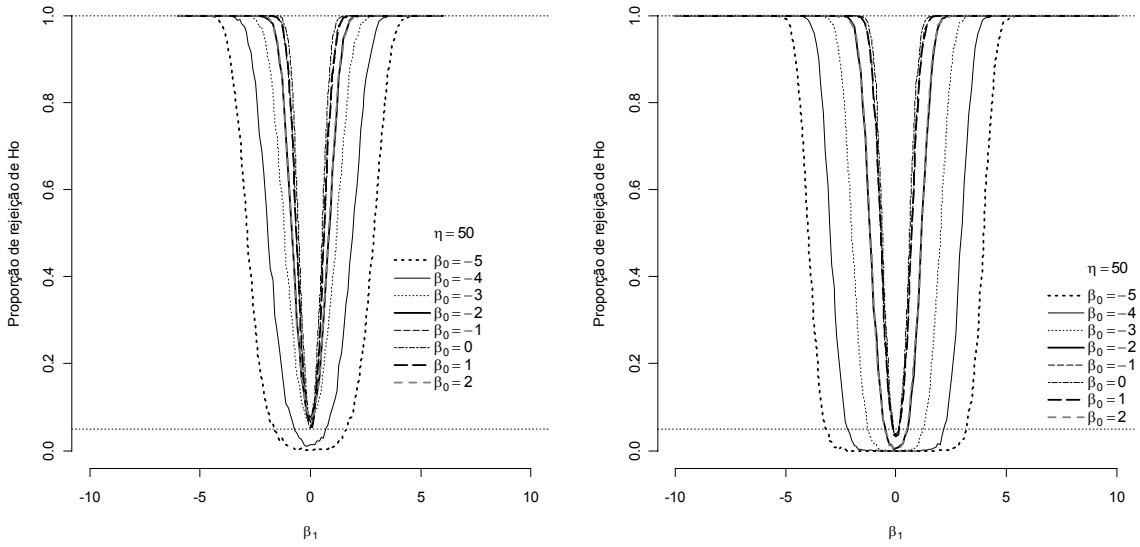
Figura A12 – Poder do teste da estatística de Wald penalizada para o tamanho amostral η=400 e diferentes valores dos parâmetros $\beta_0 \, e \, \beta_1$.

As Figuras A13 a A16 mostram, para tamanhos de amostras diferentes e numa mesma escala, a comparação entre o poder dos dois testes em questão.



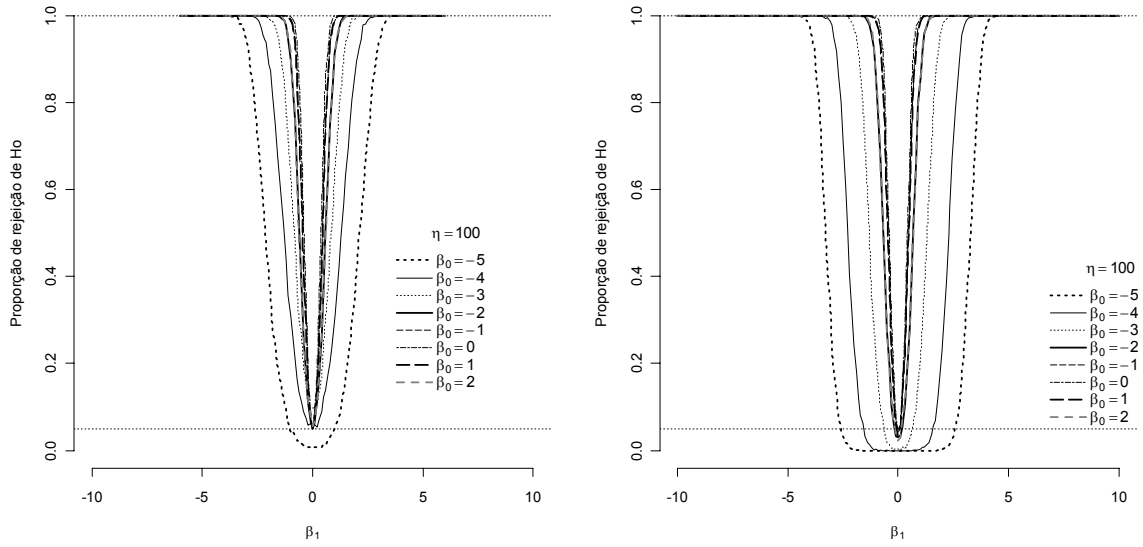
(a) (b)

Figura 13 – Função poder empírica dos testes da razão de verossimilhanças(TRV) (a) e de Wald (b) para amostras de tamanho η=30.



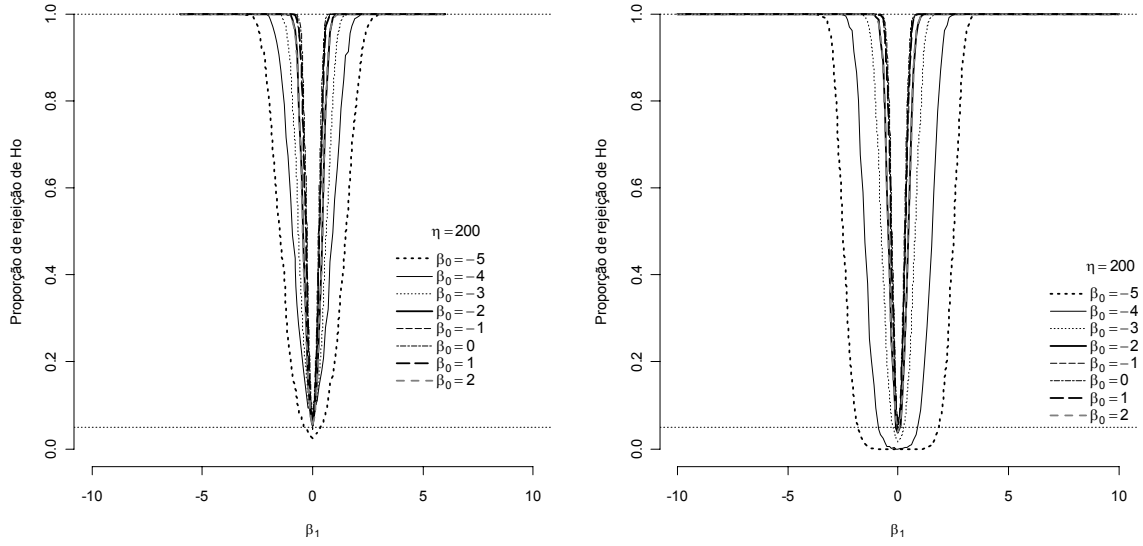
(a) (b)

Figura A14 – Função poder empírica dos testes da razão de verossimilhanças(TRV) (a) e de Wald (b) para amostras de tamanho η=50.



(a) (b)

Figura 15 – Função poder empírica dos testes da razão de verossimilhanças(TRV) (a) e de Wald (b) para amostras de tamanho η=100.



(a) (b)

Figura 16 – Função poder empírica dos testes da razão de verossimilhanças(TRV) (a) e de Wald (b) para amostras de tamanho η=200.

CÓDIGOS R PARA ANÁLISE DOS DADOS DAS APLICAÇÕES

 # Dados de craniotomia #

```
options(scipen=7)
require(RODBC)
a<-odbcConnectExcel("cran.xls")
dados1<-sqlFetch(a,"Plan1")
attach(dados1)
dados1
```

 # Modelos ajustados pelo método de máxima verossimilhança genuína #

```
fit1<-glm(cbind(N1,nao)~1,family=binomial)
summary(fit1)
fit2<-glm(cbind(N1,nao)~X1,family=binomial)
summary(fit2)
fit3<-glm(cbind(N1,nao)~X2,family=binomial)
summary(fit3)
fit4<-glm(cbind(N1,nao)~X1+X2,family=binomial)
summary(fit4)
```

 # Modelos ajustados pelo método de máxima verossimilhança penalizada #

```
require(brglm)
require(aod)
fit5<-brglm(cbind(N1,nao)~X1+X2,family=binomial)
summary(fit5)
```

 # Teste Wald #

```
wald.test(b=coef(fit5),Sigma=vcov(fit5),Terms= 2:2)
wald.test(b=coef(fit5),Sigma=vcov(fit5),Terms= 3:3)
wald.test(b=coef(fit5),Sigma=vcov(fit5),Terms= 2:3)
```

```
##########################################################################
```

 # Dados de *Adenanthera pavonina* L #

```
options(scipen=7)
require(RODBC)
a<-odbcConnectExcel("aden.xls")
dados1<-sqlFetch(a,"Plan4")
attach(dados1)
str(dados1)
```

 # Testando a interação pelo TRV #

```
fit1<-glm(cbind(sim,nao)~X1+X2+X1*X2,family=binomial,data=dados1)
summary(fit1)
fit2<-glm(cbind(sim,nao)~X1+X2,family=binomial,data=dados1)
summary(fit2)
```

 # Modelos ajustados pelo método de máxima verossimilhança genuína #

```
fit3<-glm(cbind(sim,nao)~X1,family=binomial,data=dados1)
summary(fit3)
fit4<-glm(cbind(sim,nao)~X2,family=binomial,data=dados1)
summary(fit4)
fit5<-glm(cbind(sim,nao)~X1+X2,family=binomial,data=dados1)
summary(fit5)
```

 # Testando a interação pelo teste Wald #

```
require(brglm)
require(aod)
fit6<-brglm(cbind(sim,nao)~X1*X2,family=binomial,data=dados1)
wald.test(b=coef(fit6),Sigma=vcov(fit6),Terms= 6:8)
```

 # Modelos ajustados pelo método de máxima verossimilhança penalizada #

```
fit7<-brglm(cbind(sim,nao)~X1+X2,family=binomial,data=dados1)
```

 # Teste Wald #

```
wald.test(b=coef(fit7),Sigma=vcov(fit7),Terms= 2:2)
wald.test(b=coef(fit7),Sigma=vcov(fit7),Terms= 3:5)
wald.test(b=coef(fit7),Sigma=vcov(fit7),Terms= 2:5)
```