UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS - ICEx
DEPARTAMENTO DE ESTATÍSTICA

MODELOS DE REGRESSÃO LOGÍSTICA PARA DELINEAMETENTOS HIERÁRQUICOS DESBALANCEADOS

ALUNA: RAQUEL YURI DA SILVEIRA AOKI
ORIENTADOR: FÁBIO NOGUEIRA DEMARQUI
RELATÓRIO TÉCNICO – MONOGRAFIA
DEZEMBRO DE 2014

RAQUEL YURI DA SILVEIRA AOKI

MODELOS DE REGRESSÃO LOGÍSTICA PARA DELINEAMETENTOS HIERÁRQUICOS DESBALANCEADOS

Relatório técnico, apresentado a Universidade Federal de Minas Gerais, como parte das exigências para a obtenção do título de bacharel em Estatística.

Orientador: Prof. Dr. Fabio Nogueira Demarqui.

Belo Horizonte - MG
2014

Monografia apresentada como requisito necessário para obtenção de título em Bacharel em Estatística. Qualquer citação atenderá as normas da ética científica.

NOME DO ALUNO

Monografia apresentada em ____/____/____

_____

Orientador Prof. Doutor Fabio Nogueira Demarqui

_____

1ª Examinadora Prof.ª Doutora Magda Carvalho Pires

_____

2ª Examinador Prof. Doutor Gregorio Saravia Atuncar

_____

Coordenadora Prof.ª Doutora Edna Afonso Reis

Dedico esse trabalho a minha mãe Elza, minha tia Terezinha e minha avó Rita (*in memorian). Desejo ter sido merecedora do esforço dedicado por vocês.*

AGRADECIMENTOS

Agradeço a minha família por sempre ter me incentivado a estudar e me apoiado ao longo desse caminho percorrido e por compreender minha ausência em certos momentos.

Agradeço aos professores do ensino fundamental por terem me dado as bases necessárias; aos professores de matemática do ensino fundamental e médio por terem despertado meu interesse pelas Ciências Exatas e aos professores da Universidade Federal de Minas Gerais (UFMG) por terem me dado uma nova paixão: a Estatística.

Em especial, queria deixar meu agradecimento ao meu orientador Prof. Dr. Fabio Nogueira Demarqui, pela orientação, suporte e conhecimento transmitido. Agradeço também a Prof.ª Doutora Sueli Aparecida Mingoti, a quem tenho profunda admiração, pelo conhecimento transmitido ao longo das várias disciplinas que lecionou para minha turma. A Profª Doutora Edna Afonso Reis por toda sua ajuda enquanto coordenadora do curso de Estatística e a tanto outros professores do Departamento de Estatística que de alguma forma, deixaram marcas e ensinamentos que levarei comigo para vida toda.

Agradeço ao meu namorado, João Pedro Schneider, pela paciência, suporte e incentivo dados ao longo do curso e desse trabalho.

Não posso deixar de agradecer aos meus amigos, que sempre estiveram do meu lado me apoiando, e compreendendo minhas ausências a cada final de período. E obrigada também, a aqueles que direta ou indiretamente, fizeram parte da minha formação, como minhas colegas de apartamento e a Fundação Mendes Pimentel.

Por fim, agradeço á Pró-Reitoria de Graduação da UFMG, em particular, a chefe do Setor de Estatística Carolina Silva Pena, pelo apoio e conselhos dados, além da disponibilização dos dados que foram utilizados nesse trabalho.

Resumo

Nesse trabalho, são apresentados os passos necessários para fazer o ajuste de um modelo de regressão logística hierárquico desbalanceado nos *softwares* R e SAS. Na metodologia apresentada no trabalho, é feita uma revisão dos Modelos Lineares Generalizados, com foco no modelo logístico. Também é falado sobre a análise de contrastes, a razão de chances, calculada de maneira diferente em alguns casos, e uma breve explicação sobre os modelos hierárquicos. A metodologia estudada é ilustrada através da análise de um banco de dados reais, referente ao desempenho dos alunos da Universidade Federal de Minas Gerais (UFMG) na disciplina de Cálculo Diferencial e Integral I entre os anos de 1993 e 2013. O modelo logístico hierárquico desbalanceado ajustado tinha como variável resposta a aprovação\reprovação dos alunos na disciplina selecionada, e as variáveis explicativa eram o curso, a instituição e o RSG dos alunos presentes no estudo. Dentre os resultados obtidos, observou-se um melhor desempenho dos alunos da Escola de Engenharia com relação aos demais institutos. Os contrastes indicaram que os parâmetros do modelo são bem diferentes entre si, e a razão de chances mostrou, por exemplo, que no ICEX, somente quatro cursos possuem uma chance menor de serem aprovados em cálculo I que o nível de referência adotado para essa instituição, que foi o curso de Matemática Diurno.

Palavras chaves: Modelos Hierárquicos Desbalanceados, Modelos Lineares Generalizados, SAS, R, desempenho acadêmico.

Sumário

Lista de Figuras

Lista de Tabelas

1. Introdução

Os *softwares* estatísticos são atualmente ferramentas muito poderosas e que permitem aos estatísticos utilizarem qualquer tipo de técnica que julgarem necessárias aos seus estudos e análises. Porém, muitas vezes ainda são encontradas dificuldades em fazer algumas análises, devido a falta de documentação apresentada pelos *softwares*. Nesses casos, corre-se o risco de utilizar alguma técnica de maneira errônea ou interpretar de maneira equivocada os resultados apresentados.

Uma análise que atualmente ainda gera muita confusão nos *softwares* é o ajuste de modelos hierárquicos de efeitos fixos desbalanceados. Baseado nisso, pretende-se nesse trabalho, desenvolver um pequeno roteiro de como preparar os dados e fazer o ajuste de modelos desse tipo no SAS 9.2 e no R 3.1.1 através de uma aplicação.

A base de dados selecionada para ser utilizada como aplicação se refere ao desempenho acadêmico dos estudantes da Universidade Federal de Minas Gerais (UFMG) que fizeram a disciplina de Cálculo Integral de Diferencial I.

A motivação ao para a utilização dessa base de dados é que existem, em todas as universidades, algumas disciplinas que são feitas por alunos de diversos cursos. Essas disciplinas são, em geral, do ciclo básico de algumas áreas do conhecimento, ou disciplinas multidisciplinares que podem ser feitas por estudantes de todos os cursos.

Observa-se, nesse contexto, que alguns cursos possuem uma proporção maior de aprovações que outros cursos, quando estes fazem a mesma disciplina. Portanto, pretende-se através de um modelo investigar melhor essa relação entre os cursos e a proporção de aprovação\reprovação e, através de alguns métodos estatísticos realizar comparações entre os cursos quanto ao desempenho nessa disciplina.

O modelo que pretende-se ajustar é o Modelo de Regressão Logística, que pertence a família dos Modelos Lineares Generalizados, e incluir a estrutura hierárquica desbalanceada ao modelo.

Na Seção 2 será mostrada a metodologia utilizada no trabalho, como as definições e características dos Modelo Linear Generalizado, e maneiras de verificar a adequação do ajuste, com foco no modelo de Regressão Logística. Ao fim dessa seção, é mostrado uma estrutura hierárquica e os procedimentos necessários para a análise de contrastes no caso de um modelo de Regressão Logística Hierárquica Desbalanceada.

A análise descritiva da base de dados que será utilizada é exibida na Seção 3, e o ajuste do modelo, com a preparação dos dados, resultados obtidos, análise de diagnóstico, contrastes e razão de chances é mostrada na Seção 4. Por fim, na Seção 5 tem-se a conclusão a cerca do trabalho desenvolvido.

2. Metodologia

2.1. Modelos Lineares Generalizados

Uma das etapas mais importante de uma análise estatística consiste no ajuste do modelo, que deve ser o mais parcimonioso possível, e, ao mesmo tempo, descrever de maneira satisfatória o comportamento dos dados.

No passado, várias técnicas de modelagem estatísticas eram estudadas separadamente. Mas em 1972, Nelder e Wedderburn mostraram que tais técnicas podiam ser formuladas de maneira unificada, originando os chamados Modelos Lineares Generalizados (MLG).

Nas próximas seções, serão apresentadas as definições e características desses modelos, bem como maneiras de verificar a adequação do ajuste.

2.1.1. Família Exponencial

Muitas distribuições, tais como as distribuições normal, binomial, binomial negativa, Poisson, normal inversa, multinomial, beta, entre outras pertencem a uma mesma família de distribuição, denominada família exponencial.

Para definir a família exponencial, considere $\phi > 0$ um parâmetro de dispersão e o parâmetro θ_i denominado parâmetro canônico de Y_i. Dessa maneira, tem-se que a família exponencial é caracterizada por uma função de probabilidade ou densidade da forma:

$$f(y_i;\theta_i,\phi) = \exp\left(\phi^{-1}\left[y_i\theta_i - b(\theta_i)\right] + c(y_i,\phi)\right) \quad (1)$$

em que $b(.)$ e $c(.)$ são funções conhecidas. Sabe-se que

$$E(Y_i) = \mu_i = b'(\theta_i) \text{ e } Var(Y_i) = \phi b''(\theta_i) = \phi V_i$$

E $V_i = V(\mu_i) = d\mu_i/\mu_i\theta_i$ é denominada função de variância e depende unicamente da média de μ_i. A família exponencial é importante nas construções dos MLG, pois a única exigência dessa classe de modelos é que a distribuição da variável resposta pertença a essa família.

2.1.2. Definição

Os modelos lineares generalizados podem ser usados quando se tem uma única variável aleatória Y associada a um conjunto de variáveis explicativas $x_1, \dots, x_p$. Um MLG possui três componentes que o define, que são:

I. Componente Aleatório: representado pelo conjunto de variáveis aleatórias independentes $Y_1, \dots, Y_p$ provenientes de uma mesma família de distribuição, ou seja, $E(Y_i) = \mu_i, \ \ i = 1, \dots, n$. O componente aleatório de um modelo generalizado é definido a partir da família exponencial.

II. Componente Sistemático: as variáveis explicativas entram na forma de uma soma linear de seus efeitos

$$\eta_i = \sum_{r=1}^{p} x_{ir}\beta_j = \boldsymbol{x}_i^T\beta \text{ ou } \boldsymbol{\eta} = \boldsymbol{X\beta} \quad (2)$$

Sendo **X** a matriz do modelo, **β** o vetor de parâmetros e $\boldsymbol{\eta}$ o preditor linear.

III. Função de ligação: uma função de relaciona o componente aleatório ao componente sistemático, ou seja, relaciona a média ao preditor linear da seguinte maneira:

$$\eta_i = g(\mu_i) \quad (3)$$

em que $g(.)$ é uma função monótona e diferenciável.

Portanto, definir a distribuição da variável resposta, a matriz do modelo e a função de ligação são decisões importantes, pois essas são as características principais de um MLG.

A escolha da função de ligação depende, em geral, problema em questão. Existem vários tipos de funções de ligação, tais como a potência, *probito*, identidade, recíproca, complemento *log-log*, logística, dentre outras. Se a função de ligação é escolhida de tal forma que $g(\mu_i) = \theta_i = \eta_i$, ela é chamada de ligação canônica, pois o preditor linear modela diretamente o parâmetro canônico θ_i. Nesses casos os modelos ajustados são denominados canônicos. As funções de ligação canônicas para as principais distribuições estão apresentadas na Tabela 1.

Tabela 1: Funções de ligação canônicas

Distribuição	Função de ligação canônica	
Normal: N(μ,σ²)	Identidade:	$\eta = \mu$
Poisson: P(μ)	Logarítmica:	$\eta = \log(\mu)$
Binomial: B(m,π)	Logística:	$\eta = \log\left(\frac{\pi}{1-\pi}\right) = \log\left(\frac{\mu}{m-\mu}\right)$
Binomial Negativa: BN(μ,k)		$\eta = \log\left(\frac{\mu}{\mu+k}\right)$
Gama: G(μ,ν)	Recíproca:	$\eta = \frac{1}{\mu}$
Normal Inversa:NI(π)	Recíproca do quadrado:	$\eta = \frac{1}{\mu^2}$

O uso de outras funções de ligação implica em certas restrições. Por exemplo, quando se trabalha com a distribuição de Poisson, em que $\mu > 0$, não pode ser usada uma função de ligação que poderá assumir valores negativos dependendo dos valores obtidos para $\hat{\beta}$.

2.1.3. Estimação

Muitos métodos podem ser usados para estimar os parâmetros β's, tais como o *qui-quadrado* mínimo, o Bayesiano e a estimação-M, sendo que esse último contém o método de máxima verossimilhança (MV). O método MV será utilizado nesse trabalho por possuir excelentes propriedades tais como consistência e eficiência assintótica.

Para tanto, define-se $L(\beta) = L(\beta; y)$ o logaritmo da função de verossimilhança como função de β dado o vetor y e através da Equação (1) tem-se:

$$l(\beta) = \frac{1}{\phi}\sum_{i=1}^{n}[y_i\theta_i - b(\theta_i)] + \sum_{i=1}^{n} c(y_i, \phi) \tag{4}$$

em que $\theta_i = q(\mu_i), \mu_i = g^{-1}(\eta_i)$ e $\eta_i = \sum_{r=1}^{P} x_{ir}\beta_r$. Na expressão (4) utilizando a regra da cadeia, pode-se calcular o vetor escore, formado pelas derivadas parciais de primeira ordem do logaritmo da função de verossimilhança $U(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ de dimensão *p* com elemento típico:

$$U_r = \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_r} = \sum_{i=1}^{n} \frac{dl_i}{d\theta_i}\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{d\eta_i}{d\beta_r} = \frac{1}{\phi}\sum_{i=1}^{n}(y_i - \mu_i)\frac{1}{V_i}\frac{d\mu_i}{d\eta_i}x_{ir} \quad (5)$$

em que $\mu_i = b'(\theta_i)$ e $\frac{d\mu_i}{d\theta_i} = V_i$ para $r = 1, \ldots, p$.

A estimativa de máxima verossimilhança (EMV) $\hat{\beta}$ do vetor de parâmetros de β é obtida igualando U_r a zero para $r = 1, \ldots, p$. Em geral, essas equações são não-lineares e por isso, faz-se necessário o uso de métodos numéricos para resolvê-las. Alguns desses métodos são:

- Método Newton-Raphson

O Método de Newton-Raphson é um método interativo utilizado para resolver equações não-lineares, tais como equações cuja solução determina o ponto no qual a função atinge seu máximo.

Para obter a solução, deve-se considerar um $\hat{\beta}$ que maximiza a função $L(\beta)$. Seja $U' = \left(\frac{\partial \mathrm{L}(\beta)}{\partial \beta_1}, \frac{\partial \mathrm{L}(\beta)}{\partial \beta_2}, \ldots\right)$ as derivadas do vetor escore e $h_{ab} = \partial^2 \mathrm{L}(\beta)/\partial\beta_a\partial\beta_b$ as entradas da matriz Hessiana H. Considere $U^{(t)}$ e $H^{(t)}$como U e H avaliadas na t-ésima interação de $\hat{\beta}$ denominado $\beta^{(t)}$. O passo t do processo interativo $(t = 0, 1, \ldots)$ aproxima o máximo de $L(\beta)$ próximo de $\beta^{(t)}$ por termos que seguem até a segunda ordem da expansão da série de Taylor:

$$\mathrm{L}(\beta) \approx \mathrm{L}\left(\beta^{(t)}\right) + \mathrm{U}^{(t)'}\left(\beta - \beta^{(t)}\right) + (1/2)\left(\beta - \beta^{(t)}\right)' H^{(t)}\left(\beta - \beta^{(t)}\right) \quad (6)$$

A próxima interação é obtida resolvendo $\frac{\partial \mathrm{L}(\beta)}{\partial \beta} \approx U^{(t)} + H^{(t)}\left(\beta - \beta^{(t)}\right) = 0$ em β, e pode ser expresso como

$$\beta^{(t+1)} = \beta^{(t)} - \left(H^{(t)}\right)^{-1} U^{(t)} \quad (7)$$

Assumindo $H^{(t)}$ não singular.

As interações são feitas até que as mudanças em $L(\beta^{(t)})$ entre sucessivos palpites sejam suficientemente pequenas. O estimador MV é o limite de $\beta^{(t)}$ quanto $t \to \infty$. No entanto, isso não necessariamente ocorre se $L(\beta)$ possui outro máximo local onde a derivada de $L(\beta) = 0$. Esse método é bastante útil quando as derivadas parciais de segunda ordem são facilmente obtidas.

- Método Escore de Fisher

O Escore de Fisher é um método interativo alternativo para resolver equações de verossimilhança similar ao método Newton-Raphson. O método escore de Fisher é em geral, mais simples e gera o mesmo resultado que o método de Newton-Raphson quando a função de ligação utilizada é canônica, sendo usado principalmente quando as derivadas de parciais de segunda ordem da função não são obtidas facilmente.

A diferença entre os dois métodos é que no escore de Fisher a matriz de derivadas parciais de segunda ordem é substituída pela matriz de valores esperados das derivadas parciais, ou seja, a matriz de informação observada H é substituída pela matriz de informação esperada de Fisher K.

Considere $K^{(t)}$ a aproximação t para o EMV da matriz de informação esperada. Nesse caso, $K^{(t)}$ possui elementos $-E(\partial^2 \mathrm{L}(\beta)/\partial\beta_a\partial\beta_b)$ avaliados em $\beta^{(t)}$. Assim, tem-se que:

$$\begin{aligned} \beta^{(t+1)} &= \beta^{(t)} - \left(K^{(t)}\right)^{-1} U^{(t)} \\ K^{(t)}\beta^{(t+1)} &= K^{(t)}\beta^{(t)} + U^{(t)} \end{aligned} \tag{8}$$

Para modelos com ligações não-canônicas, o método Escore de Fisher tem mais vantagens que o método de Newton-Raphson, pois produz a matriz de covariância assintótica como subproduto e a informação esperada é necessariamente não-negativa definida.

- Método interativo de mínimos quadrados ponderados

Existe uma relação entre a estimativa via ponderação dos mínimos quadrados e a estimativa utilizando Escore de Fisher para encontrar estimativas MV, que será mostrada a seguir. Considerando o elemento típico de K igual à $k_{r,s} = E(U_r, U_s) = \phi^{-1}\sum_{i=1}^{n} w_i x_{ir} x_{is}$, onde $w_i = \frac{1}{V_i}\left(\frac{d\mu_i}{d\eta_i}\right)^2$ são pesos, tem-se que a matriz de informação de Fisher para **β** tem forma:

$$\boldsymbol{K} = \phi^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}$$

sendo $\boldsymbol{W} = diag\{w_1, \ldots, w_n\}$ uma matriz diagonal de pesos que traz a informação sobre a distribuição e a função de ligação utilizada.

Considerando que a função de ligação utilizada é canônica, tem-se que $w_i = V_i$. O componente $\boldsymbol{U} = \boldsymbol{U}(\boldsymbol{\beta})$ pode ser escrito na forma $\boldsymbol{U} = \phi^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{G}(\boldsymbol{y} - \boldsymbol{\mu})$, com $\boldsymbol{G} = diag\left\{\frac{d\eta_1}{d\mu_1}, \ldots, \frac{d\eta_n}{d\mu_n}\right\} = diag\{g'(\mu_1), \ldots, g'(\mu_n)\}$ que representa as derivadas de primeira ordem da função de ligação. Substituindo **K** e **U** em (8) e eliminando o ϕ, tem-se:

$$\boldsymbol{X}^T\boldsymbol{W}^{(t)}\boldsymbol{X}\boldsymbol{\beta}^{(t+1)} = \boldsymbol{X}^T\boldsymbol{W}^{(t)}\boldsymbol{X}\boldsymbol{\beta}^{(t)} + \boldsymbol{X}^T\boldsymbol{W}^{(t)}\boldsymbol{G}^{(t)}(\boldsymbol{y} - \boldsymbol{\mu}^{(t)})$$

Ou, ainda,

$$\boldsymbol{X}^T\boldsymbol{W}^{(t)}\boldsymbol{X}\boldsymbol{\beta}^{(t+1)} = \boldsymbol{X}^T\boldsymbol{W}^{(t)}\left[\boldsymbol{\eta}^{(t)} + \boldsymbol{G}^{(t)}(\boldsymbol{y} - \boldsymbol{\mu}^{(t)})\right]$$

Definindo a variável dependente ajustada $\boldsymbol{z} = \boldsymbol{\eta} + \boldsymbol{G}(\boldsymbol{y} - \boldsymbol{\mu})$, obtem-se o seguinte resultado:

$$\boldsymbol{\beta}^{(t+1)} = \left(\boldsymbol{X}^T\boldsymbol{W}^{(t)}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}^{(t)}\boldsymbol{z}^{(t)} \quad (9)$$

A equação (9) é válida para qualquer MLG e mostra que a solução das equações de MV equivale a calcular repetidamente uma regressão linear ponderada de uma variável dependente ajustada **z** sobre a matriz **X** usando uma função de peso **W** que se modifica no processo interativo. As funções de variância e de ligação entram no processo interativo através de **W** e **z**.

Observa-se também que a equação interativa (9) não depende do parâmetro de dispersão ϕ. Sua demonstração em generalidade foi dada por Nelder e Wedderburn (1972) e atualmente é o algoritmo de estimação mais utilizado nos softwares estatísticos que realizam ajuste de modelos GLM.

2.1.4. Início e critério de parada do algoritmo de estimação

O método usual para iniciar o processo interativo é especificar uma estimativa inicial que será alterada sucessivamente até a convergência. Cada observação pode ser considerada uma estimativa do seu valor médio, isto é, $\mu_i^{(1)} = y_i$ e portanto, calcula-se:

$$\eta_i^{(1)} = g\left(\mu_i^{(1)}\right) = g(y_i) \text{ e } w_i^{(1)} = \frac{1}{V(y_i)[g'(y_i)]^2}$$

Usando $\eta_i^{(1)}$ como variável resposta, $\boldsymbol{X}$ a matriz do modelo e $\boldsymbol{W}^{(1)}$ a matriz diagonal de pesos, obtém-se o vetor $\beta^{(2)} = \left(\boldsymbol{X}^T\boldsymbol{W}^{(1)}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}^{(1)}\eta^{(1)}$. Assim sendo, o algoritmo para $t = 2,\ldots,k$ sendo $k-1$ o número necessário de interações até atingir a convergência pode ser resumido em:

(1) Obter as estimativas

$$\eta_i^{(t)} = \sum_{r=1}^{p} x_{ir}\beta_r^{(t)} \text{ e } \mu_i^{(t)} = g^{-1}\left(\eta_i^{(t)}\right)$$

(2) Obter a variável dependente ajustada

$$z_i^{(t)} = \eta_i^{(t)} + \left(y_i - \mu_i^{(t)}\right)g'\left(\mu_i^{(t)}\right)$$

e os pesos

$$w_i^{(t)} = \frac{1}{V(\mu_i^{(t)})\left[g'(\mu_i^{(t)})\right]^2}$$

(3) Calcular

$$\boldsymbol{\beta}^{(t+1)} = \left(\boldsymbol{X}^T\boldsymbol{W}^{(t)}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}^{(t)}\boldsymbol{z}^{(t)}$$

Voltar ao passo (1) com $\boldsymbol{\beta}^{(t)} = \boldsymbol{\beta}^{(t+1)}$ e repetir o processo a convergência.

Em geral, esse algoritmo é robusto e converge rapidamente, com no máximo 10 interações. Um dos critérios existentes que pode ser utilizado para verificar a convergência é

$$\sum_{r=1}^{p}\left(\frac{\beta_r^{(t+1)} - \beta_r^{(t)}}{\beta_r^{(t)}}\right)^2 < \xi$$

para ξ um valor suficientemente pequeno. Um fato que deve ser observado é se a função $g(.)$ pode não ser definida para alguns valores y_i. Por exemplo, se a função de ligação for $\eta = g(\mu) = \log(\mu)$ e forem observados valores $y_i = 0$, o processo não pode ser iniciado. Para contornar esse problema, pode-se substituir y por $y + c$ tal que $E[g(y+c)]$ seja o mais próximo possível de $g(\mu)$. Para o modelo logístico usa-se em geral $c = \frac{1-2\pi}{2}$ e $\pi = \frac{\mu}{t}$, sendo t o índice da distribuição binomial. De forma

geral, usando a expansão de Taylor até a segunda ordem para $g(y+c)$ em relação a $g(\mu)$, obtém-se $c \approx -\frac{1}{2} Var(Y) \frac{g''(\mu)}{g'(\mu)}$.

Para pequenas amostras, a equação (9) pode divergir. Embora o algoritmo convirja rapidamente, o número de interações até a convergência depende inteiramente do valor inicial de $\hat{\beta}$.

2.1.5. Comentários adicionais

Para as funções de ligação canônicas que produzem os modelos denominados canônicos, as equações de MV têm a seguinte forma:

$$\boldsymbol{X}^T \boldsymbol{y} = \boldsymbol{X}^T \hat{\mu} \tag{10}$$

em que pode-se concluir que as estimativas de MV dos ***β's*** são únicas. A equação (10) é válida para os modelos canônicos clássico de regressão, modelo log-linear, modelo logístico linear, modelo gama com função de ligação recíproca e modelo normal inverso com função de ligação recíproca ao quadrado. Para os modelos canônicos o ajuste é feito pelo algoritmo (8) com $\boldsymbol{W} = diag\{V_i\}$, $\boldsymbol{G} = diag\{V_i^{-1}\}$ e variável dependente ajustada com componente típica expressa por $z_i = \eta_i + (y_i - \mu_i)/V_i$. Nos modelos com respostas binárias, a variável resposta tem distribuição binomial $B(m_i, \pi_i)$, e o logaritmo da função de verossimilhança em (4) é expresso como

$$l(\beta) = \sum_{i=1}^{n} \left\{ y_i log\left(\frac{\mu_i}{t_i - \mu_i}\right) + t_i log\left(\frac{t_i - \mu_i}{t_i}\right) \right\} + \sum_{i=1}^{n} log\binom{t_i}{y_i}$$

em que $\mu_i = t_i \pi_i$. No caso especial do modelo logístico linear que será utilizado nesse trabalho, obtém-se $\eta_i = g(\mu_i) = \log[\mu_i / (t_i - \mu_i)]$.

As interações do processo interativo de estimação dos parâmetros são realizadas com matriz de pesos $\boldsymbol{W} = diag\left\{\frac{\mu_i(t_i - \mu_i)}{t_i}\right\}$, $\boldsymbol{G} = diag\left\{\frac{t_i}{\mu_i(t_i - \mu_i)}\right\}$ e variável dependente ajustada com componentes iguais a $z_i = \eta_i + [t_i(y_i - \mu_i)]/[\mu_i(t_i - \mu_i)]$. O algoritmo de estimação converge exceto quando ocorrem médias ajustadas próximas de zero ou ao índice t_i.

Com a obtenção da EMV de $\widehat{\boldsymbol{\beta}}$, é possível calcular as estimativas de MV dos preditores lineares $\widehat{\boldsymbol{\eta}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ e das medias $\widehat{\boldsymbol{\mu}} = g^{-1}(\widehat{\boldsymbol{\eta}})$. A EMV do vetor de parâmetro canônicos θ é simplesmente $\hat{\theta} = q(\widehat{\boldsymbol{\mu}})$.

A estrutura de covariância assintótica de $\widehat{\boldsymbol{\beta}}$ dada pelo inverso da matriz de informação de Fisher avaliada em $\theta = \hat{\theta}$, pode ser escrita por:

$$\widehat{Cov}(\widehat{\boldsymbol{\beta}}) = \phi\left(\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{W}}\boldsymbol{X}\right)^{-1} \tag{11}$$

em que $\widehat{\boldsymbol{W}}$ é o valor de matriz de pesos **W** avaliada em $\widehat{\boldsymbol{\beta}}$. A partir de (11) pode-se deduzir os intervalos de confiança assintóticos para os parâmetros β's. Observa-se que o parâmetro de dispersão ϕ é um fator multiplicativo na matriz de covariância assintótica de $\widehat{\boldsymbol{\beta}}$. Assim, se $Var(\hat{\beta}_r)$ é o elemento (r,r) da matriz $\phi\left(\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{W}}\boldsymbol{X}\right)^{-1}$, um intervalo de 95% de confiança para β_r pode ser obtida dos limites

$$\hat{\beta}_r \pm 1{,}96 Var(\hat{\beta}_r)^{1/2}$$

Na prática, uma estimativa consistente de ϕ deve ser inserida nesse intervalo.

A estrutura da covariância assintótica das estimativas de MV dos preditores lineares em $\widehat{\boldsymbol{\eta}}$ é obtida diretamente de $Cov(\widehat{\boldsymbol{\eta}}) = \boldsymbol{X}Cov(\widehat{\boldsymbol{\beta}})\boldsymbol{X}^{\boldsymbol{T}}$. Logo,

$$Cov(\widehat{\boldsymbol{\eta}}) = \phi\boldsymbol{X}\left(\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{W}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}} \tag{12}$$

Considerando a estrutura de covariância assintótica das estimativas de MV das médias em $\widehat{\boldsymbol{\mu}}$**,** que pode ser calculada expandindo $\widehat{\boldsymbol{\mu}} = g^{-1}(\widehat{\boldsymbol{\eta}})$ em uma série de Taylor, tem-se:

$$\widehat{\boldsymbol{\mu}} = \boldsymbol{\eta} + (\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})\frac{\boldsymbol{d}g^{-1}(\boldsymbol{\eta})}{\boldsymbol{d\eta}}$$

e, portanto,

$$Cov(\widehat{\boldsymbol{\mu}}) = \boldsymbol{G}^{-1}Cov(\widehat{\boldsymbol{\eta}})\boldsymbol{G}^{-1} \tag{13}$$

em que $\boldsymbol{G} = diag\{d\boldsymbol{\eta}/d\boldsymbol{\mu}\}$. Essa matriz é estimada por

$$\widehat{Cov}(\widehat{\boldsymbol{\mu}}) = \phi\boldsymbol{G}^{-1}\boldsymbol{X}\left(\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{W}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{G}^{-1}$$

As matrizes $Cov(\hat{\boldsymbol{\eta}})$ e $Cov(\hat{\boldsymbol{\mu}})$ são de ordem n^{-1}. Os erros-padrão $\hat{z}_{ii}^{1/2}$ de $\hat{\boldsymbol{\eta}}_{\boldsymbol{i}}$ e os coeficientes de correlação estimados $Corr(\hat{\eta}_i, \hat{\eta}_j) = \frac{\hat{z}_{ij}}{(\hat{z}_{ii}\hat{z}_{jj})^{1/2}}$ dos preditores lineares estimados $\hat{\eta}_1, \ldots, \hat{\eta}_n$, são resultados aproximados que dependem fortemente do tamanho da amostra. Entretanto, são guias úteis de informação sobre a confiabilidade e a independência das estimativas dos preditores lineares, além de também poderem ser usadas para obter intervalos de confiança aproximados para esses parâmetros.

2.1.6. Inferência

Em geral, a obtenção de distribuições exatas é muito complicada e por esse motivo, os resultados assintóticos são usados. Entretanto, esses resultados dependem de algumas condições de regularidade, bem como o número de observações disponíveis que, em particular, são válidas para os MLG.

Considerando $\hat{\theta}$ é um estimador consistente para um parâmetro θ e $Var(\hat{\theta})$ é a variância desse estimador, então, para amostras grandes, pode-se afirmar que:

I) $\hat{\theta}$ é assintoticamente imparcial;

II) A estatística

$Z_n = \frac{\hat{\theta}-\theta}{\sqrt{Var(\hat{\theta})}} \rightarrow Z$ quando $n \rightarrow \infty$, sendo que $Z \sim N(0,1)$ ou de forma equivalente,

$Z_n^2 = \frac{(\hat{\theta}-\theta)^2}{Var(\hat{\theta})} \rightarrow Z^2$ quando $n \rightarrow \infty$, sendo que $Z^2 \sim x_1^2$.

Se $\hat{\theta}$ é um estimador consistente de um vetor **θ** de *p* parâmetros, tem-se, assintoticamente, que

$$(\hat{\theta} - \theta)^T V^{-1} (\hat{\theta} - \theta) \sim x_p^2$$

em que V é a matriz de variâncias e covariâncias, suposta não singular. Se V é singular, usa-se uma matriz inversa generalizada ou, então, uma reparametrização de forma a se obter uma nova matriz de variância e covariâncias não singular.

Considere o vetor escore $\boldsymbol{U(\beta)} = \frac{\partial \mathrm{L}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ que tem valor esperado zero e estrutura de covariância igual à matriz de informação **K** em problemas regulares. Conforme mostrado anteriormente, a matriz de informação nos MLG é dada por $\boldsymbol{K} = \phi^{-1}\boldsymbol{X^T\widehat{W}X}$**.** O teorema central do limite aplicado a $\boldsymbol{U(\beta)}$ implica que a distribuição assintótica de $\boldsymbol{U(\beta)}$ é normal *p*-variada $N_p(\mathbf{0}, \boldsymbol{K})$. Para amostras grandes, a estatística escore definida na forma quadrática $E = \boldsymbol{U(\beta)}^T\boldsymbol{K}^{-1}\boldsymbol{U(\beta)}$ tem aproximadamente distribuição x_p^2, supondo verdadeiro o modelo com vetor de parâmetro β especificado.

Resumidamente, pode-se afirmar que:

I) O estimador $\hat{\beta}$ é assintoticamente não viesado.

II) A matriz de variância e covariância de $\boldsymbol{\widehat{\beta}}$ para amostras grandes é:

$$Cov(\boldsymbol{\widehat{\beta}}) = \boldsymbol{K^{-1}}$$

Pois $\boldsymbol{K^{-1}}$ é simétrica.

III) Para amostras grandes, tem-se a seguinte aproximação:

$$(\boldsymbol{\widehat{\beta}} - \boldsymbol{\beta})^{\boldsymbol{T}}\boldsymbol{K}(\boldsymbol{\widehat{\beta}} - \boldsymbol{\beta}) \sim x_p^2 \tag{14}$$

ou, equivalentemente,

$$\boldsymbol{\widehat{\beta}} \sim N_p(\boldsymbol{\beta}, \boldsymbol{K^{-1}}) \tag{15}$$

Pra amostras pequenas, o estimador $\boldsymbol{\widehat{\beta}}$ é viesado, e, portanto, torna-se necessário calcular seu viés de ordem n^{-1}. Além disso, a estrutura de covariância das estimativas de MV dos parâmetros lineares difere de $\boldsymbol{K^{-1}}$**.**

A distribuição assintótica normal *p*-variada $N_p(\boldsymbol{\beta}, \boldsymbol{K^{-1}})$ de $\boldsymbol{\widehat{\beta}}$ é a base da construção dos testes de significância do parâmetro e intervalos de confiança quando a amostra é suficientemente grande. O erro da aproximação $N(\boldsymbol{\beta}, \boldsymbol{K^{-1}})$ para a distribuição de $\boldsymbol{\widehat{\beta}}$ é da ordem n^{-1} em probabilidade.

Os erros-padrão dos estimadores de$\hat{\beta}_{\boldsymbol{1}}, \ldots, \hat{\beta}_{\boldsymbol{p}}$ são iguais às raízes quadradas dos elementos da diagonal principal de $\boldsymbol{\widehat{K}^{-1}}$ e fornecem informações sobre a

exatidão de tais estimadores. Usando que $Cov(\hat{\beta}_{\boldsymbol{r}}, \hat{\beta}_{\boldsymbol{s}}) = k^{r,s}$, obtem-se ao nível de 95% de confiança que os intervalos para os parametros $\hat{\beta}_{\boldsymbol{r}}'s$ podem ser escritos como:

$$\hat{\beta}_{\boldsymbol{r}} \pm 1{,}96\sqrt{\hat{k}^{r,r}}$$

em que $\hat{k}^{r,r} = \widehat{Var}(\hat{\beta}_{\boldsymbol{r}})$.

Para verificar a independência entre $\hat{\beta}'_{\boldsymbol{r}}s$ é utilizado a correlação $\rho_{rs} = \widehat{Corr}(\hat{\beta}_{\boldsymbol{r}}, \hat{\beta}_{\boldsymbol{s}}) = \frac{\hat{k}^{r,s}}{\sqrt{\hat{k}^{r,r}\hat{k}^{s,s}}}$, sendo obtida diretamente da inversa da informação de **K** avaliada em $\widehat{\boldsymbol{\beta}}$**.**

2.1.7. Função Desvio e Estatística de Pearson generalizada

O ajuste de um modelo a um conjunto de observações pode ser tratado como uma maneira de substituir **y** por um conjunto de valores estimados $\widehat{\boldsymbol{\mu}}$. Logicamente, os valores de $\widehat{\boldsymbol{\mu}}$'s não serão exatamente iguais aos **y**'. Nesse ponto surge uma questão que deve ser analisada, que é estudar o quanto esses dois valores diferem.

Admitindo uma combinação satisfatória da distribuição da variável resposta e da função de ligação, deseja-se determinar quantos termos são necessários para uma descrição razoável dos dados. Um modelo com muitas variáveis explicativas pode explicar bem os dados, porém pode dificultar a interpretação do modelo. Por outro lado, o uso de poucas variáveis explicativas simplifica a interpretação, mas pode apresentar um ajuste pobre dos dados. Portanto, o que se procura é um modelo intermediário, que não seja muito complicado e ajuste bem o modelo.

O modelo mais simples é o modelo nulo, que contém um único parâmetro representando por um valor μ comum a todos os dados. A matriz do modelo nesse caso é reduzida á uma coluna composta apenas de 1's. Esse modelo atribui toda a variação entre os y's ao componente aleatório. No outro extremo, tem-se o modelo saturado ou completo, que considerando um conjunto de dados *n* observações é um modelo que possui *n* parâmetros, ou seja, possui um parâmetro para cada observação. O modelo saturado atribui toda a variação dos dados ao componente sistemático e assim, ajusta-se perfeitamente, reproduzindo os próprios dados.

Existem também, outros dois modelos que não são tão extremos quanto os modelos nulo e o saturado: o modelo minimal que contém o menor número de termos necessários para o ajuste e o maximal que por sua vez, contem o maior número de termos. Qualquer modelo com *p* parâmetros linearmente independentes entre os modelos minimal e maximal é chamado de modelo sob pesquisa ou modelo corrente. Em geral, trabalha-se com modelos encaixados e o conjunto de matrizes dos modelos pode, então, ser formada pela inclusão sucessiva de termos do modelo minimal até se chegar ao maximal. O problema é então determinar a utilidade de um parâmetro extra no modelo corrente ou verificar a falta de ajuste induzida por sua omissão.

Para discriminar os modelos, são utilizadas medidas de discrepância para medir o ajuste do modelo. Uma medida de discrepância amplamente utilizada é o "*deviance*", com expressão dada por:

$$S_p = 2(\hat{l}_n - \hat{l}_p)$$

Sendo $\hat{l}_n$ e $\hat{l}_p$ os máximos do logaritmo da função de verossimilhança dos modelos saturados e corrente, respectivamente. Nota-se, portanto, que o modelo saturado é usado como base de medida do ajuste do modelo corrente. Do logaritmo da função de verossimilhança obtém-se:

$$\hat{l}_n = \frac{1}{\phi}\sum_{i=1}^{n}\left[y_i\tilde{\theta}_i - b(\tilde{\theta}_i)\right] + \frac{1}{\phi}\sum_{i=1}^{n} c(y_i, \phi)$$

e

$$\hat{l}_p = \frac{1}{\phi}\sum_{i=1}^{n}\left[y_i\hat{\theta}_i - b(\hat{\theta}_i)\right] + \frac{1}{\phi}\sum_{i=1}^{n} c(y_i, \phi)$$

Sendo $\tilde{\theta}_i = q(y_i)$ e $\hat{\theta}_i = q(\hat{\mu}_i)$ as estimativas de MV do parâmetro canônico sob os modelos saturados e corrente. Então:

$$S_p = \frac{D_p}{\phi} = \frac{2}{\phi}\sum_{i=1}^{n}\left[y_i\left(\tilde{\theta}_i - \hat{\theta}_i\right) + b\left(\hat{\theta}_i\right) - b(\tilde{\theta}_i)\right] \quad (16)$$

em que S_p e D_p são denominados de desvio escalonado e desvio respectivamente. Pode-se ainda escrever

$$S_p = \frac{1}{\phi}\sum_{i=1}^{n} d_i^2$$

sendo d_i^2 a medida da diferença dos logaritmos das funções de verossimilhança observada e ajustada, para a observação *i* correspondente, e é chamado componente do desvio. A soma deles mede a discrepância total entre as duas funções de verossimilhança na escala logarítmica, sendo uma medida de distância entre o modelo corrente e o modelo saturado. Verifica-se que o desvio equivale a:

$$S_p = 2\left(\hat{l}_n - \hat{l}_p\right) = constante - 2\hat{l}_p$$

Dessa forma, se um modelo é bem (mal) ajustado aos dados, com uma verossimilhança máxima grande (pequena), tem um pequeno (grande) desvio. Na prática, deseja-se modelos simples com desvios moderados, situados entre os modelos mais complicados e os que se ajustam mal aos dados.

O desvio é calculado facilmente para qualquer MLG a partir da estimativa de MV de μ dada por $\hat{\mu} = g^{-1}(\boldsymbol{X}\hat{\beta})$. O desvio é sempre maior ou igual à zero e decresce a medida que covariáveis são acrescentadas ao modelo, sendo que quanto menor seu valor, melhor é o ajuste do modelo corrente. Para realizar o teste, definem-se os graus de liberdade do desvio do modelo por $v = n - p$, ou seja, número de observações menos o posto da matriz do modelo proposto. Na Tabela 2 são apresentadas funções desvios para os principais modelos.

Tabela 2: Funções desvios de alguns modelos

Modelo	Desvio
Normal	$D_p = \sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2$
Binomial	$D_p = 2\sum_{i=1}^{n}\left[y_i log\left(\frac{y_i}{\hat{\mu}_i}\right) + (m_i - y_i) log\left(\frac{m_i - y_i}{m_i - \hat{\mu}_i}\right)\right]$
Poisson	$D_p = 2\sum_{i=1}^{n}\left[y_i log\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i)\right]$
Binomial Negativo	$D_p = 2\sum_{i=1}^{n}\left[y_i log\left(\frac{y_i}{\hat{\mu}_i}\right) + (y_i + k) log\left(\frac{y_i + k}{\hat{\mu}_i + k}\right)\right]$
Gama	$D_p = 2\sum_{i=1}^{n}\left[log\left(\frac{\hat{\mu}_i}{y_i}\right) + \frac{(y_i - \hat{\mu}_i)}{\hat{\mu}_i}\right]$
Normal Inverso	$D_p = \sum_{i=1}^{n}\frac{(y_i - \hat{\mu}_i)^2}{y_i \hat{\mu}_i^2}$

Para o modelo normal com função de ligação identidade, assumindo-se que o modelo usado é o verdadeiro e que σ^2 é conhecido, tem-se o resultado exato

$$S_p = \frac{D_p}{\sigma^2} \sim x^2_{n-p}$$

Entretanto, para modelos normais com outras funções de ligação, esse resultado é apenas uma aproximação. No geral, porém, apenas alguns resultados assintóticos estão disponíveis e, em alguns casos, o desvio não tem distribuição x^2_{n-p}, nem mesmo assintoticamente. Então, o desvio é corrigido por uma correção de Bartlett proposta para os MLG por Cordeiro (1983, 1987, 1995) tem sido usada para melhorar a sua aproximação pela distribuição x^2_{n-p} de referência. O desvio modificado $S_p = \frac{(n-p)S_p}{\hat{E}(S_p)}$ em que a correção de Bartlett é dada por $\frac{(n-p)}{\hat{E}(S_p)}$ sendo $\hat{E}(S_p)$ o valor de $E(S_p)$ avaliado em $\hat{\mu}$, para ser melhor aproximado pela distribuição de referência do que o desvio S_p conforme comprovam os estudos de simulação de Cordeiro(1993).

Assumindo que o modelo usado é verdadeiro, para a distribuição binomial, quando *n* é fixo e $m_i \to \infty$, $\forall i$ e para a distribuição de Poisson quando $\mu_i \to \infty$, $\forall i$, tem-se com $\phi = 0$:

$$S_p = D_p \sim x^2_{n-p}$$

Nos casos em que S_p depende do parâmetro de dispersão ϕ:

$$S_p \sim x^2_{n-p}, quando\ \phi \to 0$$

Ou seja, quando o parâmetro de dispersão é pequeno, a aproximação de x^2_{n-p} para S_p é satisfatória. Na prática, de maneira geral, contenta-se em testar um MLG comparando o valor de S_p com os percentiis da distribuição x^2_{n-p}. Assim, quando

$$S_p = \phi^{-1} D_p \leq\ x^2_{n-p,\alpha}$$

pode-se considerar que existem evidências, a um nível aproximado de $100\alpha\%$ de confiança que o modelo proposto está bem ajustado aos dados. Ou se o valor D_p for próximo do valor esperado $n - p$ de uma distribuição x^2_{n-p}, pode ser uma indicação de que o modelo ajustado é adequado aos dados.

Uma outra medida de discrepância do ajuste de um modelo MLG a um conjunto de dados é a estatística de Pearson X_p^2 generalizada cuja expressão é:

$$X_p^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \tag{17}$$

Sendo $V(\hat{\mu}_i)$ a função de variância estimada sob o modelo corrente. Para respostas com distribuição normal tem-se que $X_p^2 \sim \sigma^2 X_{n-p}^2$ sendo esse resultado exato somente se a função de ligação for a identidade e a variância conhecida. Quando os dados são provenientes da distribuição binomial e de Poisson, em que $\phi = 1$, X_p^2 é a estatística original de Pearson, que tem a forma:

$$X_p^2 = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i}$$

Considerando o_i a freqüência observada e e_i a freqüência esperada.

Para as distribuições não-normais, têm-se apenas resultados assintóticos para X_p^2, isto é, a distribuição X_{n-p}^2 pode ser usada somente como uma aproximação de X_p^2, e em muitos casos pode ser inadequada.

A Expressão (17) da estatística de Pearson generalizada tem uma forma equivalente dada em termos da variável dependente ajustada no algoritmo mostrado na Equação (9). Tem-se:

$$X_p^2 = (z - \hat{\eta})^2 \widehat{W} (z - \hat{\eta})$$

O desvio S_p tem a grande vantagem como medida de discrepância por ser aditivo para um conjunto de modelos encaixados, enquanto X_p^2, em geral, não tem essa propriedade, apesar de ser preferido em relação ao desvio, devido sua facilidade de interpretação.

2.1.8. Análise do Desvio e seleção do modelo

Em geral, o algoritmo de ajuste não é aplicado a um único MLG, mas a vários modelos de um conjunto bem amplo que deve ser relevante para o tipo de dados que se pretende analisar. Sempre há o ajuste de vários modelos, pois caso seja ajustado somente um modelo sem levar em conta modelos alternativos, o modelo

ajustado obtido pode não ser o mais adequado aos dados. Esse conjunto de modelos pode ser definindo uma família de funções de ligação, ou considerando diferentes opções para a escala de medição, ou adicionando\retirando vetores colunas independentes a partir de uma matriz básica original.

Uma ferramenta utilizada para fazer a seleção de modelos é a análise do desvio, uma generalização da análise de variância para os MLG, visando a partir de uma sequências de modelos encaixados obter os efeitos de fatores, covariáveis e suas interações. O desvio é usado como medida de discrepância do modelo e faz-se uma tabela de diferença de desvios.

A título de ilustração, considere $M_{p1}, M_{p2}, \dots, M_{pr}$ uma sequência de modelos encaixados de respectivas dimensões $p_1 < p_2 < \cdots < p_r$, matrizes dos modelos $X_{p1}, X_{p2}, \dots, X_{pr}$ e desvios $D_{p1}, D_{p2}, \dots, D_{pr}$, tendo os modelos a mesma distribuição e a mesma função de ligação. Essas desigualdades entre os desvios, em geral, não são encontradas na estatística de Pearson, e por isso, a comparação de modelos encaixados é feita utilizando principalmente a função desvio. Assim, para o caso de um ensaio inteiramente casualizado, com *r* repetições e tratamentos no esquema fatorial, com *a* níveis no fator A e *b* níveis no fator B, obtêm-se os resultados mostrados na Tabela 3.

Tabela 3: Exemplo de construção de uma tabela de Análise de Desvio.

Modelo	gl	desvio	Dif. de Desvios	Dif. De gl	Significado
Nulo	*rab-1*	D_1			
A	*a(rb-1)*	D_A	D_1	*a-1*	A ignorando B
A+B	*a(rb-1)-(b-1)*	D_{A+B}	D_A-D_{A+B}	*b-1*	B incluído A
A+B+A.B	*ab(r-1)*	D_{A*B}	D_{A+B}–D_{A*B}	*(a-1)(b-1)*	Interação AB incluídos A e B
Saturado	*0*	0	D_{A*B}		Resíduo

Dois termos A e B são ortogonais se a redução que A (ou B) causa no desvio D_p é a mesma, esteja B (ou A) incluindo ou não no modelo corrente. Em geral, nos MLG ocorre a não-ortogonalidade dos termos, fazendo com que a interpretação da tabela ANODEV seja mais complicada que a ANOVA usual.

Sejam os modelos M_q e $M_p (M_q \subset M_p, q < p)$, com p e q parâmetros respectivamente. A estatística $D_q - D_p$ com $(p - q)$ graus de liberdade é interpretada

como uma medida de variação dos dados, explicada pelos termos que estão em M_p e não estão em M_q, incluídos os efeitos dos termos em M_q e ignorando outros efeitos que não estão em M_p. Dessa forma tem-se, assintoticamente, para ϕ conhecido

$$S_q - S_q = \phi^{-1}(D_q - D_q) \sim x^2_{p-q}$$

que é simplesmente a estatística da razão de verossimilhanças. Se o ϕ é desconhecido, deve-se obter uma estimativa consistente $\hat{\phi}$, de preferência baseada no modelo maximal, e a inferência baseada na estatística F dada por

$$F = \frac{(D_q - D_q)/(p-q)}{\hat{\phi}} \sim F_{p-q,n-m}$$

Quando o modelo minimal é o verdadeiro, $S_q - S_q$ tem distribuição assintótica x^2_{p-q}. Porém, cada desvio isolado não é distribuído, assintoticamente, como qui-quadrado.

2.1.9. Estimação do parâmetro de dispersão

Para as distribuições Binomial e Poisson tem-se que o parâmetro de dispersão $\phi = 1$, mas no caso das distribuições normal, normal inversa e gama, o ϕ é desconhecido. Nesses casos, admite-se que ele é constante para todas as observações e é estimado pelo método do desvio, método de Pearson e pelo método de máxima verossimilhança. Sua correta estimação é importante, pois é utilizado no cálculo do erro-padrão, intervalos de confiança e testes de hipóteses dos $\hat{\beta}'s$.

O método do desvio é baseado na aproximação x^2_{n-p} para o desvio escalonado dado pela equação (16). Para um modelo bem ajustado aos dados, espera-se, que o desvio escalonado S_p tenha valor esperado igual a $n - p$. Assim, obtém-se:

$$\hat{\phi}_d = \frac{D_p}{n-p} \tag{18}$$

em que o desvio D_p é obtido como função dos dados em y e dos valores ajustados em $\hat{\mu}$ de (16). O método de Pearson é baseado na aproximação da distribuição da

estatística de Pearson X_p^2 generalizada, dada em (17), mas dividida por ϕ, pela distribuição x_{n-p}^2. Dessa forma, obtêm-se a estimativa:

$$\hat{\phi}_P = \frac{1}{n-p}\sum_{i=1}^{n}\frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \tag{19}$$

Para o modelo normal $\hat{\phi}_d = \hat{\phi}_P$. Para os demais modelos contínuos, entretanto, esses estimadores diferem em valor.

O método de máxima verossimilhança nem sempre tratável computacionalmente, embora seja sempre possível em teoria. Se o ϕ é o mesmo para todas as observações, a EMV de β independe de ϕ enquanto a matriz de variâncias e covariâncias dos $\hat{\beta}'s$ envolve esse parâmetro. Interpretando o logaritmo da função de verossimilhança $l(\beta, \phi)$ como função de β e ϕ, dado y, pode-se escrever:

$$l(\beta, \phi) = \phi^{-1}\sum_{i=1}^{n}[y_i\theta_i - b(\theta_i)] + \sum_{i=1}^{n}c(y_i, \phi) \tag{20}$$

A função escore relativa ao parâmetro ϕ é dada por

$$U_\phi = \frac{\partial \mathrm{L}(\boldsymbol{\beta}, \phi)}{\partial \phi} = -\phi^{-2}\sum_{i=1}^{n}\left[y_i\theta_i - b(\theta_i)\right] + \sum_{i=1}^{n}\frac{dc(y_i, \phi)}{d\phi}$$

Observa-se que U_ϕ é função de β através de θ (ou μ) e de ϕ, supondo y conhecido. A EMV $\widehat{\phi}$ de ϕ é obtida igualando $\frac{\partial \mathrm{L}(\hat{\boldsymbol{\beta}}, \phi)}{\partial \phi}$ a zero. Pode-se ver que a EMV $\widehat{\phi}$ é função das médias ajustadas $\hat{\mu}$ e dos dados y.

2.2. Análise de Resíduos e Técnicas de Diagnóstico

Como já citado, a escolha de um modelo linear generalizado envolve a definição da distribuição, da função de ligação e da matriz do modelo. Entretanto, na prática, mesmo após uma escolha cuidadosa do modelo, ajuste do modelo pode ser insatisfatório. Isso pode ocorrer em função de algum desvio sistemático entre os valores observados e ajustados ou então, porque um ou mais valores observados são discrepantes dos demais.

Desvios sistemáticos podem ocorrer devido a uma escolha inadequada da função de variância, ou da função de ligação e da matriz do modelo, ou ainda pela

definição errada da escala da variável dependente ou das variáveis explanatórias. Pontos discrepantes isolados podem ocorrer porque os pontos estão nos extremos da amplitude de validade da covariável, ou porque eles estão realmente errados, como por exemplo, erros de digitação ou medição.

As técnicas usadas para esse fim podem ser formais ou informais. As informais são baseadas em exames visuais de gráficos para detecção de padrões ou pontos discrepantes. As formais envolvem colocar o modelo corrente sob pesquisa em uma classe maior de inclusão de um parâmetro extra γ. As mais usadas são baseadas nos testes de razão de verossimilhança e escore.

A análise de resíduos e diagnósticos para MLG possui técnicas semelhantes àquelas utilizadas para o modelo clássico de regressão, com algumas adaptações, como por exemplo, a variância residual s^2 é substituída por uma estimativa consistente do parâmetro ϕ e a matriz de projeção **H** é definida por

$$\boldsymbol{H} = \boldsymbol{W}^{\frac{1}{2}}\boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}^{\frac{1}{2}} \tag{21}$$

Nota-se que **H** depende das variáveis explicativas, da função de ligação e da função de variância, tornando mais difícil a interpretação da medida “leverage”. Pode ser mostrado que

$$V^{-\frac{1}{2}}(\hat{\mu} - u) \cong HV^{-\frac{1}{2}}(Y - \mu)$$

em que $V = diag\{V(\mu_i)\}$. Isso mostra que H mede a influência em unidades estudentizadas de y sobre $\hat{\mu}$.

2.2.1. Pontos de Alavanca

A matriz **H** definida em (21) desempenha um papel importante na análise de resíduos em MLG e é conhecida como matriz de projeção. O elemento h_{ii} desempenha um papel importante nas técnicas de diagnóstico. Supondo que todos os pontos exerçam a mesma influência sobre os valores ajustados, podemos esperar que h_{ii}esteja próximo de $\frac{tr(H)}{n} = \frac{p}{n}$. Então, convém analisar os pontos tais que $h_{ii} > \frac{2p}{n}$, conhecidos como pontos de alavanca ou de alto *leverage*.

2.2.2. Análise de Resíduos

Os resíduos são importantes para detectar observações aberrantes e influentes, que devem ser observadas com maiores detalhes. Segundo a definição dada por COX e SNELL (1968), o resíduo R_i deve expressar uma discrepância entre a observação y_i e o seu valor ajustado $\hat{\mu}_i$ como mostrado em (22):

$$R_i = h_i(y_i, \hat{\mu}_i) \tag{22}$$

em que h_i é uma função de fácil interpretação, usualmente escolhida para estabilizar a variância ou induzir simetria na distribuição amostral de R_i. Em geral, a definição de h_i depende do tipo de anomalia que se deseja estudar. Para maiores detalhes, veja CORDEIRO e DEMÉTRIO (2007, 2010).

A definição dos resíduos através de (22) deve satisfazer, aproximadamente, propriedades tais como $E(R_i) = 0$, $Var(R_i) = constante$ e $Cov(R_i, R_j) = 0, i \neq j$, pois em muitos casos, essas condições são suficientes para especificar a forma da distribuição de R_i. Em geral, a distribuição exata de R_i não é conhecida e, portanto, trabalha-se com resultados assintóticos.

Os resíduos de Pearson é um dos mais simples utilizado nos MLG e é definido por:

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\hat{V}_i^{1/2}} \tag{23}$$

Essa quantidade é um componente da estatística de Pearson generalizada dada em (14). Sua desvantagem é que sua distribuição é bastante assimétrica para modelos não-normais.

Um outro tipo de resíduo bastantes utilizados na prática são os resíduos de Pearson Estudentizados dados por:

$$r_i^{P'} = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)(1 - \hat{h}_{ii})}} \tag{24}$$

em que h_{ii} é o i-ésimo elemento da diagonal definida em (21). Os resíduos estudentizados têm, aproximadamente, variância igual a um quando o parâmetro de dispersão $\phi \to 0$.

O resíduo de Anscombe é definido como uma transformação $N(y_i)$ da observação y_i, escolhida visando que sua distribuição esteja o mais próxima possível da distribuição normal. Foi demonstrado por BARNDORFF-NIELSEN(1978) que no caso dos MLG, $N(\mu) = \int V^{-1/3} d\mu$. Como $N'(\mu)(V/\phi)^{1/2}$ é a aproximação de primeira ordem para o desvio padrão de $N(y)$. Assim, o resíduo de Anscombe pode expresso por

$$A_i = \frac{N(y_i) - N(\hat{\mu}_i)}{N'(\hat{\mu}_i)(\hat{V}_i)^{1/2}} \tag{25}$$

Os resíduos também podem ser definidos como raízes quadradas dos componentes do desvio com sinal dado por $y_i - \hat{\mu}_i$. Tem-se:

$$r_i^D = sinal(y_i - \hat{\mu}_i)\sqrt{2}\{v(y_i) - v(\hat{\mu}_i) + g(\hat{\mu}_i)(\hat{\mu}_i - y_i)\}^{1/2} \tag{26}$$

em que a função $v(x) = xg(x) - b(q(x))$ é definida em termos das funções $b(.)$ e $g(.)$ dadas na Seção 2.1.1 . Esse resíduo representa a distância da observação y_i ao seu valor ajustado $\hat{\mu}_i$, medida na escala do logaritmo da função de verossimilhança. Tem-se, portanto, $D_p = \sum_{i=1}^{n} {r_i^D}^2$, e um valor grande de r_i^D indica que a i-ésima observação é mal ajustada pelo modelo. Os valores de r_i^D podem ser tratados aproximadamente como variáveis aleatórias normais, e ${r_i^D}^2$ comdistribuição aproximadamente x_i^2.

As vantagens da utilização desse resíduo é que o mesmo não requer o conhecimento da função normalizadora, seu calculo é simples depois que o MLG é ajustado e pode ser definido para todas as observações, mesmo as censuradas, desde que essas forneçam alguma contribuição para o logaritmo da função de verossimilhança.

A partir da definição do r_i^D em (26), obtêm-se que os componentes do desvio estudentizadas são obtidos da seguinte maneira:

$$r_i^{D'} = \frac{r_i^D}{\sqrt{1 - \hat{h}_{ii}}}$$

Para modelos bem ajustados, as diferenças entre r_i^D e r_i^P devem ser pequenas. Enquanto para modelos mal ajustados, podem ocorrer grandes diferenças entre

esses valores. Os resíduos com valores muito grandes são chamados de pontos aberrantes.

A influência das observações nas estimativas dos coeficientes no caso do modelo de Regressão Logística que será utilizado nesse trabalho é obtida através da distância de Cook, dada por:

$$Dcook_i = \frac{\hat{h}_{ii}}{(1-\hat{h}_{ii})^2}\frac{(y_i - n_i\hat{\pi}_i)^2}{n_i\hat{\pi}_i(1-\hat{\pi}_i)}$$

Uma melhor avaliação dos resíduos pode ser feita construindo gráficos, tais como:

i) Resíduos *versus* alguma função dos valores ajustados, como $\hat{\eta}_i$. O esperado é que os resíduos variem em torno de zero com amplitude constante. Desvios sistemáticos podem ter algum tipo de curvatura ou mudança sistemática da amplitude com o valor ajustado.

ii) Resíduos *versus* variáveis explicativas não incluídas, para avaliar se existe relação entre eles. O padrão nulo deste gráfico é uma distribuição em torno de zero dos resíduos de amplitude constante.

iii) Resíduos *versus* variáveis explicativas já incluídas, novamente, para avaliar se existe relação. O padrão nulo é o mesmo do gráfico anterior.

iv) Gráfico de índices para localizar observações com resíduo, "*leverage*", Distância de Cook modificada, etc, grandes.

v) Gráfico semi-normal de probabilidade.

2.2.3. Verificação da Função de Ligação

Um método utilizado de maneira informal para verificar a adequação da função de ligação usada é o gráfico da variável dependente ajustada **z** *versus* o preditor linear estimado $\hat{\eta}$. Outro gráfico que pode ser utilizado é o de variável adicionada, onde o padrão nulo indica que a função de ligação usada é adequada.

Quanto aos métodos formais, existem dois para a verificação da adequação da função de ligação:

i) O primeiro consiste em adicionar $u = \hat{\eta}^2$ como uma covariável extra e examinar a mudança ocorrida no desvio, o que é equivalente ao teste de razão de verossimilhanças. Diminuições drásticas são evidências de inadequação da função de ligação. O teste de escore também pode ser utilizado nesse método.

ii) Outro método formal é indexar a família de funções de ligação por um parâmetro λ e fazer o teste de hipótese $H_0: \lambda = \lambda_0$, utilizando os testes de razão de verossimilhança e o escore.

A verificação da adequação da função de ligação também pode ser afetada pela falha em estabelecer escalas corretas para as variáveis explicativas no preditor linear. Em particular, se o teste fosse construído pela adição de $\hat{\eta}^2$ ao preditor linear produz uma redução significativa do desvio do modelo, pode ser um indicativo de uma função de ligação não está adequada. Além disso, pontos atípicos também podem influenciar. Para maiores detalhes, ver CORDEIRO e DEMETRIOS (2007).

2.3. Regressão Logística

Amplamente utilizado na área biomédica, genética, social e no marketing, o modelo de regressão logística é considerado um dos principais métodos de modelagem estatística de dados segundo PAULA (2004). Uma das principais explicações encontradas para sua larga utilização é a facilidade de interpretação dos seus resultados. Esse modelo faz parte dos modelos da classe MLG (Modelos Lineares Generalizados) e é tão utilizado que faz até com que alguns pesquisadores dicotomizem seus resultados.

Nesse trabalho de final de curso, será apresentado um ajuste e uma análise dos resultados de um modelo de regressão logística múltipla para dados com resposta binária, isso é, que possuem dois resultados possíveis comumente chamados de sucesso e fracasso.

2.3.1. Modelo

Considerando a variável resposta binária Y e a covariável X, diz-se que $\pi(x) = P(Y = 1|X = x) = 1 - P(Y = 0|X = x)$ representa a probabilidade do

“sucesso” do evento estudado (Y = 1). Desta forma, o modelo de regressão logística com a função de ligação *logit* é:

$$logit\left\{\frac{\pi(\boldsymbol{x})}{1-\pi(\boldsymbol{x})}\right\} = \beta_1 + \beta_2 x_2 + \cdots + \beta_p x_p \tag{27}$$

em que $\boldsymbol{x} = (1, x_2, \ldots, x_p)^T$contém os valores observados das variáveis explicativas. Os $\beta' s$ são estimados através do método de Newton-Raphson.

Os parâmetros estimados no ajuste do modelo de regressão logística múltipla seguem as mesmas propriedades dos MLG mostradas na seção 2.1.6. Através desses resultados, é possível estudar a significância dos parâmetros e obter intervalos de confiança.

A seleção do melhor ajuste e o estudo da adequação do modelo corrente é feita de maneira análoga aos dos demais modelos da classe MLG.

2.3.2. Razão de Chances

Como dito anteriormente, uma das grandes vantagens do modelo logístico é a facilidade de sua interpretação, que pode ser feita diretamente através dos seus coeficientes. Para exemplificar, sabe-se que de modo geral, a chance de sucesso de um experimento quando X=x é dada por:

$$\frac{\pi(x)}{1-\pi(x)} = e^{\alpha+\beta x} \tag{28}$$

A partir desse resultado, define-se como razão de chances a ocorrência de sucesso para X = x+1 com relação ao valor X=x, ou seja:

$$RC(x+1, x) = \frac{\frac{\pi(x+1)}{1-\pi(x+1)}}{\frac{\pi(x)}{1-\pi(x)}} = \frac{e^{\alpha+\beta(x+1)}}{e^{\alpha+\beta x}} = e^{\beta} \tag{29}$$

Dessa maneira, a interpretação do coeficiente de β pode ser feita da seguinte forma: o aumento de 1 unidade em X aumenta\reduz e^{β} vezes a chance da ocorrência do “fracasso”.

2.4. Modelos Hierárquicos

Em muitos estudos, é comum observar a presença de uma estrutura de hierarquia nos dados. Nesses casos, utilizam-se os modelos hierárquicos, que permitem a modelagem conjunta dos dados considerando seus níveis. Os estudos educacionais são exemplos de áreas que utilizam muito modelos hierárquicos, quando, por exemplo, desejam estudar o desempenho de um aluno considerando sua turma, escola, cidade em estudos de âmbito estadual ou nacional. Pode ser facilmente visto que esses fatores (turma, escola, cidade, etc.) constituem uma estrutura hierárquica, justificando o uso de tal tipo de modelagem.

Os modelos hierárquicos permitem a determinação direta dos níveis na variável resposta estudada. Um modelo estatístico com dois níveis possui o delineamento no caso balanceado como mostrado na Figura 1:

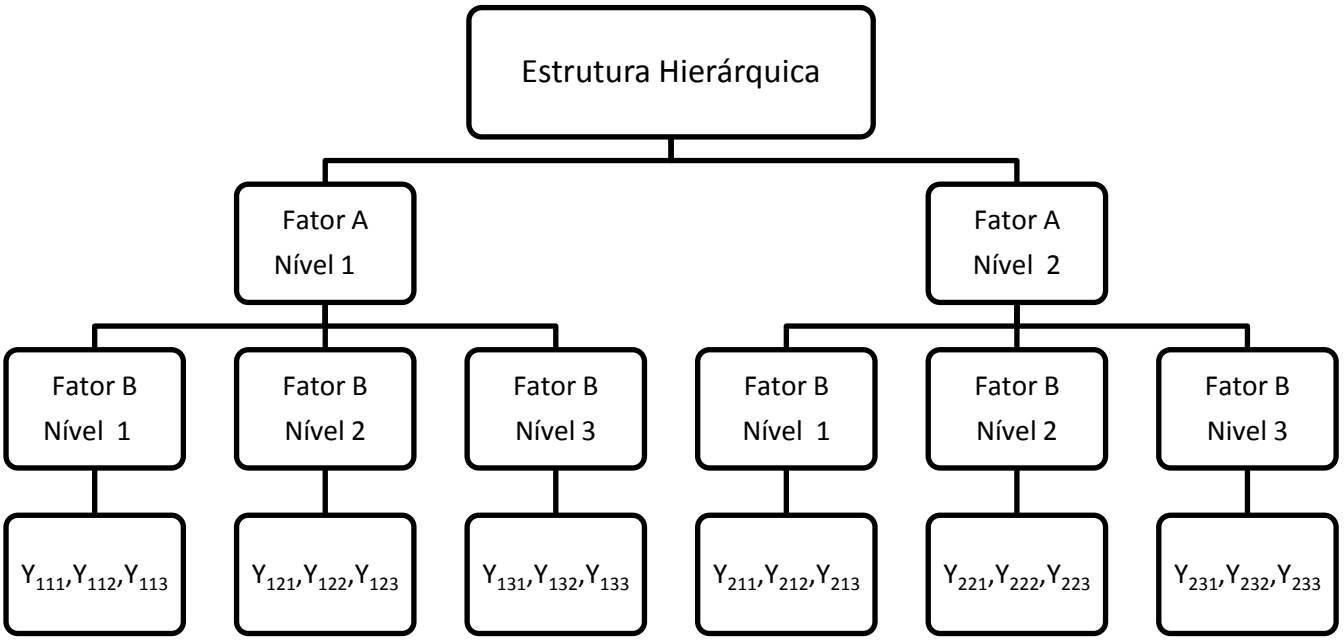

Figura 1: Esquema de um delineamento hierárquico balanceado.

Um modelo linear com dois fatores hierárquicos pode ser escrito da seguinte maneira:

$$y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \epsilon_{(ij)k} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, n \end{cases} \quad (30)$$

em que tem-se a níveis do fator A, b níveis do fator B e n réplicas. O índice $j(i)$ indica que o j-ésimo nível do fator B está abaixo do i-ésimo nível do fator A. Dessa forma, é conveniente pensar que as réplicas são combinações entre os níveis dos fatores A e B. Esse delineamento é considerado balanceado, pois tem-se a mesma quantidade de níveis do fator B para cada nível do fator A e um número de réplicas igual em todos os níveis. Nesse tipo de modelo, não há interação entre os fatores A e B.

Os modelos hierárquicos são aplicados nos MLG da mesma maneira que é aplicado nos modelos lineares: os elementos são agrupados em blocos e a distribuição de probabilidade é aplicada a cada um desses blocos.

Considerando o modelo de regressão logística hierárquica que irá ser ajustado, há uma mudança no cálculo da Razão de Chance. No caso dos subníveis, a Razão de Chance se procede da mesma maneira mostrada na Seção 2.3.2. Mas o cálculo entre os níveis deve ser feito a partir da soma das probabilidades de ocorrência de cada subnível.

Seja $\pi_{j(i)}$ a probabilidade de um elemento do subnível *j* no nível *i* ser aprovado. Tem-se que $\pi_i = \sum_j \pi_{j(i)}$, em que π_i é a probabilidade de um aluno do nível *i* ser aprovado. Dessa forma, pode-se escrever a Razão de Chance da seguinte maneira quando se tem modelos hierárquicos:

$$RC(Nível\ i, Nível\ i') = \frac{\frac{\pi_i}{1-\pi_i}}{\frac{\pi_{i'}}{1-\pi_{i'}}} = \frac{\pi_i}{1-\pi_i} . \frac{1-\pi_{i'}}{\pi_{i'}} \quad (31)$$

em que i' se refere ao nível ao qual a nível i será comparado.

2.5. Contrastes

Através dos contrastes é possível obter um teste com hipóteses específicas para os parâmetros do modelo. As estatísticas são desenvolvidas baseando-se que assintoticamente, os coeficientes seguem uma distribuição normal multivariada com

média igual aos coeficientes do modelo e variância igual a matriz de covariância do modelo.

Os contrastes podem comparar mais de um par de parâmetros a cada teste. Considerando que $\boldsymbol{\theta}$ é um vetor *p*-variado, o contraste pode ser escrito como uma combinação linear da seguinte maneira:

$$\Gamma = \boldsymbol{c\theta} = \sum_{i=1}^{n} c_i \theta_i$$

em que $\boldsymbol{c_{1xp}}$ é o vetor de contrastes com soma zero ($\sum_{i=1}^{n} c_i = 0$). Dessa forma, as hipóteses podem ser reescritas da seguinte maneira:

$$\begin{aligned} H_0 &: \boldsymbol{c\theta} = \sum_{i=1}^{n} c_i \theta_i = a \\ H_1 &: \boldsymbol{c\theta} = \sum_{i=1}^{n} c_i \theta_i \neq a \end{aligned} \tag{32}$$

Seja $\hat{\Gamma} = \boldsymbol{c\hat{\theta}}$. Tem-se, sob a hipótese nula, que $E(\hat{\Gamma}) = a$ e $Var(\hat{\Gamma}) = \boldsymbol{c}Cov(\theta)^{-1}\boldsymbol{c'}$. Dessa forma, a estatística W pode ser escrita como:

$$W = (\boldsymbol{c\theta} - \boldsymbol{a})'(\boldsymbol{c}Cov(\theta)^{-1}\boldsymbol{c'})^{-1}(\boldsymbol{c\theta} - \boldsymbol{a})$$

em que $W \sim x_1^2$.

Esse teste é conhecido como teste de Wald. Caso o valor *p* da estatística de teste seja menor que o nível de significância adotado, rejeita-se a hipótese nula. No caso que será estudado nesse trabalho, as hipóteses testarão a situação onde $a = 0$.

No caso de modelos hierárquicos, os coeficientes do modelo só podem ser comparados entre os elementos pertencentes á um mesmo nível. Isto é, os coeficientes de níveis superiores somente podem ser comparados com outros coeficientes também pertencentes ao nível superior. E os subníveis somente podem ser comparados com subníveis pertencentes ao mesmo nível superior ao qual pertence.

3. Analise Descritiva

A base de dados que será utilizada é referente a aprovação e reprovação dos alunos de 28 cursos de graduação da UFMG na disciplina de Cálculo Diferencial e Integral I entre os anos de 1993 e 2013. Os dados foram disponibilizados pela Pró-Reitoria de Graduação (Prograd) da UFMG.

Parte desses alunos fez a disciplina mais de uma vez, devido a reprovações. Nesses casos, considerou-se somente a primeira vez em que o aluno cursou a disciplina, para ter uma melhor base comparativa com os demais estudantes.

Muitos outros cursos além dos 28 selecionados tiveram alunos que fizeram a disciplina de Cálculo Diferencial e Integral I ao longo do período estudado. Mas para evitar trabalhar com cursos com uma quantidade pequena de alunos, foi definido o critério que somente seriam estudados os cursos onde pelo menos 50 alunos fizeram a disciplina. Os cursos foram agrupados em três grupos, de acordo com a instituição ao qual pertence. Nesse sentido, o grupo da Escola de Engenharia possui 13 cursos, o grupo do Instituto de Ciências Exatas (ICEX) possui 12 cursos. Os 2 cursos da Faculdade de Ciências Econômicas (FACE) e o curso do Instituto Geociências (IGC) presentes no estudo foram agrupados em um único grupo denominado Outros.

Na Tabela 4 é mostrada a quantidade de alunos de cada um dos cursos estudados. A partir da tabela, observa-se que do total de 23705 estudantes da UFMG, o curso com maior quantidade de alunos é a Engenharia Civil com 3892 alunos (16,42%), seguido do curso de Engenharia Elétrica com 1872 (7,90%). O curso com a menor quantidade de alunos é Química Tecnológica, com apenas 147 (0,62%). Esse curso que foi criado em 2010 e por isso ainda conta com uma quantidade pequena de alunos que fizeram a disciplina.

Entre as Instituições de Ensino, a Engenharia é a com maior número de alunos, com 14310 (60,37%) do total seguido pelo Instituto de Ciências Exatas (ICEX) com 8109 alunos (34,21%).

Tabela 4: Instituições e cursos dos alunos presentes na base de dados.

Instituição	Curso	Total	
		Freq.	%
ICEX	Matemática Diurno	1006	4,24%
	Ciência da Computação	1507	6,36%
	Ciências Atuariais	295	1,24%
	Estatística	690	2,91%
	Física Diurno	902	3,81%
	Física Noturno	648	2,73%
	Matemática Computacional	265	1,12%
	Matemática Noturno	683	2,88%
	Química Diurno	845	3,56%
	Química Noturno	674	2,84%
	Química Tecnológica	147	0,62%
	Sistemas de Informação	447	1,89%
	Total	**8109**	**34,21%**
E. Engenharia	Engenharia Aeroespacial	233	0,98%
	Engenharia Ambiental	230	0,97%
	Engenharia Civil	3892	16,42%
	Engenharia de Controle e Automação Diurno	1102	4,65%
	Engenharia de Controle e Automação Noturno	251	1,06%
	Engenharia de Minas	954	4,02%
	Engenharia de Produção	1005	4,24%
	Engenharia de Sistemas	183	0,77%
	Engenharia Elétrica	1872	7,90%
	Engenharia Mecânica Diurno	1568	6,61%
	Engenharia Mecânica Noturno	979	4,13%
	Engenharia Metalúrgica	968	4,08%
	Engenharia Química	1073	4,53%
	Total	**14310**	**60,37%**
Outros	Ciências Econômicas	451	1,90%
	Controladoria e Finanças	199	0,84%
	Geologia	636	2,68%
	Total	**1286**	**5,43%**
TOTAL		**23705**	**100%**

Na Figura 2 é mostrada a distribuição dos alunos ao longo dos anos de 1993 e 2013. Observa-se uma tendência crescente de alunos que fizeram a disciplina de Cálculo Integral e Diferencial I, sendo que nos últimos quatro anos a quantidade de alunos é o dobro da quantidade de alunos nos 5 primeiros anos. Esse aumento do número de vagas se deve ao REUNI, um programa de expansão da educação superior, com o objetivo de reestruturar e expandir as universidades federais.

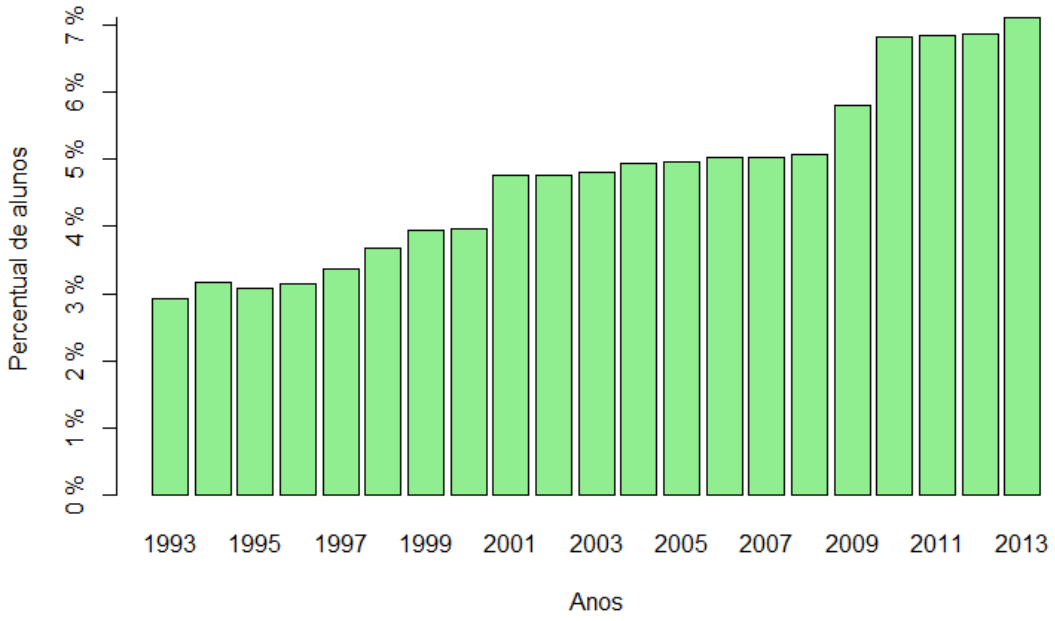

Figura 2: Distribuição dos alunos ao longo dos anos.

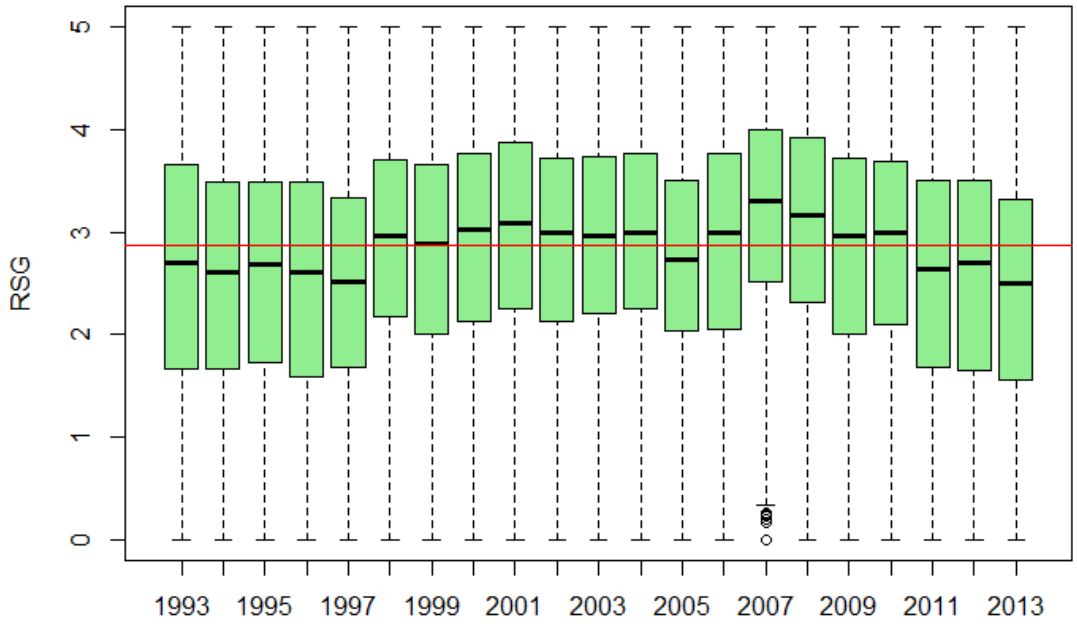

Figura 3: Boxplot entre o RSG do período em que foi feita a disciplina Cálculo Integral de Diferencial I *versus* o ano em que foi feita a disciplina.

Na Figura 3 é mostrado um *boxplot* com o RSG do período em que foi feita a disciplina de cálculo *versus* o ano em que foi feita a disciplina e em vermelho é mostrado o RSG mediano de todos os alunos. Todas as medianas oscilam entre os valores 3.3 e 2.5, mas observa-se uma diminuição dessa medida nos últimos anos em comparação com os demais, sendo que o ano de 2013 foi o menor RSG mediano. Considerando todos os anos conjuntamente, o menor valor observado de RSG foi 0 e o maior 5. Do total de alunos, 75% tiveram RSG igual ou maior que 2.00, e 25% igual ou maior que 3.66, sendo o valor mediano igual à 2.87. Nota-se também que nos três últimos anos e nos cinco primeiros, a mediana dos anos é menor que a geral.

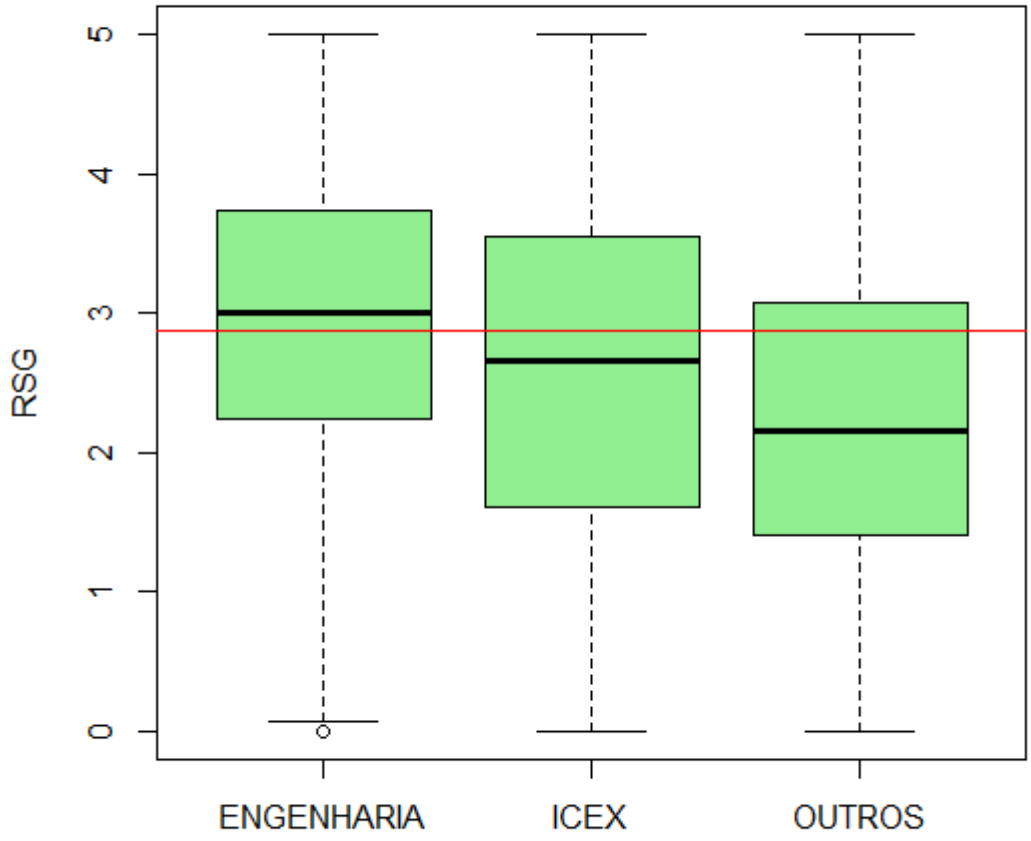

Figura 4: RSG *versus* a Instituição de Ensino.

A Figura 4 mostra os *boxplots* do RSG dos alunos do estudo no período em que foi feita pela primeira vez a disciplina de Cálculo Diferencial e Integral I *versus* a

Instituição de Ensino e a linha em vermelho representa o valor da mediana geral. Observa-se que a Escola de Engenharia é a instituição com maior RSG mediano, seguido pelo Instituto de Ciências Exatas e seguido pelos Outros, que engloba cursos da FACE e do Instituto de Geociências (IGC). A única instituição com RSG mediano superior ao geral é a Escola de Engenharia.

Tabela 5: Coeficiente de Variação e porcentagem de aprovação e reprovação por instituição.

Instituição	CV (RSG)	Reprovações		Aprovações		Total
		Freq.	%	Freq.	%	
E. Engenharia	37,93	3382	23,63%	10928	76,37%	14310
ICEX	51,77	3049	37,60%	5060	62,40%	8109
Outros	53,91	617	47,98%	669	52,02%	1286

Analisando a A Figura 4 mostra os *boxplots* do RSG dos alunos do estudo no período em que foi feita pela primeira vez a disciplina de Cálculo Diferencial e Integral I *versus* a Instituição de Ensino e a linha em vermelho representa o valor da mediana geral. Observa-se que a Escola de Engenharia é a instituição com maior RSG mediano, seguido pelo Instituto de Ciências Exatas e seguido pelos Outros, que engloba cursos da FACE e do Instituto de Geociências (IGC). A única instituição com RSG mediano superior ao geral é a Escola de Engenharia.

Tabela 5, observa-se que a E. Engenharia tem o menor coeficiente de variação que as demais instituições, ou seja, possui uma maior homogeneidade quanto ao RSG. Em seguida, tem-se a quantidade de reprovações e aprovações por instituição. A instituição Outros é que a possui maior percentual de reprovação, com 47,98% de alunos reprovados. Já a E. Engenharia possui o maior percentual de aprovações, com 76,37% de seus alunos aprovados.

4. Ajuste do Modelo

O modelo que será ajustado nessa seção é um modelo de regressão logística hierárquica desbalanceada, pois a variável em estudo que é a aprovação\reprovação dos alunos na disciplina de Cálculo I é binária (1 = Aprovação; 0 = Reprovação) e existe uma estrutura hierárquica entre os cursos e a instituição, que são variáveis explicativas do modelo.

O modelo que será estimada é da seguinte forma:

$$logit\left\{\frac{\pi(\boldsymbol{x})}{1-\pi(\boldsymbol{x})}\right\} = \beta_0 + \beta_i x_1 + \beta_{1(i)} x_2 + \cdots + \beta_{j\,(i)} x_p$$

em que $\beta_{j\,(i)}$ se refere ao j-ésimo curso da i-ésima instituição.

A primeira dificuldade encontrada foi conseguir realizar o ajuste corretamente em algum *software* estatístico, pois como cada curso existe somente em uma instituição, o modelo hierárquico é desbalanceado.

No ajuste de um modelo hierárquico, os *softwares* esperam que no segundo nível os nomes se repitam. Isto é, se no nível superior ‘A’ tem-se os subníveis ‘a’, ‘b’, e ‘c’, os *softwares* esperam que em outro nível superior ‘B’, os subníveis sejam ‘a’, ‘b’ e ‘c’, mesmo que nem todos estejam presentes, e não uma segunda situação onde os subníveis são ‘d’ e ‘e’. Como os cursos não se repetem para mais de uma instituição, estamos nessa segunda situação, e, portanto, para ajustar o modelo tanto no SAS quanto no R, serão necessárias algumas mudanças nos dados. Além disso, também foram encontrados problemas para realizar as comparações múltiplas entre os coeficientes e no cálculo da razão de chances entre os níveis superiores. As soluções obtidas para todas essas questões serão exibidas nas próximas seções.

4.1. Preparação dos dados

Para que os *softwares* entendam a estrutura hierárquica dos dados, foi necessário em cada instituição codificar os nomes dos cursos, para que sejam os mesmos em todas as instituições. Por exemplo, após essa codificação, o ICEX, a E.

Engenharia e o Outros possuem um subnível com o nome ‘A’, mas no primeiro ele representa o curso de Matemática Diurno, no segundo Engenharia Aeroespacial e no terceiro Ciências Econômicas.

Com essa codificação os *softwares* passam a realizar corretamente o ajuste hierárquico e não há nenhuma perda de informação ou interferência no modelo por alguns cursos terem o mesmo nome. Na Tabela 6 é mostrada a codificação que cada curso recebeu.

Tabela 6: Codificação dos cursos em cada instituição.

Instituição	Curso	Codificação	Instituição	Curso	Codificação
ICEX	Matemática Diurno	A	E. Engenharia	Engenharia Aeroespacial	A
	Ciência da Computação	B		Engenharia Ambiental	B
	Ciências Atuariais	C		Engenharia Civil	C
	Estatística	D		Engenharia de Controle e Automação Diurno	D
	Física Diurno	E			
	Física Noturno	F		Engenharia de Controle e Automação Noturno	E
	Matemática Computacional	G			
	Matemática Noturno	H		Engenharia de Minas	F
	Química Diurno	I		Engenharia de Produção	G
	Química Noturno	J		Engenharia de Sistemas	H
	Química Tecnológica	K		Engenharia Elétrica	I
	Sistemas de Informação	L		Engenharia Mecânica Diurno	J
Outros	Ciências Econômicas	A		Engenharia Mecânica Noturno	K
	Controladoria e Finanças	B		Engenharia Metalúrgica	L
	Geologia	C		Engenharia Química	M

Mesmo após a codificação, os dados apresentam um desbalanceamento, pois a quantidade de subníveis varia de instituição para instituição. Mas esse desbalanceamento não impede o ajuste de ser feito, diferentemente de quando não tinha as codificações para os cursos. Para entender o que os *softwares* fazem nesses casos de desbalanceamento, lembre que a Escola de Engenharia possui 13 subníveis (13 cursos), o ICEX 12 e a instituição denominada como Outros (FACE e IGC) que possui somente 3. Quando o SAS ou o R ajusta o modelo, ele cruza as instituições com todos os subníveis existentes, no caso 13 (lembre da codificação mostrada na Tabela 6). No caso da Engenharia, todos esses 13 coeficientes estarão preenchidos, já ele possui exatamente 13 cursos. O mesmo não ocorre com o ICEX

e os Outros, pois eles possuem um número menor de cursos, sendo que no primeiro haverá 12 coeficientes preenchidos e no segundo 3.

O SAS e o R divergem na forma com que cada um trata as caselas vazias dos cruzamentos que não existem na prática (como o cruzamento entre o curso M e a instituição Outros) e que são exibidos nas saídas computacionais. O SAS atribui 0 ou um ponto às caselas vazias, enquanto o R coloca NA. Nesse sentido, a saída do SAS é mais intuitiva do que a do R, pois os NA's passam a impressão que o modelo não foi bem ajustado ou que há algum erro.

4.2. Modelo

Após a preparação dos dados e o esclarecimento de como o SAS e o R lidam com o desbalanceamento, será mostrado nessa seção quais os comandos necessários para o ajuste e os resultados obtidos.

Os comandos para o ajuste no SAS estão mostrados na Figura 5. A *procedure* utilizada foi a *genmod*. O 'class' indica dentre as variáveis explicativas, quais são fatores, e o comando entre parênteses altera o nível de referência da variável. Dentre as instituições, o ICEX será o nível de referênciae dentre os cursos, o 'A' será a referência. Essa alteração é feita no SAS e também será feita no R, para que os resultados possam ser comparados. O comando 'model' faz o ajuste do modelo. Observa-se o indicativo que o modelo será hierárquico é a variável instituição ('*inst*') estar entre parênteses ao lado da variável curso, que é o subnível. As demais variáveis entram normalmente no modelo, assim como a variável instituição. Por fim, o 'link=logit', indica que o modelo ao qual os dados serão ajustados será o logístico. Nessa figura ainda tem-se os comandos para fazer o contraste, mas os comentários sobre eles serão feitos posteriormente.

```
proc genmod DATA=dados;
        class inst(ref="ICEX" param=ref) curso(ref="A" param=ref);
        model ap = inst curso(inst) rsg/ dist = bin
                                        link = logit;

        estimate 'ICEX-ENGENHARIA'      inst -1  0  1 ;
        estimate 'ICEX-OUTROS'          inst  0 -1  1 ;
        estimate 'ENGENHARIA-OUTROS'    inst  1 -1  0 ;
        output out=out ;
run;
```

Figura 5: Comandos para o ajuste de um modelo logístico hierárquico no SAS.

Os comandos para fazer o ajuste do modelo de Regressão Logístico Hierárquico R são mostrados na Figura 6. O comando ‘relevel’ é utilizado para mudar o nível de referência das variáveis explicativas que são fatores. Assim como no ajuste feito no SAS, o ICEX e o curso ‘A’ serão as referências. O comando ‘glm’ faz o ajuste do modelo e a estrutura hierárquica é indicada pela através da barra ‘ / ’, e as demais variáveis explicativas entraram normalmente no modelo. Observa-se que diferentemente do SAS, no R a variável instituição só é escrita uma vez no modelo. O comando ‘family = binomial(link=‘logit’))’ indica que o tipo de modelo que será ajustado é o logístico.

```
dados$curso <- relevel(dados$curso, ref="A")
dados$inst  <- relevel(dados$inst, ref="ICEX")
m1 = glm(ap ~ inst/curso+rsg,data = dados,
        family = binomial(link = 'logit'))
```

Figura 6: Comandos para o ajuste de um modelo logístico hierárquico no R.

O ajuste realizado nos dois *softwares* através dos comandos mostrados nas Figura 5 e Figura 6 geram resultados iguais, como o esperado. Na Tabela 7 são mostrados os valores estimados retirados da saída do R.

A partir dos valores estimados para os coeficientes do modelo, é possível fazer comparações com o nível de referência para cada variável. Caso o valor estimado de algum parâmetro seja positivo\negativo e significativo, conclui-se que esse parâmetro, dentro do contexto estudado, possui mais\menos probabilidade de serem aprovados em Cálculo I que o nível de referência. Considerando as instituições, tem-se que o coeficiente da Engenharia é 2,548 (valor p = 0,000), de onde pode-se inferir que os alunos da Escola de Engenharia tem maior probabilidade de serem aprovados em Cálculo I do que os alunos do ICEX. Já o coeficiente de Outros é negativo e vale - 0,535 (valor p = 0,008), portanto, os alunos de outras instituições tem menor probabilidade de serem aprovados que os alunos do ICEX.

Os cursos do ICEX, por exemplo, serão comparados com o curso de Matemática Diurno, que na codificação foi o ‘A’ tomado como referência. Nota-se que apenas os cursos de Estatística, Química Diurno e Química Tecnológica foram significativos e negativos, ou seja, tem menor probabilidade de serem aprovados que o curso de Matemática Diurno. O curso de Sistemas de Informação também é

negativo, porém, não é significativo. Nesse caso, afirma-se somente que há indícios que a probabilidade de aprovação em cálculo I dos alunos de Sistemas de Informação seja menor que o dos alunos do nível de referência. A mesma interpretação pode ser feita para os cursos das demais instituições. O coeficiente do RSG é positivo e significativamente diferente de zero, e no seu caso, a interpretação que se faz é que quanto maior o RSG, maior a probabilidade de aprovação do aluno em Cálculo I.

Tabela 7: Ajuste do modelo no R.

Coeficiente	**Estimativa**	**Erro Padrão**	**Valor z**	**Valor *p***
Intercepto	-7,026	0,168	-41,914	0,000 ***
Engenharia	2,548	0,329	7,737	0,000 ***
Outros	-0,535	0,202	-2,646	0,008 **
RSG	3,135	0,046	68,008	0,000 ***
ICEX:Ciência da Computação	1,126	0,167	6,734	0,000 ***
ICEX:Ciência Atuariais	0,906	0,240	3,770	0,000 ***
ICEX:Estatística	-0,758	0,182	-4,159	0,000 ***
ICEX:Física D	0,665	0,180	3,694	0,000 ***
ICEX:Física N	1,323	0,199	6,665	0,000 ***
ICEX:Matemática Computacional	0,432	0,271	1,598	0,110
ICEX:Matemática N	1,655	0,198	8,374	0,000 ***
ICEX:Química D	-0,326	0,178	-1,834	0,067 .
ICEX:Química N	0,103	0,181	0,572	0,568
ICEX:Química Tecnológica	-0,614	0,269	-2,287	0,022 *
ICEX:Sistemas de Informação	-0,147	0,204	-0,721	0,471
E.Engenharia:E. Ambiental	-4,164	0,359	-11,609	0,000 ***
E.Engenharia:E. Civil	-2,098	0,308	-6,817	0,000 ***
E.Engenharia:E. Controle e A. D	-1,011	0,334	-3,032	0,002 **
E.Engenharia:E. Controle e A. N	-0,815	0,399	-2,043	0,041 *
E.Engenharia:E. Minas	-1,944	0,322	-6,035	0,000 ***
E.Engenharia:E. Produção	-2,295	0,324	-7,076	0,000 ***
E.Engenharia:E. Sistemas	-0,452	0,424	-1,067	0,286
E.Engenharia:E. Elétrica	-2,168	0,319	-6,793	0,000 ***
E.Engenharia:E. Mecânica D	-1,797	0,316	-5,680	0,000 ***
E.Engenharia:E. Mecânica N	-1,508	0,327	-4,614	0,000 ***
E.Engenharia:E. Metalúrgica	-2,367	0,323	-7,333	0,000 ***
E.Engenharia:E. Química	-2,672	0,341	-7,847	0,000 ***
Outros:Controladoria e Finanças	1,082	0,268	4,044	0,000 ***
Outros:Geologia	1,425	0,202	7,048	0,000 ***

O valor do *deviance* do modelo foi 11314 com 23635 graus de liberdade com um valor *p* igual à 1. Observa-se que apenas 5 coeficientes referentes a cursos não foram significativamente diferentes de 0.

4.3. Análise de Diagnóstico

Nessa seção será analisando a qualidade do ajuste do modelo logístico hierárquico obtido na Seção 4.2. Essa análise será feita graficamente, a partir dos gráficos de diagnósticos presentes na Figura 7.

Pode-se observar que os gráficos indicam a presença de pontos de alavancas, influentes e aberrantes, além de um possível problema na função de ligação, indicando que o modelo pode não estar bem ajustado. Entretanto, deve-se levar em consideração o fato da base de dados ser muito grande (23705 observações), o que gera uma quantidade maior de pontos considerados de alavanca, influentes e aberrantes e que o valor *p* da estatística do *deviance* foi igual a 1.

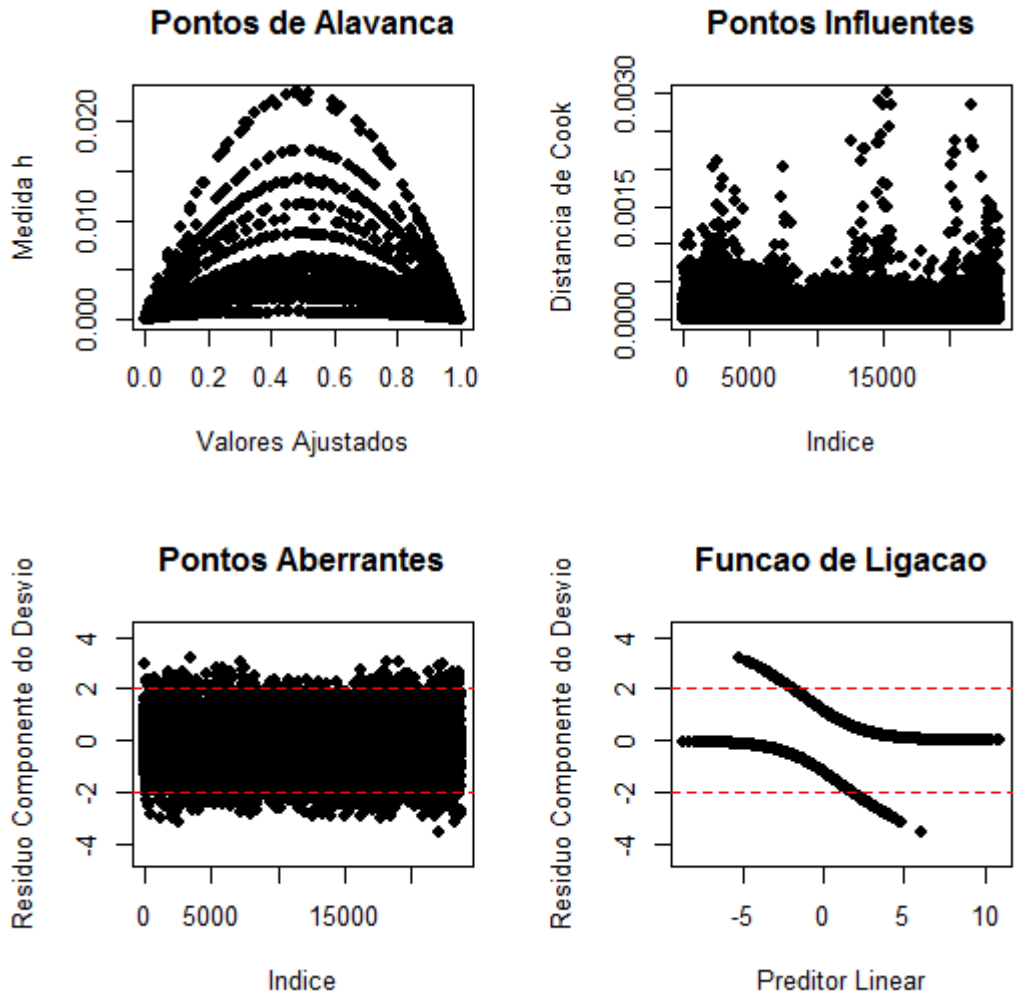

Figura 7: Gráficos de diagnóstico.

Como o objetivo principal do trabalho era estudar a utilização de modelos hierárquicos em *softwares*, decidiu-se manter o modelo ajustado, por ser o mais parcimonioso e gerar resultados razoáveis. Além disso, para obter um modelo que se ajuste melhor aos dados, seria necessário obter um maior número de covariáveis explicativas do modelo, atualmente não disponíveis, e estudar melhor a análise de resíduos em grandes bases de dados.

Porém, mantendo esse ajuste, deve-se ressaltar que as conclusões e interpretações feitas a partir de seus resultados devem ser feitas com cuidado. Serão feitas a seguir, a título de ilustração, os resultados obtidos na análise de contrastes e na razão de chances. Acredita-se que será de grande proveito exibir como essas análises são feitas no caso dos modelos logísticos hierárquicos.

4.4. Contrastes

O contraste permite fazer comparações entre os coeficientes do modelo, seja dois a dois, ou entre um número maior de coeficientes. No caso do modelo hierárquico desbalanceado em estudo, o nível superior pode ser comparado entre si, mas os subníveis só podem ser comparados dentro de um mesmo nível, pois subníveis de níveis diferentes representam cursos diferentes, e, portanto, não faz sentido comparações somente entre os subníveis desconsiderando o nível ao qual ele pertence. Da mesma forma, não faz sentido comparar níveis com subníveis (instituições com cursos).

No SAS, como mostrado na Figura 5, é relativamente fácil obter essas comparações entre as instituições, porém, não foi encontrada durante o período de estudo nenhuma maneira de obter as comparações entre os cursos por instituições.

No R existem algumas funções em pacotes que realizam as comparações entre os coeficientes de um modelo, inclusive de modelos hierárquicos, porém não funcionam corretamente devido ao fato do modelo ser desbalanceado. Foi encontrada somente uma função chamada ‘*wald.test’* presente no pacote ‘aod’ que realiza corretamente a análise de contraste. Considerando o estudo em questão, para comparar as instituições, selecionam-se somente os coeficientes referentes às instituições e o intercepto; da matriz de covariância do modelo, são selecionadas as linhas e colunas referentes aos coeficiente selecionados para serem comparados.

Outro parâmetro que deve ser passado é o vetor de contraste, sendo que essa passagem é feita de maneira diferente no SAS e no R. No primeiro, o nível de referência fica na última coluna da matriz de contrastes e deve ser preenchido com o número 1, e os demais parâmetros que serão comparados à ele recebem o -1.Quando a comparação de parâmetros não o envolve, ele recebe o valor 0. Já no R, o nível de referência fica na primeira coluna da matriz de contraste e é sempre 0. Quando se deseja comparar um curso com a referência, esse número recebe o valor 1, quando deseja-se comparar outros dois parâmetros entre si, um recebe o +1, outroo -1, e o nível de referência permanece com o valor 0.

```
contraste = cbind(0,1,0)

wald.test(b = coef, Sigma = cov, L =contraste,
                                        ncol = length(coef))
```

Figura 8: Código do R para a realização de contrastes.

A Figura 8 mostra o código do R para fazer o contraste entre o ICEX e a E. Engenharia. Na primeira linha tem o vetor de contraste. De acordo com a parametrização do R, nesse vetor a primeira coluna é referente ao ICEX, a segunda à E. Engenharia e a terceira a Outros. Como desejamos comparar a E. Engenharia ao ICEX, colocamos 1 na posição referente à E. Engenharia e 0 nas demais. Em “b = coef”, entra-se com o vetor de coeficientes que serão comparados, no caso, os coeficientes das instituições. No parâmetro “Sigma = cov”, é colocada a matriz de covariância dos coeficientes selecionados, em “L” entra-se com os contrastes e em “ncol” com o comprimento do vetor de coeficientes.

Os resultados apresentados nos dois softwares para as comparações entre as instituições foram idênticos. Abaixo na Tabela 8 será mostrado o resultado retirado do SAS. Observa-se que ao nível de 5% de significância, os coeficientes de todas as comparações diferem significativamente.

Tabela 8: Contrastes entre instituições.

Comparação	Estatística x^2	Valor *p*	Contraste
ICEX -- E. Engenharia	59,86	<,0001	0 , 1, 0
ICEX -- Outros	7,00	0,0081	0, 0, 1
E. Engenharia -- Outros	81,42	<,0001	0, 1, -1

Os contrastes entre cursos foram feitos somente no R e a função para seu cálculo encontra-se em anexo. Considerando os cursos do ICEX, os resultados são mostrados na Tabela 9.

Tabela 9: Contrastes dos cursos do ICEX.

Cursos	Estatística x^2	Valor *p*		Cursos	Estatística x^2	Valor *p*	
Ciência da Comp.--Matemática D.	45,35	0,00	***	Estatística--Matemática N.	147,05	0,00	***
Ciências Atuariais--Matemática D.	14,21	0,00	***	Estatística--Química D.	5,90	0,01	*
Estatística--Matemática D.	17,29	0,00	***	Estatística--Química N.	22,54	0,00	***
Física D.--Matemática D.	13,64	0,00	***	Estatística--Química Tec.	0,28	0,59	
Física N.--Matemática D.	44,41	0,00	***	Estatística--Sistemas de Inf.	8,92	0,00	**
Matemática Comp.--Matemática D.	2,55	0,11		Física D.--Física N.	11,29	0,00	***
Matemática N.--Matemática D.	70,12	0,00	***	Física D.--Matemática Comp.	0,74	0,38	
Química D.--Matemática D.	3,36	0,06	.	Física D.--Matemática N.	25,89	0,00	***
Química N.--Matemática D.	0,32	0,56		Física D.--Química D.	31,58	0,00	***
Química Tec.--Matemática D.	5,22	0,02	*	Física D.--Química N.	9,88	0,00	**
Sistemas de Inf.--Matemática D.	0,52	0,47		Física D.--Química Tec.	22,82	0,00	***
Ciência da Comp.--Ciências Atuariais	0,92	0,33		Física D.--Sistemas de Inf.	16,13	0,00	***
Ciência da Comp.—Estatística	125,34	0,00	***	Física N.--Matemática Comp.	10,00	0,00	**
Ciência da Comp.--Física D.	7,84	0,00	**	Física N.--Matemática N.	2,48	0,11	
Ciência da Comp.--Física N.	1,13	0,28		Física N.--Química D.	71,27	0,00	***
Ciência da Comp.--Matemática Comp.	7,08	0,00	**	Física N.--Química N.	38,28	0,00	***
Ciência da Comp.--Matemática N.	8,34	0,00	**	Física N.--Química Tec.	47,62	0,00	***
Ciência da Comp.--Química D.	79,22	0,00	***	Física N.--Sistemas de Inf.	45,21	0,00	***
Ciência da Comp.--Química N.	38,05	0,00	***	Matemática Comp.--Matemática N.	18,95	0,00	***
Ciência da Comp.--Química Tec.	45,05	0,00	***	Matemática Comp.--Química D.	8,00	0,00	**
Ciência da Comp.--Sistemas de Inf.	44,40	0,00	***	Matemática Comp.--Química N.	1,48	0,22	
Ciências Atuariais – Estatística	47,58	0,00	***	Matemática Comp.--Química Tec.	9,74	0,00	**
Ciências Atuariais--Física D.	1,02	0,31		Matemática Comp.--Sistemas de Inf.	4,10	0,04	*
Ciências Atuariais--Física N.	2,73	0,09	.	Matemática N.--Química D.	103,66	0,00	***
Ciências Atuariais--Matemática Comp.	2,29	0,13		Matemática N.--Química N.	62,59	0,00	***
Ciências Atuariais--Matemática N.	8,90	0,00	**	Matemática N.--Química Tec.	65,58	0,00	***
Ciências Atuariais--Química D.	26,86	0,00	***	Matemática N.--Sistemas de Inf.	68,50	0,00	***
Ciências Atuariais--Química N.	11,25	0,00	***	Química D.--Química N.	5,89	0,01	*
Ciências Atuariais--Química Tec.	23,79	0,00	***	Química D.--Química Tec.	1,17	0,27	
Ciências Atuariais--Sistemas de Inf.	16,75	0,00	***	Química D.--Sistemas de Inf.	0,79	0,37	
Estatística--Física D.	61,75	0,00	***	Química N.--Química Tec.	7,16	0,00	**
Estatística--Física N.	108,62	0,00	***	Química N.--Sistemas de Inf.	1,52	0,21	
Estatística--Matemática Comp.	19,28	0,00	***	Química Tec.--Sistemas de Inf.	2,70	0,10	

Observa-se que das 64 comparações, apenas 17 foram não significativas, ou seja, indicaram não haver diferenças entre os parâmetros estudados.

Em anexo, encontram-se as comparações entre os parâmetros das demais instituições. Observa-se que das 78 comparações dois a dois entre os cursos da E. Engenharia, somente 14 não foram significativas e 64 foram significativamente diferentes. Considerando a FACE e o IGC, o curso de Ciências Econômicas difere significativamente do curso de Controladoria e Finanças e de Geologia, mas esses não diferem entre si.

4.5. Razão de Chances

A Razão de Chances entre os institutos foi calculada como mostrado na Seção 2.4 e os resultados obtidos são mostrados na Tabela 10. Observa-se que, considerando-se a disciplina de Cálculo Diferencial e Integral I, a chance de um aluno pertencente a E. Engenharia ser aprovado é 6,692 vezes maior que a de um aluno do ICEX ser aprovado. A mesma interpretação pode ser feita considerando as demais razões de chances, isto é, a chance de um aluno da FACE e do IGC ser aprovado (reprovado) é 0,169 (5,917) vezes a de um aluno do ICEX e a chance de um aluno da E. Engenharia ser aprovado é 39,688 vezes a de um aluno da FACE ou IGC.

Tabela 10: Razão de Chances dos níveis principais.

Coeficiente	Instituto de Referência	Razão de Chances
E. Engenharia	ICEX	6,692
FACE e IGC	ICEX	0,169
E. Engenharia	FACE e IGC	39,688

Para os subníveis, a Razão de Chances referente a chance de ser aprovado na disciplina de Cálculo Diferencial e Integral I é calculada da maneira usual, como mostrado na Seção 2.3.2 e os resultados são mostrados na Tabela 11. Os subníveis da E. Engenharia são comparados ao curso de Eng. Aeroespacial, e observa-se que a chance de todos os alunos dos demais cursos serem aprovados é menor que a dos alunos do curso de Eng. Aeroespacial. Um exemplo é o curso de Eng. De Produção, cuja chance de ser aprovado (reprovado) é 0,101 (9,900) vezes dos alunos do curso de Eng. Aeroespacial.

Considerando os cursos do ICEX, observa-se que a chance de um aluno da Ciência da Computação ser aprovado é 3,084 vezes a de um aluno da Matemática Diurno, e a chance de um aluno da Estatística ser aprovado (reprovado) é 0,468 (2,136) vezes a de um aluno da Matemática Diurno.

Entre os cursos pertencentes a FACE e ao IGC agrupados na categoria Outros, observa-se que a chance de um aluno do curso de Controladoria e Finanças ser aprovado é 2,951 vezes a de um aluno da Ciências Econômicas.

Tabela 11: Razão de Chances dos subníveis.

Coeficiente	Curso de Referência	Razão de Chance
E.Engenharia:E. Ambiental	Engenharia Aeroespacial	0,016
E.Engenharia:E. Civil		0,123
E.Engenharia:E. Controle e A. D		0,364
E.Engenharia:E. Controle e A. N		0,443
E.Engenharia:E. Minas		0,143
E.Engenharia:E. Produção		0,101
E.Engenharia:E. Sistemas		0,636
E.Engenharia:E. Elétrica		0,114
E.Engenharia:E. Mecânica D		0,166
E.Engenharia:E. Mecânica N		0,221
E.Engenharia:E. Metalúrgica		0,094
E.Engenharia:E. Química		0,069
ICEX:Ciência da Computação	Matemática Diurno	3,084
ICEX:Ciência Atuariais		2,475
ICEX:Estatística		0,468
ICEX:Física D		1,945
ICEX:Física N		3,754
ICEX:Matemática Computacional		1,541
ICEX:Matemática N		5,233
ICEX:Química D		0,722
ICEX:Química N		1,109
ICEX:Química Tecnológica		0,541
ICEX:Sistemas de Informação		0,863
Outros:Controladoria e Finanças	Ciências Econômicas	2,951
Outros:Geologia		4,159

5. Conclusão

Os *softwares* R e SAS fazem de maneira correta o ajuste dos modelos hierárquicos desbalanceados, mas ambos impõem algumas dificuldades. Primeiro, para que os programas utilizados entendam de maneira correta a estrutura hierárquica dos dados desbalanceados, é necessário fazer a recodificação dos nomes dos subníveis. Ainda assim, a saída apresentada pelos dois *softwares* faz o cruzamento entre os níveis e todos subníveis, mesmo que na prática devido ao desbalanceamento alguns subníveis não existam. O SAS, nesses casos, adota '0' e '.' para indicar que não possui informação nesses casos, e o R utiliza 'NA'. Neste caso, a saída do SAS é melhor que a do R, pois a do R passa a impressão de mau ajuste ou erro, quando está correto.

Outra análise que foi feita com certa dificuldade nos dois *softwares* foi a análise de comparações múltiplas através de contrastes. No SAS o contraste é feito no momento do ajuste do modelo, sendo relativamente fácil fazer o contraste entre os níveis. Porém, não foi encontrada em nenhum momento ao longo da pesquisa uma forma de se fazer o contraste entre os subníveis, respeitando os níveis ao qual esses subníveis pertencem. No R, as funções normalmente utilizadas para fazer contrastes, como a "*glht*", não funcionam devido a forma que o R apresenta sua saída, com 'NA's. Entretanto, utilizando-se a função '*wald.teste',* do pacote 'aod', é possível realizar a análise de contrastes tanto para os níveis quando para os subníveis. Nessa função, passam-se os coeficientes e a matriz de covariância que serão comparados e o vetor contraste.

A razão de chances entre os subníveis é calculada da maneira usual, como mostrada na Seção 2.3.2, porém, entre os níveis, cálculos adicionais mostrados na Seção 2.4 fazem necessários. Essa análise foi feita somente no R, por ser mais fácil de programar cálculos como esses que o SAS.

Considerando a aplicação utilizada, a análise descritiva mostrou um aumento de alunos que fizera a disciplina de Cálculo Diferencial e Integral I, e indicou que a Escola de Engenharia possui um maior percentual de aprovações nessa disciplina

que o Instituto de Ciências Exatas (ICEx) e que os Outros, composto por cursos pertencentes à Faculdade de Ciências Econômicas (FACE) e pelo Instituto de Geociências (IGC). O ajuste do modelo foi feito e sua análise de diagnóstico indicou que melhorias deveriam ser feitas. Porém, seria necessário um tempo maior para estudar onde melhorias deveriam ser feitas, tentar obter novas covariáveis que expliquem melhor a aprovação\reprovação dos alunos, e estudar melhor a análise de resíduos em grandes bases de dados. Considerando isso, as interpretações dos resultados devem ser feitas com cautela, e foram mostrados, a título de ilustração, a análise de contraste e razão de chances do modelo ajustado.

O modelo indicou que a E. Engenharia possui uma maior proporção de aprovações que o ICEx, e este por sua vez, possui mais que o Outros (FACE e IGC). Os contrastes indicaram diferenças significativas entre os parâmetros dos três institutos e a razão de chances mostrou que a chance de um aluno da Escola de Engenharia ser aprovado em cálculo I é aproximadamente 6 vezes maior que a de um aluno do ICEx, e aproximadamente 39 vezes maior que a de um aluno da FACE ou IGC.

Considerando os cursos, a análise de contrastes indicou um número bem maior de diferencias significativas entre os parâmetros, com 49 de 64 não significativas no ICEx, 64 de 78 na E. Engenharia e 2 de 3 em Outros. Como exemplos de resultados, destaca-se que no ICEx, houve diferença significativa entre os parâmetros do curso de Física Diurno e Noturno, e na E. Engenharia, não houve diferença significativa entre os cursos de Engenharia de Controle e Automação Diurno e Noturno. Em outros, a diferença não significativa foi entre os cursos de Controladoria e Finanças e Geologia.

A razão de chances entre os cursos mostrou que na Escola de Engenharia, todos os alunos possuem uma chance menor de serem aprovados que os alunos do curso de Engenharia Aeroespacial. No ICEx, somente os alunos de 4 cursos apresentam uma chance menor de serem aprovados que os alunos do curso de referência que é Matemática Diurno. Considerando os Outros, os alunos dos cursos de Controladoria e Finanças e Geologia possuem uma maior chance de serem aprovados que os alunos do curso de Ciências Econômicas.

De maneira geral, acredita-se que embora o ajuste deva ser melhorado, os resultados obtidos foram satisfatórios. E espera-se que com essa pesquisa, no futuro, estatísticos e estudantes da área que necessitem fazer análises de modelos hierárquicos desbalanceados consigam fazê-lo de maneira mais fácil.

Referências Bibliográficas

CORDEIRO, G. M.; DEMÉTRIO, C.G.B. Modelos Lineares Generalizados e Extensões. Piracicaba-SP, 2010. 249 p

CORDEIRO, G. M.; DEMÉTRIO, C.G.B. Modelos Lineares Generalizados: Minicurso para o 12º SEAGRO e a 52ª Reunião Anual da RBRAS. 2007. 161 p.

AGRESTI, A. Categorical Data Analysis. 2ª Edição. Editora Wiley-Interscience, 2002. 710 p.

AGRESTI, A. Analysis of Ordinal Categorical Data. 2ª Edição. Editora Wiley, 2010. 388 p.

PAULA, G.A. Modelos de Regressão: com apoio computacional. Universidade de São Paulo, 2013. 428 p.

MONTGOMERY, D.C. Design and Analysis of Experiments. 7ª Edição. Editora John Wiley & Sons, 2008. 680 p.

BIASE, N. G; FERREITA, D. F. Comparações Múltiplas e testes simultâneos para parâmetros binomiais de K populações independentes. Revista Brasileira Biometria, São Paulo, v.27, n.3, p.301-323, 2009.

BOOMSMA, A. Regression Diagnostics with R. University of Groningen, 2014. 23 p.

OLIVEIRA, S. Inferência e Análise de Resíduos e de Diagnóstico em Modelos Lineares Generalizados. Juiz de Fora-MG, 2013. 68 p.

CORRENTE, J. E.; NOGUEIRA, M. C. S.; COSTA, B. M. Contrastes Ortogonais na Análise do Controle de Volatilização de Amônia em Compostagem. Scientia Agricola, v.58, n.2, p.407-412, 2001.

POLETO, F. Z. Funções em S-Plus/R para análises de MLGs. Disponível em : http://www.poleto.com/funcoes.html. Acessado em: 22/10/2013.

O que é o REUNI. Disponível em: http://reuni.mec.gov.br/o-que-e-o-reuni. Acessado em: 20/11/14.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.URL http://www.R-project.org/.

SAS Institute Inc. 2008.SAS/STAT® 9.2. User's Guide. Cary, NC: SAS Institute Inc.

Anexo

Código para fazer o ajuste do modelo com dados codificado no SAS.

```
proc genmod DATA=dados;
      class inst(ref="ICEX" param=ref) curso(ref="A" param=ref);
      model ap = inst curso(inst) rsg/ dist = bin
                                link = logit;
      estimate 'ICEX-ENGENHARIA'      inst -1  0  1 ;
      estimate 'ICEX-OUTROS'          inst  0 -1  1 ;
      estimate 'ENGENHARIA-OUTROS'    inst  1 -1  0 ;
           output out=out ;
run;
```

Código para fazer o ajuste do modelo com dados codificados no R.

```
dados$curso <- relevel(dados$curso, ref="A")
dados$inst <- relevel(dados$inst, ref="ICEX")
m1 = glm(ap ~ inst/curso+rsg, data = dados,family = binomial(link =
'logit'))
summary(m1)
```

Função para fazer comparações múltiplas no R.

```
if(!require(aod)){install.packages("aod") }
require(aod)

tab.contraste<-    function(modelo,    nomes.variaveis,    contrastes,
nomes.cursos){
      cov = vcov(modelo)[nomes.variaveis, nomes.variaveis]
      # fazendo para o primeira linha
      teste = wald.test(b = coef(modelo)[nomes.variaveis],
            Sigma   =   cov,   L   =matrix(c(contrastes[1,]),   ncol   =
dim(contrastes)[2]))
      valor.p= round(teste$result$chi2[3],4)
      chisq  = round(teste$result$chi2[1],4)
      L = teste$L;    col = c()
      for ( i in 1:length(L)){
           if(L[i]!= 0) {col = c(col, i)}}
      if( length(col)==1){col = c(col, 1)}
      comparacao = c(paste(nomes.cursos[col[1]],nomes.cursos[col[2]],
sep = "--"))
      contraste.l = paste(L[1], L[2], sep = " ,")
      for( i in 3:length(L)){
```

```
          contraste.l = paste(contraste.l, L[i], sep = " ,")}
     resultado = data.frame(comparacao, chisq, valor.p, contraste.l)
     # fazendo comparacao para as demais
     for ( i in 2:dim(contrastes)[1]){
          teste = wald.test(b = coef(modelo)[nomes.variaveis],
               Sigma = cov, L =matrix(c(contrastes[i,]), ncol =
dim(contrastes)[2]))
          valor.p= round(teste$result$chi2[3],4)
          chisq  = round(teste$result$chi2[1],4)
          L     = teste$L
          col = c()
          for ( j in 1:length(L)){
               if(L[j]!= 0) {col = c(col, j)}}
          if( length(col)==1){col = c(col, 1)}
          comparacao                                                =
c(paste(nomes.cursos[col[1]],nomes.cursos[col[2]], sep = "--"))
          contraste.l = paste(L[1], L[2], sep = " ,")
          for( i in 3:length(L)){
               contraste.l = paste(contraste.l, L[i], sep = " ,")}

          aux = data.frame(comparacao, chisq, valor.p, contraste.l)
          resultado = rbind(resultado, aux)}
     # arrumando tabela
     a<- c(-0.1,0.001, 0.01, 0.05, 0.1, 1)
     sign<- as.numeric(cut(resultado$valor.p,a))
     sign<- gsub(1,"***",sign)
     sign<- gsub(2,"**",sign)
     sign<- gsub(3,"*",sign)
     sign<- gsub(4,".",sign)
     sign<- gsub(5,"",sign)
     resultado = data.frame( resultado, sign)
     resultado = resultado[, c(1,2,3,5)]; rownames(resultado) = NULL
     names(resultado)   =   c("Cursos",   "Chi-squared   test","p-
values","")
     sig.code<- "Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05
‘.’ 0.1 ‘ ’ 1"
     return ( list(resultado, sig.code, nomes.variaveis))
}
```

Código para fazer os gráficos de análise de diagnóstico.

```
residuos <- function(fit.model){
X <- model.matrix(fit.model)
n <- nrow(X)
p <- ncol(X)
h = hatvalues(fit.model)
```

```
ts <- resid(fit.model,type="pearson")/sqrt(1-h)
td <- resid(fit.model,type="deviance")/sqrt(1-h)
inf = influence.measures(fit.model)
di <- inf$infmat[,"cook.d"]
a <- max(td)
b <- min(td)
par(mfrow=c(2,2))
plot(fitted(fit.model),h,xlab="Valores Ajustados", ylab="Medida h",
main="Pontos de Alavanca", pch=16)
#
plot(di,xlab="Indice", ylab="Distancia de Cook",
main="Pontos Influentes",pch=16)
#
plot(td,xlab="Indice", ylab="Residuo Componente do Desvio",
main="Pontos Aberrantes", ylim=c(b-1,a+1), pch=16)
abline(2,0,lty=2,col = "red")
abline(-2,0,lty=2,col = "red")
#
plot(predict(fit.model),td,xlab="Preditor Linear",
ylab="Residuo Componente do Desvio",
main="Funcao de Ligacao", ylim=c(b-1,a+1), pch=16)
abline(2,0,lty=2, col = "red")
abline(-2,0,lty=2,col = "red")
}
residuos(m1)
```

Tabela 12: Contrastes dos cursos das instituições FACE e IGC no R.

Cursos	Estatísticax^2	Valor *p*
Controladoria e Finanças -- Ciências Econômicas	16,35	0,00 ***
Geologia -- Ciências Econômicas	49,62	0,00 ***
Controladoria e Finanças -- Geologia	1,88	0,17

Tabela 13: Contrastes entre os cursos da Escola de Engenharia no R.

Cursos	Estatís-ticax^2	Valor *p*	Cursos	Estatís-ticax^2	Valor *p*
Eng. Ambiental--Eng. Aeroespacial	134,76	0,00 ***	Eng. de Cont. e A. D.--Eng. Mecânica N.	7,02	0,01 **
Eng. Civil--Eng. Aeroespacial	46,47	0,00 ***	Eng. de Cont. e A. D.--Eng. Metalúrgica	56,84	0,00 ***
Eng. de Cont. e A. D.--Eng. Aeroespacial	9,19	0,00 **	Eng. de Cont. e A. D.--Eng. Química	63,35	0,00 ***
Eng. de Cont. e A. N.--Eng. Aeroespacial	4,17	0,04 *	Eng. de Cont. e A. N.--Eng. de Minas	15,85	0,00 ***
Eng. de Minas--Eng. Aeroespacial	36,41	0,00 ***	Eng. de Cont. e A. N.--Eng. de Produção	26,86	0,00 ***
Eng. de Produção --Eng. Aeroespacial	50,06	0,00 ***	Eng. de Cont. e A. N.--Eng. de Sistemas	0,84	0,36
Eng. de Sistemas--Eng. Aeroespacial	1,13	0,28	Eng. de Cont. e A. N.--Eng. Elétrica	23,42	0,00 ***
Eng. Elétrica --Eng. Aeroespacial	46,15	0,00 ***	Eng. de Cont. e A. N.--Eng. Mecânica D.	12,59	0,00 ***

Eng. Mecânica D.--Eng. Aeroespacial	32,26	0,00	***	Eng. de Cont. e A. N.--Eng. Mecânica N.	5,76	0,01	*
Eng. Mecânica N.--Eng. Aeroespacial	21,28	0,00	***	Eng. de Cont. e A. N.--Eng. Metalúrgica	29,88	0,00	***
Eng. Metalúrgica --Eng. Aeroespacial	53,77	0,00	***	Eng. de Cont. e A. N.--Eng. Química	37,45	0,00	***
Eng. Química --Eng. Aeroespacial	61,58	0,00	***	Eng. de Minas--Eng. de Produção	4,82	0,03	*
Eng. Ambiental--Eng. Civil	110,90	0,00	***	Eng. de Minas--Eng. de Sistemas	22,01	0,00	***
Eng. Ambiental--Eng. de Cont. e A. D.	177,76	0,00	***	Eng. de Minas--Eng. Elétrica	2,26	0,13	
Eng. Ambiental--Eng. de Cont. e A. N.	107,20	0,00	***	Eng. de Minas--Eng. Mecânica D.	1,03	0,31	
Eng. Ambiental--Eng. de Minas	102,10	0,00	***	Eng. de Minas--Eng. Mecânica N.	6,88	0,01	**
Eng. Ambiental--Eng. de Produção	72,44	0,00	***	Eng. de Minas--Eng. Metalúrgica	7,25	0,00	**
Eng. Ambiental--Eng. de Sistemas	109,68	0,00	***	Eng. de Minas--Eng. Química	14,77	0,00	***
Eng. Ambiental--Eng. Elétrica	89,31	0,00	***	Eng. de Produção --Eng. de Sistemas	33,17	0,00	***
Eng. Ambiental--Eng. Mecânica D.	127,18	0,00	***	Eng. de Produção --Eng. Elétrica	0,71	0,39	
Eng. Ambiental--Eng. Mecânica N.	138,62	0,00	***	Eng. de Produção --Eng. Mecânica D.	11,49	0,00	***
Eng. Ambiental--Eng. Metalúrgica	67,62	0,00	***	Eng. de Produção --Eng. Mecânica N.	21,80	0,00	***
Eng. Ambiental--Eng. Química	38,56	0,00	***	Eng. de Produção --Eng. Metalúrgica	0,20	0,65	
Eng. Civil--Eng. de Cont. e A. D.	51,52	0,00	***	Eng. de Produção --Eng. Química	3,93	0,04	*
Eng. Civil--Eng. de Cont. e A. N.	23,12	0,00	***	Eng. de Sistemas--Eng. Elétrica	29,73	0,00	***
Eng. Civil--Eng. de Minas	1,54	0,21		Eng. de Sistemas--Eng. Mecânica D.	18,58	0,00	***
Eng. Civil--Eng. de Produção	2,42	0,12		Eng. de Sistemas--Eng. Mecânica N.	10,71	0,00	**
Eng. Civil--Eng. de Sistemas	29,45	0,00	***	Eng. de Sistemas--Eng. Metalúrgica	36,16	0,00	***
Eng. Civil--Eng. Elétrica	0,38	0,53		Eng. de Sistemas--Eng. Química	43,61	0,00	***
Eng. Civil--Eng. Mecânica D.	7,92	0,00	**	Eng. Elétrica --Eng. Mecânica D.	7,58	0,00	**
Eng. Civil--Eng. Mecânica N.	19,08	0,00	***	Eng. Elétrica --Eng. Mecânica N.	17,45	0,00	***
Eng. Civil--Eng. Metalúrgica	4,74	0,02	*	Eng. Elétrica --Eng. Metalúrgica	1,81	0,18	
Eng. Civil--Eng. Química	12,55	0,00	***	Eng. Elétrica --Eng. Química	7,81	0,00	**
Eng. de Cont. e A. D.--Eng. de Cont. e A. N.	0,43	0,50		Eng. Mecânica D.--Eng. Mecânica N.	3,52	0,06	.
Eng. de Cont. e A. D.--Eng. de Minas	27,01	0,00	***	Eng. Mecânica D.--Eng. Metalúrgica	15,63	0,00	***
Eng. de Cont. e A. D.--Eng. de Produção	49,66	0,00	***	Eng. Mecânica D.--Eng. Química	24,08	0,00	***
Eng. de Cont. e A. D.--Eng. de Sistemas	2,88	0,09	.	Eng. Mecânica N.--Eng. Metalúrgica	26,73	0,00	***
Eng. de Cont. e A. D.--Eng. Elétrica	44,99	0,00	***	Eng. Mecânica N.--Eng. Química	35,08	0,00	***
Eng. de Cont. e A. D.--Eng. Mecânica D.	21,77	0,00	***	Eng. Metalúrgica --Eng. Química	2,61	0,10	