
Universidade Federal de Pernambuco
Centro de Ciências Exatas e da Natureza
Departamento de Estatística

**“Modelos estatísticos para análise de dados
longitudinais categorizados ordinais”**

Sílvia Patrícia Barreto Santana

Orientação

Prof^a Dr^a Viviana Giampaoli

Co-orientação

Prof^a Dr^a Maria Cristina Falcão Raposo

Área de Concentração
Estatística Aplicada

Dissertação apresentada ao Departamento de Estatística da Universidade Federal
de Pernambuco para obtenção do grau de Mestre em Estatística

Recife/PE
Janeiro/2004

“Modelos estatísticos para análise de dados longitudinais categorizados ordinais”

Sílvia Patrícia Barreto Santana

Dissertação apresentada ao Departamento de Estatística da UFPE, para a obtenção do grau de Mestre em Estatística, Área de Concentração: Estatística Aplicada.

Recife/PE
Janeiro/2004

DEDICATÓRIA

À

DEUS

inteligência suprema e causa primária de todas as coisas.

Aos meus pais, **Manoel da Conceição Santana** e **Patrícia Maria Barreto Santos** e a minha avó **Marieta** (*in memoriam*) pela força, incentivo e exemplos de uma filosofia de vida sincera, honesta e equilibrada, sem os quais não seria possível a obtenção de mais uma conquista .

Ao meu companheiro **Washington**, por se manter sempre dedicado, paciente e compreensível nos momentos difíceis devido a minha ausência e por acreditar no meu sucesso.

Aos meus irmãos **Sérgio**, **Sidnei**, **Sandro** e **Renato** pelo incentivo, força, carinho e atenção e por se fazerem presentes nos momentos de carência afetiva.

À minha tia e madrinha **Carmelita** pelo incentivo, apoio e muito carinho desprendido, nos momentos delicados.

AGRADECIMENTOS

À professora Viviana Giampaoli pela orientação paciente, empenho, estímulo e confiança na realização deste trabalho e também pela amizade, compreensão e carinho.

À professora Maria Cristina Falcão Raposo pela co-orientação e contribuição indispensável em conjunto para a concretização deste trabalho, além de tudo isso, pela presteza, amizade, carinho e muita compreensão.

A todos os professores do Programa de Mestrado em Estatística da UFPE.

Ao professor Pedro Israel, do Departamento de Nutrição da UFPE, pelo banco de dados cedido e pelas sugestões oferecidas para a análise dos dados.

A Livia pelos anos de amizade, carinho e preocupação.

A Patricia Leone pela amizade, apoio e carinho com que me recebeu em Recife.

A Gilson, pela amizade existente desde a graduação e a convivência compartilhada no primeiro ano.

A Tatiene pela atenção, cumplicidade e carinho, e acima de tudo uma amizade sincera.

A Patricia Leal pela confiança, carinho, companheirismo, irmandade, atenção, presteza.

A Tarciana pelo incentivo a confiança e a determinação.

A Moisés pelos momentos de descontração e histórias mirabolantes criadas e praticamente vivenciadas.

A Keila, pelo exemplo de dedicação e determinação.

A Raydonal pela amizade, carinho e atenção.

A João Marcelo, Felipe e Bartolomeu pela amizade e carinho.

A Sandra pela convivência compartilhada no segundo ano.

Aos colegas da minha turma do mestrado: Gilson, Tatiene, Patrícia, Tarciana, Keila, Moisés, Cristina, Raydonal, João Marcelo, Felipe e Bartolomeu pelos momentos vividos de tristezas, diversões e alegrias.

A todos os colegas do mestrado Gecy, Sandra Maria, Andréia, Junior, André, Lenaldo, Tatiane, Cherubino.

À secretária Valéria, pela amizade, atenção, carinho, confiança e acima de tudo competência.

À secretária Adriana pela atenção.

Aos participantes da banca examinadora pelas sugestões.

A seu Cícero e a seu Antônio.

A todos que, de forma direta ou indireta, contribuíram para a realização deste trabalho.

A minha família: tia Carmem, tia Gilsa, Gilmar, tio Marques, Ana Vitória, Nirvana, Núbia, Eró, Camila, tio Clóvis, tia Nalvinha, Márcio.

À CAPES pelo suporte financeiro.

Resumo

Modelos para dados longitudinais são modelos de regressão com respostas correlacionadas no tempo, observadas da mesma unidade amostral, podendo estas respostas serem consideradas de alguma forma grupos (*clusters*). A correlação existente deve ser levada em consideração. Uma forma de modelar esta correlação entre as observações ao longo do tempo é através das Equações de Estimação Generalizadas (GEEs). Liang & Zeger (1986) propuseram uma forma relativamente simples de tratar dados longitudinais através de Modelos Lineares Generalizados (MLG). O objetivo deste trabalho é estudar este tipo de modelo estatístico e seus métodos de estimação associados para análise de dados longitudinais categorizados ordinais e fazer uma análise de um conjunto de dados longitudinais considerando que os dados são correlacionados. Foi realizada uma aplicação com dados reais proveniente de um estudo de coorte de 477 crianças residentes nos municípios da zona da mata meridional de Pernambuco acompanhadas ao nascer e aos 2, 4, 6, 9, 12,15 e 18 meses de vida, estimando-se as probabilidades de desnutrição em relação a certas condições sócio-demográficas e biológicas, aqui denominada “condição geral” e mais a situação de aleitamento da criança. Foram construídos modelos marginais nos quais a resposta é uma variável aleatória categorizada ordinal. Através dos modelos encontrados foi verificada a importância do aleitamento materno no estado nutricional da criança. Para as análises foram utilizados os pacotes estatísticos SAS e R. Os resultados indicam que as probabilidades estimadas utilizando os dois pacotes computacionais são bastantes próximas, apesar do pacote SAS utilizar o método GEE1, e o R utilizar o método GEE2, que são métodos diferentes provenientes de extensões do GEE. As estimativas oferecidas pelo R são sempre um pouco menores que as oferecidas pelo SAS.

Palavras-chave: Dados longitudinais, equações de estimação generalizadas, modelos lineares generalizados, modelo marginal.

Abstract

Models for longitudinal data are models of regression with answers correlated in the time, observed of the same sample unit. These answers can to be considered in some sence groups (clusters). The existing correlation must be taken in consideration. A manner to modeling this correlation is through the Equations Estimating Generalized (GEE). Liang & Zeger (1986) proposed a relatively simple form to treat longitudinal data through Linear Models Generalized (GLM). The objective of this work are study these statistical model and their methods of estimation associate for analysis of longitudinal data classified ordinal, and to do an analysis of a longitudinal data set considering that the data are correlated. An application was carried out using real data proceeding from a study of coorte of 477 resident children in cities of the southern bush of Pernambuco obtained folloied to the rising and from 2, 4, 6, 9, 12, 15 and 18 months of life, estimating the probabilities of malnutrition in relation the certain partner-demographic and biological conditions, here called “general condition” and more the situation of breastfeeding of the child. Marginal models were constructed with response variate categorical ordinal. We verified the importance of the breastfeeding in the child’s state nutricional. For the analyses statistical the packages SAS and R had been used. The results indicate that the estimated probabilities using the two computational packages are sufficiently near, in despite that package SAS to use the method GEE1 and R to use the method GEE2, that are different methods proceeding from extensions of the GEE.

Words-key: Equations estimating generalized, linear models generalized, longitudinal data, model marginal.

SUMÁRIO

	Página
1. Introdução	1
1.1. Considerações iniciais	1
1.2. Organização dos capítulos.....	3
1.3. Suporte computacional.....	3
2. Modelos para variáveis categorizadas	5
2.1. Considerações iniciais	5
2.2. Modelo linear generalizado	6
2.2.1. Ligação canônica	11
2.2.2. Estimação dos parâmetros	12
2.2.3. Função desvio	16
2.2.4. Testes de hipóteses	17
2.3. Modelos para variáveis de resposta binária	19
2.3.1. Modelo loglinear	19
2.3.2. Modelo de probabilidade linear	22
2.4. Modelos para variáveis de resposta multinomial	23
2.4.1. Distribuição multinomial	23
2.4.2. Modelo loglinear	24
2.4.3. Logits para respostas ordinais	26
3. Modelos para dados longitudinais	29
3.1. Considerações iniciais	29
3.2. Método de quase-verossimilhança	31
3.3. Equações de estimação para respostas independentes	33
3.4. Equações de estimação generalizadas	34
3.4.1. Estimação de α e ϕ	34
3.5. Estimação para GEE2	39
3.5.1. Algoritmo de estimação	41
3.5.2. Propriedades assintóticas $(\hat{\beta}_{gee2}, \hat{\alpha}_{gee2})$	42
3.6. Estimação para GEE1	43
3.6.1. Algoritmo de estimação	44
3.6.2. Propriedades assintóticas $(\hat{\beta}_{gee1}, \hat{\alpha}_{gee1})$	45

4. Modelos Marginais	46
4.1. Considerações iniciais	46
5. Aplicação	51
5.1. Organização dos dados e análise preliminar	51
5.2. Modelo ajustado usando o SAS	57
5.3. Modelo ajustado usando o R considerando intercepto constante no tempo	65
5.4. Modelo ajustado usando o R considerando intercepto não constante no tempo	68
6. Conclusão	75
Apêndice	78
Referências bibliográficas	84

Capítulo 1

Introdução

1.1 Considerações iniciais

O recente desenvolvimento de métodos estatísticos para análise com dados categorizados tem estimulado a sofisticação metodológica em vários campos de interesse como: Agronomia, Biologia, Ciências da Educação, Economia, etc., em especial nas ciências sociais para medir atitudes e opiniões em várias questões e características demográficas tais como gênero, raça e classe social; em ciências da saúde escalas categorizadas são também usadas para medir, por exemplo, fatores como severidade de um ferimento, grau de recuperação a partir da cirurgia e estágio da doença. Estes métodos estatísticos para dados categorizados tem apenas recentemente alcançado o mesmo nível de sofisticação, que a metodologia para dados contínuos (Agresti, 2002).

Investigação científica em que o interesse é voltado para a avaliação do comportamento de uma ou mais variáveis respostas ao longo de uma dimensão específica como o tempo, a distância de uma certa origem ou a dosagem de uma determinada substância, são exemplos de dados provenientes de *estudos longitudinais* (Andrade & Singer, 1986). Exemplos práticos de tais estudos incluem:

- (1) pesquisas prospectivas (isto é, a unidade de investigação é exposta ao fator de risco e observada ao longo de um período) destinadas a explorar a influência de diferentes tratamentos no crescimento de crianças em idade escolar, por exemplo, dietas alimentares;
- (2) ensaios clínicos para investigar os efeitos da variação na dosagem de uma ou mais drogas em um ou mais sintomas de uma determinada moléstia;
- (3) experimentos planejados para estudar a relação entre a intensidade de poluição sonora e a distância de uma ou mais fontes de equipamentos industriais.

A pesquisa longitudinal envolve observações de um conjunto de *unidades de investigação* (por exemplo, indivíduos) classificados em diferentes *subpopulações* segundo um ou mais fatores ou tratamentos, ao longo de diversas *condições de avaliação* (por exemplo, doses) que representam as *unidades de observação*. Assim, as unidades de investigação constituem as unidades naturais para a análise e as unidades de observação correspondentes podem

ser consideradas como perfis multivariados de respostas associadas às mesmas, o *tempo* é considerado como a dimensão ao longo do qual esses perfis são observados (Andrade & Singer, 1986).

Uma das características mais importantes para a análise de estudos longitudinais é o processo de obtenção de dados (*planejamento do estudo*), dado que o método estatístico a ser utilizado é parcialmente determinado pelo mesmo. Nesse tipo de estudo, se planeja observar cada unidade de investigação em todas as ocasiões para as quais o estudo foi globalmente dimensionado. É interessante notar que planejamentos longitudinais constituem um caso especial dos chamados *planejamentos com medidas repetidas* (Koch *et al.*, 1977).

Um possível objetivo da análise do conjunto de dados longitudinais é descrever a esperança marginal da resposta como uma função das variáveis preditoras. Devido às observações repetidas feitas em cada sujeito, a correlação é prevista entre as medidas dos sujeitos, podendo ser levado em conta para obter uma análise estatística correta (Zeger & Liang, 1986). O modelo marginal descreve a distribuição da ocasião (ou tempo) da variável resposta e a dependência dessa distribuição com outras variáveis independentes (Ware *et al.*, 1988). Harville (1977) e Laird & Ware (1982) desenvolveram modelos de efeitos aleatórios em que observações repetidas ao longo do tempo para um sujeito, que são medidas correlacionadas, são termos não observáveis variando aleatoriamente entre os indivíduos, podendo ser acrescido no modelo como um componente aleatório.

Mais precisamente, após a publicação do trabalho de Liang & Zeger (1986) que propuseram uma forma relativamente simples de tratar dados longitudinais através de Modelos Lineares Generalizados (MLG), ocorreram uma série de publicações no assunto e em particular envolvendo diretamente modelos de quase-verossimilhança. A ausência de distribuições multivariadas não-gaussianas com a mesma estrutura de correlação da distribuição normal multivariada, tem transformado o trabalho de Liang & Zeger num dos mais citados em Estatística desde 1986.

No contexto da regressão com uma única observação no tempo para cada sujeito, modelos lineares generalizados e teoria de quase-verossimilhança (McCullagh & Nelder, 1983b; Wedderburn, 1974) tem estendido modelos lineares a partir do caso gaussiano para uma ampla classe de variáveis respostas. No enfoque de quase-verossimilhança, uma transformação conhecida da esperança marginal da resposta é assumida ser uma função linear das covariáveis. Em vez da especificação da distribuição da variável dependente, assumimos que sua variância é uma função conhecida de sua esperança. Por exemplo, com respostas binárias podemos assumir que o logit da probabilidade de resposta, π_i , depende linearmente das covariáveis e a variância é exatamente $\pi_i(1 - \pi_i)$. Esta especificação parcial da distribuição da resposta leva a simples técnica para análise de regressão de dados gaussianos, gama, Poisson, binomial, multinomial (categórico e ordinal) (McCullagh & Nelder, 1983b). Diversos autores tem recentemente proposto técnicas de regressão para respostas multinomiais longitudinais

(Clayton, 1992; Gange *et al.*, 1993; Kim *et al.*, 1993; Miller *et al.*, 1993), eles têm adaptado o enfoque da função de estimação de Liang e Zeger (1986) para modelos de “odds” (chances) proporcionais para respostas ordinais agrupadas.

O objetivo desta dissertação é estudar e descrever alguns modelos estatísticos e seus métodos de estimação associados para análise de dados longitudinais categorizados ordinais, para tanto faz-se necessário rever alguns conceitos sobre MLG. Foi feita uma aplicação para modelar a probabilidade de ocorrência de desnutrição a partir de um conjunto de dados provenientes de um estudo longitudinal constituído de uma coorte de 477 crianças de famílias residentes em áreas urbanas da zona da mata meridional do estado de Pernambuco, com informações completas aos 2, 4, 6, 9, 12, 15 e 18 meses de vida da criança. Foram usados modelos marginais de logits acumulados comparando diferentes critérios de estimação dos parâmetros, onde o interesse é obter estimativas consistentes levando em consideração que os dados eram correlacionados.

1.2. Organização dos capítulos

A dissertação está organizada da seguinte forma: neste capítulo uma breve revisão da idéia básica sobre os dados categorizados e longitudinais, bem como definições pertinentes aos assuntos que serão abordados nos capítulos seguintes. O Capítulo 2 discute alguns conceitos de variáveis categorizadas e uma breve revisão dos modelos lineares generalizados para variáveis respostas do tipo discreta ou contínua. O Capítulo 3 trata de modelos para dados longitudinais. O Capítulo 4 apresenta o enfoque sobre modelos marginais, onde os parâmetros de regressão têm interpretação a nível de média populacional. No Capítulo 5 encontra-se o resultado de uma aplicação considerando o enfoque dos modelos marginais. Uma discussão sobre os resultados é apresentada.

1.3. Suporte computacional

Para a realização desta dissertação foi utilizado o pacote R, que é uma linguagem de alto nível, de domínio público, desenvolvida para análise, manipulação e apresentação gráfica de dados, é um ambiente computacional preferido por muitos pois inclui um leque de métodos estatísticos tradicionais e modernos. A versão utilizada foi 1.8 (<http://www.r-project.org>).

Também foram utilizadas outras três plataformas computacionais que foram o EPI-info, SPSS 8.0 e o SAS 8.0, que são três pacotes estatísticos mais comumente usados para análise de dados em ciências sociais e ciências da saúde, além do \TeX , que é um sistema de tipografia digital para publicação científica criado por Donald Knuth na década de 70. A utilização do EPI-info ocorreu na fase de construção do banco para o cálculo da medida antropométrica

WAZ - weight for age (escore Z do peso por idade). O SPSS foi utilizado na fase seguinte para a organização do banco de dados, agrupamento das variáveis pertinentes ao estudo e também para uma análise exploratória inicial dos dados.

O SAS oferece um grande leque de métodos para dados de medidas repetidas categorizadas. A versão utilizada permite o uso do procedimento chamado PROC GENMOD que ajusta os modelos de logits acumulados para dados longitudinais, e também realiza análise GEE (sob independência, para distribuição multinomial) para modelos marginais usando uma função de ligação, incluindo logit acumulado e probit. Este procedimento não fornece estimativas pelo método de máxima verossimilhança para o ajuste de modelos marginais. Também foi utilizado o procedimento PROC LOGISTIC na construção de modelos de regressão logística para cada tempo.

A utilização do R ocorreu na fase de elaboração dos gráficos, e também na construção dos modelos de dados agrupados com ordinalidade, foi utilizado o objeto “ORDGEE” do pacote “GEESE”, o qual produz um ajuste tanto a nível de média populacional como a nível de *cluster* (grupo). O SAS utiliza o método GEE1, e o R utiliza ambos os métodos GEE1 e GEE2, que são métodos diferentes provenientes de extensões do GEE.

Capítulo 2

Modelos para variáveis categorizadas

2.1 Considerações iniciais

A Análise de Dados Categorizados visa evidenciar e interpretar a informação relevante que está contida em dados discretos provenientes de contagens de eventos ou de unidades de investigação (pessoas, lugares, objetos) possuindo certas características ou atributos definidos pela combinação das categorias de duas ou mais variáveis de interesse, ou ainda, apenas categorias de uma variável (Paulino & Singer, 1997).

A classificação do tipo da variável depende de como foi medida a característica de interesse, determinando assim a maneira que será analisada estatisticamente. Por exemplo, a variável “educação” é nominal quando se tem interesse em saber qual é o tipo de escola, tal como escola pública e privada, e é ordinal quando medida por níveis de educação, tais como, ensino fundamental, ensino médio, ensino superior; e intervalar quando medida por números de anos de estudo, usando valores inteiros 0,1,2,3

As variáveis de interesse num estudo podem ser classificadas como discretas ou contínuas de acordo com o número de valores que elas podem alcançar. Podemos chamar de discretas as variáveis que podem assumir valores referentes a uma contagem, enquanto que as contínuas são valores provenientes de mensuração, ou seja, os valores das variáveis discretas constituem um subconjunto dos números inteiros e os das contínuas um subconjunto dos números reais. Contudo, é possível ocorrer a situação em que a variável por natureza seja contínua e após uma categorização esta se torne discreta.

As variáveis podem ser classificadas de uma outra forma como quantitativas, se as respostas forem numéricas admitindo, portanto, operações aritméticas e qualitativas se as respostas forem atributos. Quanto aos níveis de mensuração (ou escala) as variáveis podem ser classificadas como:

Nominal - São variáveis para as quais os níveis não têm uma ordenação natural e a ordem da listagem das categorias é irrelevante para a análise estatística. Exemplos de variáveis nominais são: tipo de residência (casa, apartamento, outro), filiação religiosa (Católica, Protestante, outras), estado civil (solteiro, casado, divorciado, viúvo);

Ordinal - São variáveis que têm níveis ordenados, por exemplo, classe social (baixa, média, alta), atitude para legalização do aborto (desaprova fortemente, desaprova, aprova, aprova fortemente). Variáveis ordinais têm suas categorias ordenadas, porém a distância

absoluta entre as categorias são desconhecidas;

Intervalar - São variáveis que têm distância numérica entre qualquer dois níveis da escala, por exemplo, pressão sanguínea;

Razão - São observações numéricas que têm as características de escala intervalar porém, possui um ponto **zero verdadeiro** e uma unidade de medida absoluta, por exemplo, rendimentos.

Segundo Agresti (2002), na hierarquia da medida, variáveis intervalares são ditas serem de níveis superiores, variáveis ordinais são de níveis intermediários e variáveis nominais são de níveis inferiores. Métodos estatísticos adequados para variáveis de um tipo podem também ser usado para variáveis de níveis superiores, mas não de nível mais baixo. Por exemplo, métodos estatísticos para variáveis ordinais podem ser usados para variáveis intervalares (usando somente a ordem dos níveis e não suas distâncias); eles não podem ser utilizados com variáveis nominais, desde que as categorias de tais variáveis não tenham importância na ordenação. Em geral é melhor aplicar métodos apropriados para cada tipo de variável.

A classificação das variáveis ordinais no sentido de serem quantitativas ou qualitativas não é fácil discernir. Elas são freqüentemente tratadas como qualitativas, sendo analisadas usando métodos para variáveis nominais. Mas, em muitos casos as variáveis ordinais se parecem mais com variáveis intervalares do que com variáveis nominais, pois possuem características quantitativas importantes.

Os métodos não-paramétricos são apropriados para analisar variáveis na escala nominal ou ordinal de medida, ou ainda dados na escala intervalar ou de razão, desde que a distribuição da variável aleatória que produziu os dados não esteja inteiramente especificada.

Os métodos desenvolvidos adequados a análise de dados discretos relativos a uma, duas ou mais variáveis definidas qualitativamente através de um número finito de valores designados por níveis ou categorias, utiliza as denominações de *dados categorizados* e de *variáveis categorizadas*. Para o número de categorias 2, 3 ou maior que 3, as variáveis se dizem dicotômicas (ou binárias), tricotômicas ou policotômicas, respectivamente (Agresti, 2002).

Uma variável categorizada pode ser classificada como variável resposta (ou dependente) ou explicativa (ou independente). A relação entre a(s) variável(eis) dependente(s) e a(s) variável(eis) independente(s) pode ser explicada através de modelos matemáticos probabilísticos. A família de modelos lineares generalizados contém relevantes modelos para a análise de dados categorizados.

2.2 Modelo linear generalizado

Esta classe de modelos é uma extensão do modelo linear clássico e foi introduzida por

Nelder & Wedderburn (1972). Eles mostraram que uma série de modelos comumente estudados separadamente podem ser reunidos sob o nome de Modelos Lineares Generalizados (MLG).

Alguns casos especiais de MLG são:

- Modelo clássico de regressão com erro normal;
- Modelo clássico de análise de variância e de covariância com erro normal;
- Modelo de análise de variância com efeitos aleatórios, sendo um caso particular do modelo gama;
- Modelo logístico para proporções, sendo um caso particular os modelos para dados binários;
- Modelo de regressão de Poisson;
- Modelo para análise de sobrevivência, utilizado quando a variável dependente é uma duração de sobrevivência ;
- Modelo de Box & Cox;
- Alguns modelos de séries temporais.

O modelo linear tradicional é da forma

$$y_i = \mathbf{x}_i' \beta + \epsilon_i,$$

onde y_i é o valor observado da variável resposta Y_i para a i -ésima unidade de investigação, ($i = 1, \dots, n$). A quantidade \mathbf{x}_i é um vetor coluna de covariáveis ou variáveis explicativas para a unidade de investigação i e β é o vetor de parâmetros desconhecidos a ser estimado. Os ϵ_i são assumidos variáveis independentes, aleatórias, com distribuição normal de média zero ($E(\epsilon_i) = 0$) e variância constante ($\text{Var}(\epsilon_i) = \sigma^2$). Logo o valor esperado de Y_i , denotado por μ_i , é

$$E(Y_i) = \mu_i = \mathbf{x}_i' \beta.$$

No entanto existem alguns problemas para os quais o modelo linear tradicional não é apropriado, por exemplo:

- A suposição de normalidade pode não ser adequada para modelar contagens ou proporções;
- Se a média dos dados é restrita naturalmente a uma amplitude de valores, o modelo linear tradicional pode não ser apropriado desde que o preditor linear $\mathbf{x}_i' \beta$ pode assumir qualquer valor na reta real. Um exemplo seria a média de proporções que assume um valor no intervalo $[0,1]$;
- Pode não ser real assumir homocedasticidade, ou seja, que a variância dos erros seja constante para todas as observações.

O Modelo Linear Generalizado (MLG) tem como objetivo descrever a dependência da resposta média $\mu_i = E(Y_i)$ em relação as variáveis explicativas através de três componentes abordados abaixo. O modelo pode ser considerado como uma estrutura formada por um

componente aleatório, um componente sistemático e uma função de ligação, definidos a seguir:

a) Para o componente aleatório, suponha Y_1, \dots, Y_n variáveis aleatórias (v.a.) independentes cada uma com função de densidade da forma

$$f(y_i, \theta_i, \phi) = \exp\{[y_i\theta_i - b(\theta_i)]/a(\phi) + c(y_i, \phi)\}, \quad (2.1)$$

onde o parâmetro θ_i é conhecido como parâmetro natural ou canônico. A função $a(\phi)$ tem a forma $a(\phi) = \phi/w_i$, onde w_i é um peso conhecido, ϕ é o parâmetro de dispersão e $b(\cdot)$ e $c(\cdot, \cdot)$ são funções conhecidas. A expressão (2.1) acima é conhecida como família exponencial uniparamétrica com parâmetro canônico θ_i , se ϕ é conhecido; se ϕ é desconhecido então pode ou não pertencer a família exponencial biparamétrica (McCullagh & Nelder, 1983a).

Para a determinação dos dois primeiros momentos da variável Y_i , considere $l(y_i, \theta_i, \phi) = \log f(y_i, \theta_i, \phi)$ que denota a contribuição da i -ésima unidade de investigação para o logaritmo da função de verossimilhança. Então

$$l(y_i, \theta_i, \phi) = [y_i\theta_i - b(\theta_i)]/a(\phi) + c(y_i, \phi) \quad (2.2)$$

e, derivando em relação a θ_i

$$\partial l / \partial \theta_i = [y_i - \frac{\partial b(\theta_i)}{\partial \theta_i}] / a(\phi), \quad \partial^2 l / \partial \theta_i^2 = -\frac{\partial^2 b(\theta_i)}{\partial \theta_i^2} / a(\phi).$$

Aplicando a verossimilhança, considerando as relações abaixo, temos que

$$E\left(\frac{\partial l}{\partial \theta_i}\right) = 0 \quad \text{e} \quad E\left(\frac{\partial^2 l}{\partial \theta_i^2}\right) + E\left(\frac{\partial l}{\partial \theta_i}\right)^2 = 0$$

ou ainda,

$$-E\left(\frac{\partial^2 l}{\partial \theta_i^2}\right) = E\left(\frac{\partial l}{\partial \theta_i}\right)^2,$$

que sob condições de regularidade satisfaz a família exponencial.

Então, a partir da expressão (2.2) tem-se que a média e a variância são dadas, respectivamente, por

$$E(Y_i) = \mu_i = \frac{\partial b(\theta_i)}{\partial \theta_i} \quad \text{e} \quad \text{Var}(Y_i) = \frac{\partial^2 b(\theta_i)}{\partial \theta_i^2} a(\phi), \quad (2.3)$$

que na verdade pode ser expressa como uma função de sua média, isto é,

$$\text{Var}(Y_i) = a(\phi)V(\mu_i) = a(\phi)V_i,$$

com $V(\cdot)$ representando a função de variância, que caracteriza a distribuição (dada a função de variância, tem-se uma classe de distribuições correspondentes, e vice-versa). Por exemplo, a função de variância $V(\mu) = \mu(1 - \mu)$ caracteriza as distribuições binomiais com probabilidade de sucesso μ ou $(1 - \mu)$. O fator de escala, $a(\phi)$, é uma constante conhecida para alguns membros da família de modelos lineares generalizados, enquanto para outros é um parâmetro adicional a ser estimado.

b) O componente sistemático é o preditor linear. Considere a estrutura linear nos parâmetros, que relaciona η_i com as variáveis explicativas, dado por

$$\eta_i = \mathbf{x}'_i \beta, \quad i = 1, \dots, n$$

e portanto,

$$\eta_i = \sum_{j=1}^p x_{ij} \beta_j,$$

onde β_j é o j -ésimo componente do vetor $\beta = (\beta_1, \dots, \beta_p)'$ de parâmetros, η_i é a i -ésima componente do vetor $\eta = (\eta_1, \dots, \eta_n)'$ de preditores lineares e $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$ é a i -ésima linha da matriz \mathbf{x} do modelo, sendo esta conhecida.

c) A função de ligação relaciona-se com o valor esperado (μ_i) de Y_i e com o preditor linear por

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

ou ainda,

$$g(\mu_i) = \mathbf{x}'_i \beta,$$

onde $g(\cdot)$ é uma função monótona e diferenciável. Isto é, a ligação entre a média e o preditor não é necessariamente mais a identidade, podendo assumir qualquer forma monótona não-linear, que descreva a relação existente entre o preditor linear, η_i , e o valor esperado (μ_i) da variável Y_i .

Para construir um MLG é preciso decidir pela escolha da distribuição da variável resposta, a matriz do modelo e a função de ligação. No modelo linear clássico tem-se $\eta = \mu$ que é chamada ligação identidade. Essa ligação é adequada no sentido em que ambos η e μ podem assumir valores na linha real.

Quando se trabalha com, por exemplo, a distribuição de Poisson em que $\mu > 0$, surgem algumas restrições, neste caso, a função de ligação identidade não deve ser usada, pois $\hat{\eta}$ poderá assumir valores negativos dependendo dos valores obtidos de $\hat{\beta}$ (Demétrio, 2002).

Embora as funções de ligações levem as propriedades estatísticas desejáveis para o modelo, principalmente no caso de amostras pequenas, não existe, em geral, razão em princípio para que os efeitos sistemáticos do modelo devam ser aditivos na escala dada por tais funções (McCullagh & Nelder, 1989).

Quando o parâmetro natural e o preditor linear coincidem, a função de ligação é chamada função de ligação canônica. Neste caso especial, tem-se que

$$\theta_i = \eta_i = \sum_{j=1}^p x_{ij}\beta_j, \quad i = 1, \dots, n.$$

Para estas funções, denominadas funções de ligação naturais, obtém-se, sempre que existirem, estimativas de máxima verossimilhança (MV) únicas para os parâmetros β_1, \dots, β_p .

A seguir estão apresentadas algumas distribuições com parametrização na respectiva família exponencial, como por exemplo:

1. Normal

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

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

onde $-\infty < \mu < \infty$, $-\infty < y < \infty$ e $\sigma^2 > 0$.

Portanto, $\theta = \mu$, $b(\theta) = \theta^2/2$, $a(\phi) = \sigma^2$, $\phi = \sigma^2$, $w = 1$, e $c(y, \phi) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{y^2}{2\sigma^2}$ e função de variância dada por $V(\mu) = 1$.

2. Poisson

No caso de Y ter distribuição de Poisson com parâmetro μ , a densidade é dada por

$$e^{-\mu}\mu^y/y! = \exp\{y\log\mu - \mu - \log(y!)\}, \quad (2.4)$$

onde $\mu > 0$ e $y = 0, 1, \dots$. Fazendo $\log\mu = \theta$, $b(\theta) = e^\theta$, $a(\phi) = 1$, $\phi = 1$, $w = 1$ e $c(y, \phi) = -\log(y!)$ e portanto $V(\mu) = \mu$.

3. Binomial

Seja Y^* a proporção de sucessos em n ensaios independentes, cada um com probabilidade de ocorrência μ . Assumindo que $nY^* \sim B(n, \mu)$, a densidade de nY^* fica então expressa na forma

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

ou equivalentemente,

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

onde $0 < \mu < 1$ e $0 < y^* < 1$. Fazendo $\theta = \log\{\mu/(1 - \mu)\}$, $b(\theta) = \log(1 + e^\theta)$, $a(\phi) = 1/n$, $\phi = 1/n$, $w = 1$ e $c(y^*, \phi) = \log\binom{n}{ny^*}$, sendo a função de variância dada por $V(\mu) = \mu(1 - \mu)$.

2.2.1 Ligação canônica

O logaritmo da função de verossimilhança de um MLG com respostas independentes pode ser expresso na forma, onde $L(\theta; y) = \sum_i^n l(y_i, \theta_i, \phi)$ e portanto,

$$L(\theta; y) = \sum_{i=1}^n \{y_i \theta_i - b(\theta_i)\} / a(\phi) + \sum_{i=1}^n c(y_i, \phi). \quad (2.5)$$

Um caso particular importante ocorre quando o parâmetro canônico (θ_i) coincide com o preditor linear, isto é, quando $\theta_i = \eta_i = \sum_{j=1}^p x_{ij} \beta_j$. Nesse caso, $L(\theta; y) = L(\beta; y)$ fica dado por

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

e as quantidades

$$s_j = \frac{1}{a(\phi)} \sum_{i=1}^n y_i x_{ij}, \quad j = 1, \dots, p,$$

são estatísticas suficientes de dimensão mínima para os parâmetros β_1, \dots, β_p .

Daí $L(\beta; y)$ fica então na forma

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

Logo, pelo teorema da fatoração, a estatística $s = (s_1, \dots, s_p)'$ é suficiente minimal para o vetor $\beta = (\beta_1, \dots, \beta_p)'$.

As ligações que correspondem a tais estatísticas são chamadas de ligações canônicas e desempenham um papel importante na teoria dos MLG's. Portanto, por exemplo, as ligações canônicas para as distribuições Normal, Poisson e binomial são dadas por

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

respectivamente.

Uma das vantagens de usar ligações canônicas é que as mesmas garantem a concavidade de $L(\beta; y)$ e conseqüentemente muitos resultados assintóticos são obtidos mais facilmente. Por exemplo, a concavidade de $L(\beta; y)$ garante a unicidade da estimativa de máxima verossimilhança de β , quando essa existe.

Ao se considerar por exemplo a família exponencial da distribuição binomial podem ser consideradas como funções de ligação:

1. Logit

$$\eta_1 = \log\left\{\frac{\mu}{1-\mu}\right\};$$

2. Probit

$$\eta_2 = \Phi^{-1}(\mu),$$

onde $\Phi(\cdot)$ é a função distribuição Normal acumulada;

3. Complemento log-log

$$\eta_3 = \log\{-\log(1-\mu)\};$$

4. Log-log

$$\eta_4 = -\log\{-\log(\mu)\},$$

a qual é equivalente natural da função complemento log-log. É raramente usada porque seu comportamento não é apropriado para $\mu < 1/2$, que é a principal razão de interesse.

2.2.2 Estimação dos parâmetros

Uma propriedade da família exponencial de distribuições é que seus elementos satisfazem as condições de regularidades suficientes para assegurar que o máximo global do logaritmo da função de verossimilhança $L(\theta, y)$ é obtido da solução do sistema de equações

$$\mathbf{U}(\theta) = \frac{\partial L}{\partial \theta} = \mathbf{0} \text{ ou } \mathbf{U}(\beta) = \frac{\partial L}{\partial \beta} = \mathbf{0}.$$

Tem-se então, que a função escore é dada por

$$U_j(\beta_j) = \sum_{i=1}^n \frac{\partial L}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial l(\theta_i, y_i)}{\partial \beta_j}, \text{ para } j = 1, \dots, p.$$

Mas $L = \log f(\theta_1, \dots, \theta_n)$, onde $\theta_i = \int V_i^{-1} d\mu_i = h(\mu_i)$, em que V_i é a função de variância e $h(\mu_i)$ uma função conhecida de μ_i , e $\mu_i = g^{-1}(\eta_i)$ e portanto $\eta_i = \sum_{j=1}^p x_{ij}\beta_j$ e, então pela regra da cadeia, tem-se

$$U_j(\beta_j) = \sum_{i=1}^n \frac{\partial l}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j}, \text{ para } j = 1, \dots, p,$$

onde $\frac{\partial l}{\partial \theta_i} = [y_i - b'(\theta_i)]/a_i(\phi) = [y_i - \mu_i]/a_i(\phi)$, fazendo $a_i(\phi) = \frac{\phi}{w_i}$, $\frac{\partial \mu_i}{\partial \theta_i} = \frac{\partial^2 b(\theta_i)}{\partial \theta_i^2} = V_i$, com $\frac{\partial \mu_i}{\partial \eta_i}$ dependendo da função de ligação g e $\frac{\partial \eta_i}{\partial \beta_j} = x_{ij}$ e substituindo as derivadas das relações apresentadas temos que em qualquer modelo linear generalizado, os coeficientes de regressão β podem ser estimados pela solução da mesma equação escore dada abaixo

$$U_j(\beta_j) = \sum_{i=1}^n \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \frac{[y_i - \mu_i]}{a_i(\phi)V_i} x_{ij} = 0, \text{ para } j = 1, \dots, p$$

ou na forma matricial, obtém-se

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

onde \mathbf{X} é uma matriz $n \times p$ de posto completo, $\mathbf{W} = \text{diag}\{W_1, \dots, W_n\}$ é a matriz de pesos com componentes $W_i = (\partial \mu_i / \partial \eta_i)^2 w_i / V_i$, $\mathbf{V} = \text{diag}\{V_1, \dots, V_n\}$, $\mathbf{y} = (y_1, \dots, y_n)'$ e $\mu = (\mu_1, \dots, \mu_n)'$.

As equações escores, em geral, não são lineares e, nesse caso, são resolvidas por processos iterativos como os de Newton-Raphson.

O método de Newton-Raphson é bastante útil, porém isso nem sempre é possível utilizá-los e nos casos de modelos lineares generalizados, quando as derivadas de segunda ordem são obtidas facilmente, usa-se o método de escore de Fisher que, em geral, é mais simples (coincidindo com o método de Newton-Raphson no caso de funções de ligações canônicas). Este método envolve a substituição da matriz de derivadas parciais de segunda ordem pela matriz de valores esperados das derivadas parciais.

Considerando

$$U_j(\beta_j) = \sum_{i=1}^n \frac{\partial L}{\partial \beta_j} = \sum_{i=1}^n \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \frac{[y_i - \mu_i]}{a_i(\phi)V_i} x_{ij},$$

temos que,

$$\begin{aligned} \frac{\partial^2 L(\beta, \mathbf{y})}{\partial \beta \partial \beta'} &= \sum_{i=1}^n \frac{\partial L^2}{\partial \beta_j \partial \beta_k} = \sum_{i=1}^n \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \frac{[y_i - \mu_i]^2}{[a_i(\phi)]^2 [V_i]^2} x_{ij} x_{ik} \\ &= \sum_{i=1}^n \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \frac{a_i(\phi) V_i}{[a_i(\phi)]^2 [V_i]^2} x_{ij} x_{ik} \\ &= \sum_{i=1}^n \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \frac{1}{a_i(\phi) V_i} x_{ij} x_{ik} = \sum_{i=1}^n \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \frac{w_i}{\phi V_i} x_{ij} x_{ik} \end{aligned}$$

E, portanto, a matriz de informação de Fisher é dada por

$$\mathbf{K}(\beta) = E\left\{ -\frac{\partial^2 L(\beta, \mathbf{y})}{\partial\beta\partial\beta'} \right\} = E\left\{ \frac{\partial L(\beta, \mathbf{y})\partial L(\beta, \mathbf{y})}{\partial\beta\partial\beta'} \right\} = \phi^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}.$$

Para as funções de quase-verossimilhanças, esta matriz representa o mesmo papel da informação de Fisher para as equações de verossimilhanças ordinárias. Em particular, sob usual condições dos autovalores de (2.6), a matriz de covariância assintótica para $\hat{\beta}$ é

$$\text{cov}(\hat{\beta}) \simeq (\mathbf{K}(\beta))^{-1} = \hat{\phi}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}.$$

Assim, o erro padrão “Naive” para $\hat{\beta}$, obtido a partir da estimativa “Naive” de $\text{cov}(\hat{\beta})$ é

$$\hat{\sigma}^2(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}. \quad (2.6)$$

Em particular para as ligações canônicas, em que $\mathbf{W} = \mathbf{V}$, tem-se

$$\mathbf{U}(\beta) = \frac{\partial L(\beta, \mathbf{y})}{\partial\beta} = \phi^{-1}\mathbf{X}'(\mathbf{y} - \mu),$$

e

$$\mathbf{K}(\beta) = E\left\{ -\frac{\partial^2 L(\beta, \mathbf{y})}{\partial\beta\partial\beta'} \right\} = \phi^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}.$$

A solução $\hat{\beta}$, que é uma estimativa de máxima verossimilhança, é obtida numericamente por processos iterativos do tipo Newton-Raphson.

O processo iterativo de Newton-Raphson para a solução de uma equação do tipo $h(x) = 0$ é baseado na aproximação de Taylor para a função $h(x)$ nas vizinhanças do ponto x_0 , ou seja,

$$h(x) \approx h(x_0) + (x - x_0)h'(x_0) = 0,$$

originando

$$x = x_0 - \frac{h(x_0)}{h'(x_0)},$$

ou, de forma mais geral

$$x^{(m+1)} = x^{(m)} - \frac{h(x^{(m)})}{h'(x^{(m)})},$$

sendo $x^{(m+1)}$ o valor de x no passo m , $h(x^{(m)})$ a função $h(x)$ avaliada em $x^{(m)}$ e $h'(x^{(m)})$ a derivada da função $h(x)$ avaliada em $x^{(m)}$.

Portanto, a utilização do processo iterativo de Newton-Raphson para a obtenção da estimativa de máxima verossimilhança de β consiste em expandir a função escore $\mathbf{U}(\beta)$ em torno de um valor inicial $\beta^{(0)}$, tal que

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

onde $\mathbf{U}'(\beta)$ denota a primeira derivada de $\mathbf{U}(\beta)$ com respeito a β . Assim, repetindo-se o procedimento acima, chega-se ao processo iterativo

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

$m = 0, 1, \dots$ Como a matriz $-\mathbf{U}'(\beta)$ pode não ser positiva definida, a aplicação do método de “scoring” de Fisher substituindo a matriz $-\mathbf{U}'(\beta)$ pelo correspondente valor esperado, pode ser mais conveniente. Isso resulta o seguinte processo iterativo

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

$m = 0, \dots$ Trabalhando um pouco o lado direito da expressão acima multiplicando ambos os lados pela matriz de informação de Fisher, $\mathbf{K}(\beta)^{(m)}$, temos

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

ou

$$\phi^{-1}\mathbf{X}'\mathbf{W}^{(m)}\mathbf{X}\beta^{(m+1)} = \phi^{-1}\mathbf{X}'\mathbf{W}^{(m)}\mathbf{X}\beta^{(m)} + \phi^{-1}\mathbf{X}'\mathbf{W}^{(m)1/2}\mathbf{V}^{(m)-1/2}(\mathbf{y} - \mu),$$

$$\mathbf{X}'\mathbf{W}^{(m)}\mathbf{X}\beta^{(m+1)} = \mathbf{X}'\mathbf{W}^{(m)}\mathbf{X}\beta^{(m)} + \mathbf{X}'\mathbf{W}^{(m)1/2}\mathbf{V}^{(m)-1/2}(\mathbf{y} - \mu),$$

$$\mathbf{X}'\mathbf{W}^{(m)}\mathbf{X}\beta^{(m+1)} = \mathbf{X}'\mathbf{W}^{(m)}[\mathbf{X}\beta^{(m)} + \mathbf{W}^{(m)-1/2}\mathbf{V}^{(m)-1/2}(\mathbf{y} - \mu)],$$

$$\mathbf{X}'\mathbf{W}^{(m)}\mathbf{X}\beta^{(m+1)} = \mathbf{X}'\mathbf{W}^{(m)}[\eta^{(m)} + \mathbf{W}^{(m)-1/2}\mathbf{V}^{(m)-1/2}(\mathbf{y} - \mu)],$$

e definindo $\mathbf{Z}^{(m)} = \eta^{(m)} + \mathbf{W}^{(m)-1/2}\mathbf{V}^{(m)-1/2}(\mathbf{y} - \mu)$, obtém-se

$$\mathbf{X}'\mathbf{W}^{(m)}\mathbf{X}\beta^{(m+1)} = \mathbf{X}'\mathbf{W}^{(m)}\mathbf{Z}^{(m)},$$

que equivale a um processo iterativo de mínimos quadrados ponderados da seguinte forma

$$\beta^{(m+1)} = (\mathbf{X}'\mathbf{W}^{(m)}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{(m)}\mathbf{Z}^{(m)}, \quad (2.7)$$

onde $m = 0, 1, \dots$ e $\mathbf{Z} = \eta + \mathbf{W}^{-1/2}\mathbf{V}^{-1/2}(\mathbf{y} - \mu)$. Note que \mathbf{Z} desempenha o papel de uma variável dependente modificada, enquanto que \mathbf{W} é uma matriz de pesos que muda a cada passo do processo iterativo. A convergência de (2.7) ocorre em um número finito de passos, independente dos valores iniciais utilizados. É usual iniciar (2.7) com $\eta^{(0)} = g(\mathbf{y})$.

Para o caso do modelo Binomial logístico linear, temos que

$$\mathbf{Z} = \eta + \mathbf{W}^{-1/2}\mathbf{W}^{-1/2}(\mathbf{y} - \mu) = \eta + \mathbf{W}^{-1}(\mathbf{y} - \mu) = \eta + (\mathbf{y} - n\mu)/n\mu(1 - \mu).$$

Sob condições de regularidade temos que $\hat{\beta}$ é um estimador consistente e eficiente de β e que

$$\sqrt{n}(\hat{\beta} - \beta) \rightarrow N(\mathbf{0}, \phi^{-1}(\Sigma(\beta))^{-1}) \text{ quando } n \rightarrow \infty,$$

onde

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

sendo $\Sigma(\beta)$ uma matriz definida positiva.

E as condições para que exista $\Sigma(\beta)$ e seja definida positiva são

$$n_i/n \rightarrow a_i > 0, \quad n \rightarrow \infty$$

e $\sum_{i=1}^g \mathbf{x}_i \mathbf{x}_i'$ seja de posto completo, onde $n = n_1 + \dots + n_g$.

Sob certas condições de regularidade temos que

$$\sqrt{n}(\hat{\phi} - \phi) \rightarrow N(\mathbf{0}, \sigma_\phi^2),$$

quando $n \rightarrow \infty$, onde $\sigma_\phi^2 = \lim_{n \rightarrow \infty} -n[\sum_{i=1}^n c''(y_i, \phi)]^{-1}$, ou seja um estimador consistente para $\text{Var}(\hat{\phi})$ é $[\sum_{i=1}^n -c''(y_i, \phi)]^{-1}$.

2.2.3 Função desvio

Seja o logaritmo da função de verossimilhança da família exponencial de Y_1, \dots, Y_n , variáveis aleatórias, definido em (2.5) agora escrito como função da média, isto é:

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

onde $\mu_i = g^{-1}(\eta_i)$ e $\eta_i = \mathbf{x}_i' \beta$. Considerando o número de componentes do parâmetro β (p) igual ao número de observações n , temos então o modelo saturado e a função $L(\mu; \mathbf{y})$ é estimada por

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

isto é, a estimativa de máxima verossimilhança de μ_i fica nesse caso expressa por $\hat{\mu}_i^0 = y_i$.

Seja a estimativa de $L(\mu; \mathbf{y})$ dada por $L(\hat{\mu}; \mathbf{y})$, quando $p < n$. A estimativa de máxima verossimilhança de μ_i será dada por $\hat{\mu}_i = g^{-1}(\hat{\eta}_i)$, onde $\hat{\eta}_i = \mathbf{x}_i' \hat{\beta}$.

A qualidade do ajuste de um MLG pode ser avaliada através da *função desvio*

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

que é a distância entre o logaritmo da função de verossimilhança do modelo saturado (com n parâmetros) e do modelo sob investigação (com p parâmetros) avaliado na estimativa

de máxima verossimilhança $\hat{\beta}$. Um valor pequeno para a função desvio (ou “deviance”) indica que para um número menor de parâmetros, obtém-se um ajuste tão bom quanto o ajuste no modelo saturado. Se denotarmos por $\hat{\theta}_i = \theta_i(\hat{\mu}_i)$ e $\hat{\theta}_i^0 = \theta_i(\hat{\mu}_i^0)$ as estimativas de máxima verossimilhança de θ para os modelos com p parâmetros ($p < n$) e saturado ($p = n$), respectivamente, temos que a função $D(\mathbf{y}; \hat{\mu})$ fica dada por

$$D(\mathbf{y}; \hat{\mu}) = 2 \sum_{i=1}^n w_i \{y_i(\hat{\theta}_i^0 - \hat{\theta}_i) + [b(\hat{\theta}_i) - b(\hat{\theta}_i^0)]\},$$

que é a chamada *deviance* para o modelo corrente.

A seguir apresentamos a função desvio para alguns casos particulares:

1. Normal

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

2. Poisson

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

3. Binomial

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

2.2.4 Testes de hipóteses

Os métodos de inferência nos modelos lineares generalizados baseiam-se, fundamentalmente, na teoria de máxima verossimilhança. Os teste a seguir mencionados são assintoticamente equivalentes, sendo a razão de verossimilhanças o critério que define um teste uniformemente mais poderoso (Demétrio, 2002). De acordo com essa teoria, existem três estatísticas mais importantes para testar hipóteses relativas aos parâmetros β 's, que são deduzidas de distribuições assintóticas.

A primeira denominada estatística de Wald (1943) é baseada na distribuição assintótica de $\hat{\beta}$ e é geralmente mais usada no caso de hipóteses relativas a um único coeficiente β_j .

A segunda estatística é a razão de verossimilhanças e é geralmente preferida no caso de hipóteses relativas a vários coeficientes de β 's. A terceira estatística, chamada estatística escore, é conveniente em situações em que a hipótese alternativa é mais complicada que a nula.

Suponha as hipóteses simples: $H_0 : \beta = \beta^0$ contra $H_1 : \beta \neq \beta^0$, onde β^0 é um vetor p -dimensional conhecido e ϕ também conhecido e $\hat{\mu}^0$ é a estimativa sob a hipótese nula.

Para a análise do desvio, suponha para o vetor de parâmetros β a partição $\beta = (\beta_1', \beta_2')'$, onde β_1 é um vetor q -dimensional enquanto β_2 tem dimensão $p - q$ e ϕ é conhecido ou fixo. Portanto, podemos estar interessados em testar as hipóteses $H_0 : \beta_1 = 0$ contra $H_1 : \beta_1 \neq 0$. As funções desvios correspondentes aos modelos sob H_0 e H_1 que serão denotadas por $D(\mathbf{y}; \hat{\mu}^0)$ e $D(\mathbf{y}; \hat{\mu})$, respectivamente, onde $\hat{\mu}^0$ é a estimativa de máxima verossimilhança que fica nesse caso dada por

$$\xi_{RV} = \{D(\mathbf{y}; \hat{\mu}^0) - D(\mathbf{y}; \hat{\mu})\} / \phi, \quad (2.8)$$

que é a diferença entre duas funções desvios.

Podendo ainda a estatística ξ_{RV} ser definida como

$$\xi_{RV} = 2\{L(\hat{\beta}; \mathbf{y}) - L(\beta^0; \mathbf{y})\}.$$

Se as diferenças são grandes, então, H_0 é rejeitada. Como mencionado anteriormente, sob a hipótese nula, $\xi_{RV} \sim \mathcal{X}_q^2$ quando $n \rightarrow \infty$. Portanto, H_0 é rejeitada, a um nível de $(1 - \alpha)\%$ de confiança se $\xi_{RV} > \mathcal{X}_{q, 1-\alpha}^2$, onde $\hat{\mu}^0 = \mathbf{g}^{-1}(\hat{\eta}^0)$, $\hat{\eta}^0 = \mathbf{X}'\beta^0$. Portanto ξ_{RV} é a estatística do teste da razão de verossimilhanças.

De forma similar, para o caso de hipótese simples, é possível definir a estatística

$$F = \frac{\{D(\mathbf{y}; \hat{\mu}^0) - D(\mathbf{y}; \hat{\mu})\} / q}{D(\mathbf{y}; \hat{\mu}) / (n - p)}, \quad (2.9)$$

cuja distribuição assintótica é uma $F_{q, (n-p)}$ quando o denominador de (2.9) é uma estimativa consistente de ϕ^{-1} . A vantagem de utilizar a estatística F em relação a estatística de máxima verossimilhança é que a primeira não depende do parâmetro de dispersão.

O teste de Wald é baseado na distribuição normal assintótica de $\hat{\beta}$ e é uma generalização da estatística t de Student (Wald, 1943).

A estatística de teste é definida por

$$\xi_W = [\hat{\beta} - \beta^0]' \{\hat{V}ar(\hat{\beta})\}^{-1} [\hat{\beta} - \beta^0],$$

onde $\hat{V}ar(\hat{\beta})$ é a matriz de variância-covariância assintótica estimada de $\hat{\beta}$. Sob a hipótese nula, $\xi_W \sim \mathcal{X}_q^2$ quando $n \rightarrow \infty$. Portanto, H_0 é rejeitada, a um nível de $(1 - \alpha)\%$ de confiança, se $\xi_W > \mathcal{X}_{q, 1-\alpha}^2$.

O teste Escore é definido quando $\mathbf{U}'(\hat{\beta}) = 0$ por

$$\xi_{SR} = \mathbf{U}'(\beta^0)\{\hat{Var}_0(\hat{\beta})\}\mathbf{U}(\beta^0),$$

onde $\hat{Var}_0(\hat{\beta})$ é a variância assintótica estimada de $\hat{\beta}$ sob H_0 . Sob a hipótese nula, $\xi_{SR} \sim \chi_q^2$ quando $n \rightarrow \infty$. Portanto, H_0 é rejeitada, a um nível de $(1 - \alpha)\%$ de confiança, se $\xi_{SR} > \chi_{q,1-\alpha}^2$.

2.3 Modelos para variáveis de resposta binária

Um dos casos particulares dos MLGs são os modelos para variáveis que apresentam apenas duas categorias ou que foram de alguma forma dicotomizadas, as variáveis que assumem valores 0 ou 1 são chamadas variável “dummy”. Um dos mais importantes modelos é o de regressão logística, baseado na transformação *logit* para proporção. Existem também mais alguns modelos para dados binários, os modelos de probabilidade linear, os modelos *probit*, os modelos *tobit* (de regressão censurada) e *valor-extremo*.

Variáveis com duas categorias que podem ser classificadas como sucesso ou fracasso representando as possibilidades de respostas como, por exemplo 0 e 1, podem ser caracterizadas pela distribuição de Bernoulli. Comumente é chamado de “sucesso” o resultado mais importante da resposta ou aquele resultado que se pretende relacionar com as outras variáveis de interesse. A distribuição de Bernoulli para variável aleatória binária Y_i de parâmetro π_i especifica as probabilidades como $\Pr(Y_i = 1) = \pi_i$ e $\Pr(Y_i = 0) = 1 - \pi_i$, para o qual temos que $E(Y_i) = \mu_i = \pi_i$, pois por definição,

$$E(Y_i) = 1(\pi_i) + 0(1 - \pi_i) = \pi_i,$$

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

$$\text{Var}(Y_i) = E(Y_i^2) - [E(Y_i)]^2 = 1^2(\pi_i) + 0^2(1 - \pi_i) - [\pi_i]^2 = (\pi_i) - [\pi_i]^2 = \pi_i(1 - \pi_i),$$

isto é, $\text{Var}(Y_i) = \pi_i(1 - \pi_i)$. Então para esta variável aleatória a função de probabilidade é

$$\begin{aligned} f(y_i; \pi_i) &= \pi_i^{y_i} (1 - \pi_i)^{1-y_i} = \pi_i^{y_i} \frac{(1 - \pi_i)}{(1 - \pi_i)^{y_i}} \\ &= (1 - \pi_i) [\pi_i / (1 - \pi_i)]^{y_i} \\ &= (1 - \pi_i) \exp \left[y_i \log \left(\frac{\pi_i}{1 - \pi_i} \right) \right] \\ &= \exp \left[y_i \log \left(\frac{\pi_i}{1 - \pi_i} \right) + \log(1 - \pi_i) \right], \quad i = 1, \dots, n, \end{aligned}$$

para $y_i = 0$ ou $y_i = 1$. Portanto, esta função é da forma (2.1) logo esta distribuição está na família exponencial natural.

Quando a função de ligação é a identidade tem-se o modelo de probabilidade linear, porém isso as vezes não é conveniente e outras funções de ligação podem ser apropriadas.

2.3.1 Modelo loglinear

Na análise de variáveis qualitativas, os modelos mais usuais são os modelos *loglineares*, isto é, modelos em que se especifica que os logaritmos das frequências esperadas são uma

combinação linear de certos parâmetros. Esses modelos são de grande utilidade na descrição de padrões de associação entre variáveis categorizadas. Os modelos podem ser do tipo loglinear geral ou de decisão binária.

Quando é natural considerar uma variável como resposta e outras como variáveis explicativas (explanatórias), certos modelos loglineares são equivalentes a modelos *logit* para aquela variável resposta. Em muitas situações, modelar os efeitos das variáveis explicativas nas respostas é mais importante do que modelar o relacionamento entre variáveis explicativas.

Se considerarmos, por exemplo, variáveis aleatórias Y_i independentes com distribuição Poisson, isto é, com função de densidade dada por (2.4) temos que a função de ligação do valor esperado de Y_i com o preditor linear é

$$\log \mu_i = \eta_i = \mathbf{x}'_i \beta, \quad i = 1, \dots, n, \quad (2.10)$$

que é chamado modelo *loglinear*. Muitos modelos loglineares não têm expressões de forma fechada para as estimativas de MV e para a obtenção das estimativas pode-se utilizar o algoritmo iterativo método Newton-Raphson descrito anteriormente. Quando as variáveis explicativas são categorizadas os modelos logit têm representações equivalentes aos modelos loglineares. Os modelos logit e probit são modelos loglineares particulares e serão apresentados a seguir.

Modelo logit

Seja o parâmetro natural $\theta_i(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right)$, o logaritmo da razão das chances da resposta 1, é chamado o *logit* de π_i . Os MLG que usam a ligação *logit* são chamados *modelos logits*, para os quais o preditor linear é dada por

$$\eta_i = \mathbf{x}'_i \beta = \log\left(\frac{\pi_i}{1-\pi_i}\right),$$

onde $\pi_i(\mathbf{x})$ denota probabilidade de $y_i = 1$ para os valores das variáveis explicativas, ou seja, o grupo de preditores $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$.

No caso particular em que se tem $\mathbf{x} = (1, x_1)$ e quando a relação entre \mathbf{x} e $\pi_i(\mathbf{x})$ não é linear, ou seja, quando a relação se apresenta de forma curvilínea e monótona, então a função de ligação que relaciona o valor esperado de Y_i ao componente linear, isto é, $g(\mu_i) = \mathbf{x}'_i \beta$, da forma

$$\pi_i(\mathbf{x}) = \frac{\exp(\beta_0 + \beta_1 x_1)}{1 + \exp(\beta_0 + \beta_1 x_1)}.$$

Daí,

$$[1 + \exp(\beta_0 + \beta_1 x_1)]\pi_i(\mathbf{x}) = \exp(\beta_0 + \beta_1 x_1)$$

$$\pi_i(\mathbf{x}) + \exp(\beta_0 + \beta_1 x_1)\pi_i(\mathbf{x}) = \exp(\beta_0 + \beta_1 x_1)$$

$$\pi_i(\mathbf{x}) = \exp(\beta_0 + \beta_1 x_1) - \exp(\beta_0 + \beta_1 x_1)\pi_i(\mathbf{x})$$

$$\begin{aligned} \pi_i(\mathbf{x}) &= \exp(\beta_0 + \beta_1 x_1)[1 - \pi_i(\mathbf{x})] \\ \frac{\pi_i(\mathbf{x})}{[1 - \pi_i(\mathbf{x})]} &= \exp(\beta_0 + \beta_1 x_1), \end{aligned} \quad (2.11)$$

ou, equivalentemente a

$$\log \left[\frac{\pi_i(\mathbf{x})}{1 - \pi_i(\mathbf{x})} \right] = \beta_0 + \beta_1 x_1 = \mathbf{x}(\beta_0, \beta_1)^T,$$

que é chamada função de *regressão logística*. Assim, a função de ligação é o logaritmo das chances (odds), o *logit*.

A fórmula fornece uma interpretação simples para β . As chances aumentam multiplicando por e^{β_1} para cada aumento na unidade em x_1 . Para centrar o preditor em torno de 0 (isto é, substituindo x por $(x - \bar{x})$) β_0 torna-se o logit da média, e assim $e^{\beta_0} / (1 + e^{\beta_0}) = \pi(\bar{x})$.

Generalizando para múltiplas variáveis explicativas, o modelo de regressão logística para valores $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)'$ de p variáveis explicativas é

$$\text{logit}[\pi_i(\mathbf{x})] = \log \left[\frac{\pi_i(\mathbf{x})}{1 - \pi_i(\mathbf{x})} \right] = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \dots + \beta_p \mathbf{x}_p. \quad (2.12)$$

Assim quando a variável resposta é categorizada, modelos com ligação *logit* são um caso particular dos modelos *loglineares*.

Modelo probit

Se consideramos agora a relação entre \mathbf{x} e $\pi(\mathbf{x})$ da forma

$$\pi(\mathbf{x}) = \Pr(Z^* \leq \beta_0 + \beta_1 x_1) = \Phi(\beta_0 + \beta_1 x_1) \quad (2.13),$$

temos o denominado *Modelo Probit*, onde Z^* é uma variável aleatória com distribuição Normal Padrão e Φ é a densidade normal padrão acumulada, β_0 e β_1 são parâmetros desconhecidos a serem estimados. O modelo probit é uma função não linear para um conjunto linear de parâmetros e nesse caso a ligação do valor esperado com o preditor linear tem a seguinte forma

$$\eta = \Phi^{-1}(\pi(\mathbf{x})),$$

onde $\eta = \beta_0 + \beta_1 \mathbf{x}_1$ e portanto,

$$\text{probit}[\pi(\mathbf{x})] = \Phi^{-1}[\pi(\mathbf{x})] = (\beta_0 + \beta_1 \mathbf{x}_1),$$

é um MLG com função de ligação igual a inversa de Φ .

Para este modelo, a curva para $\pi(\mathbf{x})$ ou para $[1 - \pi(\mathbf{x})]$, quando $\beta_1 < 0$ tem a aparência da densidade da normal com média $-\beta_0/\beta_1$ e desvio padrão $\sigma = 1/|\beta_1|$. O modelo *probit* é não-linear nos parâmetros e restringe $\pi(\mathbf{x})$ ao intervalo (0,1).

2.3.2 Modelo de probabilidade linear

Quando a variável dependente é contínua é correto interpretarmos a regressão especificando $E(Y_i/\mathbf{x})$, porém esta interpretação pode não ser adequada quando Y_i é uma variável binária.

Seja o modelo de regressão

$$Y_i = \beta_0 + \beta_1 \mathbf{x}_{1i} + \epsilon_i, \text{ para } i = 1, \dots, n,$$

e chamado *modelo de probabilidade linear*, sendo $\mathbf{x} = (1, \mathbf{x}_1)$. A esperança condicional de Y_i dado \mathbf{x} , $E(Y_i/\mathbf{x})$, pode ser interpretada como a probabilidade condicional de que o evento ocorrerá dado \mathbf{x} ; ou seja, $\Pr(Y_i = 1/\mathbf{x})$ (Agresti, 2002).

Admitindo que $E(\epsilon_i) = 0$, obtém-se

$$E(Y_i/\mathbf{x}) = \pi_i(\mathbf{x}) = \beta_0 + \beta_1 \mathbf{x}_{1i}.$$

Quando as observações em Y_i são independentes, este modelo é um MLG com função de ligação identidade.

O modelo de probabilidade linear apresenta um problema estrutural, que é o fato da probabilidade $\pi(\mathbf{x})$ não ser restrita no intervalo (0,1), para valores de \mathbf{x} suficientemente pequenos ou grandes. Uma mudança em \mathbf{x} pode ter pequeno impacto quando $\pi(\mathbf{x})$ é próximo de 0 ou 1 do que quando $\pi(\mathbf{x})$ é próximo de 0.5. Usualmente é esperado um relacionamento não-linear entre $\pi(\mathbf{x})$ e \mathbf{x} (Agresti, 2002). Os modelos logit e probit, asseguram que as probabilidades estimadas se situam de fato entre os limites 0 e 1.

2.4 Modelos para variáveis de resposta multinomial

Quando a resposta de um indivíduo em um estudo é uma dentre várias respostas possíveis, dizemos que a variável resposta é policotômica. Os c valores possíveis da variável resposta Y são chamados de categorias de respostas. Existem modelos estatísticos para diversos tipos de respostas policotômicas ou escalas de medidas. Se as categorias de respostas têm alguma ordenação, esta deve ser considerada utilizando diferentes classes de funções de ligação para tipos diferentes de escalas de respostas. Quando as categorias são ordenadas, em algumas situações, é preferível trabalhar com probabilidades acumuladas para as respostas

$$\gamma_1 = \pi_1, \quad \gamma_2 = \pi_1 + \pi_2, \quad \dots, \quad \gamma_c \equiv 1, \quad k = 1, \dots, c.$$

Portanto, é possível obter modelos baseados em probabilidades acumuladas $\gamma_k = \Pr(Y \leq k)$ em vez das probabilidades de cada categoria π_k .

2.4.1 Distribuição multinomial

A distribuição multinomial é uma maneira natural de modelar a distribuição da variável resposta policotômica que pode ser considerada como uma generalização da distribuição binomial. Suponha que indivíduos de uma mesma população de interesse possuem uma e somente uma das c categorias. Se a população é efetivamente grande e se uma amostra aleatória simples de tamanho n for tomada, então o número de indivíduos observados possuindo os atributos de 1 até c segue a distribuição multinomial.

Seja o vetor de resposta denotado por Y , onde $Y = (Y_1, \dots, Y_c)^T$, com c categorias de resposta ($k = 1, \dots, c$), em uma amostra aleatória de n indivíduos ($i = 1, \dots, n$), então

$$\begin{aligned} \Pr(Y_1 = y_1, \dots, Y_c = y_c; n, \pi) &= \binom{n}{\mathbf{y}} \pi_1^{y_1} \dots \pi_c^{y_c} \\ &= \frac{n!}{y_1! \dots y_c!} \pi_1^{y_1} \dots \pi_c^{y_c} \\ &= \frac{n!}{\prod_{k=1}^c y_k!} \prod_{k=1}^c \pi_k^{y_k}, \end{aligned}$$

com $\pi = (\pi_1, \dots, \pi_c)$ são as probabilidades de sucesso de cada categoria, tais que $\sum_{k=1}^c \pi_k = 1$, $\pi_k = \mu_k/\mu$, $\mu = \sum_{k=1}^c \mu_k$ e $\sum_{k=1}^c y_k = n$. Tomando o logaritmo, temos que

$$\begin{aligned} l(n, \pi) &= \log\left(\frac{n!}{\prod_{k=1}^c y_k!}\right) + \sum_{k=1}^c y_k \log \pi_k \\ &= \sum_{k=1}^c y_k \log \pi_k + \log n! - \sum_{k=1}^c \log y_k!. \end{aligned}$$

Portanto, esta distribuição está na família exponencial natural.

A distribuição multinomial surge também a partir de variáveis aleatórias com distribuição Poisson.

Seja $Y_{ik} \sim P(\mu_{ik})$ para $i = 1, \dots, n$ e $k = 1, \dots, c$ então

$$e^{-\mu_{ik}} \mu_{ik}^{y_{ik}} / y_{ik}! = \exp\{y_{ik} \log \mu_{ik} - \mu_{ik} - \log(y_{ik}!)\}$$

e, portanto,

$$\begin{aligned} \prod_i^n \prod_k^c \frac{e^{-\mu_{ik}} (\mu_{ik})^{y_{ik}}}{y_{ik}!} &= \prod_i^n \prod_k^c \frac{e^{-\mu_{ik}} (\mu_{ik})^{y_{ik}}}{y_{ik}!} \frac{y_{i\cdot}! (\mu_{i\cdot})^{y_{i\cdot}}}{y_{i\cdot}! (\mu_{i\cdot})^{y_{i\cdot}}} \\ &= \prod_i^n \frac{e^{-\sum_k^c \mu_{ik}} (\mu_{i\cdot})^{y_{i\cdot}}}{y_{i\cdot}!} \prod_i^n \prod_k^c \frac{(\mu_{ik})^{y_{ik}}}{y_{ik}!} \prod_i^n \frac{y_{i\cdot}!}{(\mu_{i\cdot})^{y_{i\cdot}}}, \end{aligned}$$

fazendo $m_i = y_{i\cdot}$ e $\tau_i = \mu_{i\cdot}$, temos

$$\underbrace{\prod_i^n \frac{e^{-\tau_i} \tau_i^{m_i}}{m_i!}}_A \underbrace{\prod_i^n y_{i\cdot}! \prod_k^c \frac{(\frac{\mu_{ik}}{\mu_{i\cdot}})^{y_{ik}}}{y_{ik}!}}_B.$$

A distribuição se decompõe em dois produtos o primeiro fator (A) corresponde a uma distribuição conjunta de variáveis aleatórias com distribuição de Poisson de parâmetros τ_i .

O segundo fator (B) corresponde a uma forma multinomial, isto é, da distribuição de uma amostra de variáveis aleatórias com distribuição multinomial de parâmetros $y_{i\cdot}$ e $(\pi_{i1}, \dots, \pi_{ic})$ com $\pi_{ik} = \frac{\mu_{ik}}{\mu_{i\cdot}}$.

Com a distribuição de $\sum_{k=1}^c Y_{ik} \sim P(\tau_i)$ da expressão anterior obtém-se que a distribuição condicional de $Y_i = (Y_{i1}, \dots, Y_{ic})$ dado o valor de n , em que $n = \sum_{k=1}^c y_{ik}$, que está fixo, é multinomial com parâmetros $\pi_{ik} = \mu_{ik} / \mu_{i\cdot}$.

A relação entre modelos loglineares para frequências e modelos de resposta multinomial para proporções origina do fato de que distribuições binomial e multinomial podem ser obtidas a partir de um grupo de variáveis aleatórias independentes de Poisson, determinando previamente o tamanho da amostra n .

2.4.2 Modelo loglinear

Seja agora $\pi_{ik} = \mu_{ik} / \sum_{i,k} \mu_{ik}$ e considere o modelo loglinear, do tipo (2.10) para cada categoria, isto é,

$$\log \mu_{ik} = \alpha_i^* + \mathbf{x}'_{ik} \beta, \quad (2.14)$$

para $i = 1, \dots, n$ e $k = 1, \dots, c$ e, portanto, $\mu_{ik} = \exp(\alpha_i^* + \mathbf{x}'_{ik}\beta)$ onde $\mu_{ik} = E(Y_{ik})$, \mathbf{x}_{ik} são vetores p -dimensional conhecidos, β é o vetor $(p \times 1)$ dos parâmetros de interesse e $\alpha_1^*, \dots, \alpha_n^*$ são parâmetros de perturbação (“nuisance”). Sob este modelo a dimensão do espaço paramétrico aumenta, $(n + p)$, quando $n \rightarrow \infty$ para p fixado. Conseqüentemente nas estimativas de máxima verossimilhança não podem ser garantidas a eficiência e consistência no limite quando $n \rightarrow \infty$. Por outro lado o logaritmo da verossimilhança condicional obtida abaixo depende somente de β e não de α^* e da teoria assintótica padrão aplicada diretamente para a verossimilhança condicional (McCullagh & Nelder, 1989).

E portanto, tomando o logaritmo da verossimilhança da variável Y_{ik} com distribuição de Poisson, temos que

$$\begin{aligned}
L(\alpha^*, \beta) &= \sum_{i=1}^n \sum_{k=1}^c \{y_{ik} \log \mu_{ik} - \mu_{ik}\} \\
&= \sum_{i=1}^n \sum_{k=1}^c y_{ik} \log \mu_{ik} - \sum_{i=1}^n \sum_{k=1}^c \mu_{ik} \\
&= \sum_{i=1}^n \sum_{k=1}^c y_{ik} (\alpha_i^* + \mathbf{x}'_{ik}\beta) - \sum_{i=1}^n \sum_{k=1}^c \exp(\alpha_i^* + \mathbf{x}'_{ik}\beta) \\
&= \sum_{i=1}^n \sum_{k=1}^c y_{ik} \alpha_i^* + \sum_{i=1}^n \sum_{k=1}^c y_{ik} \mathbf{x}'_{ik}\beta - \sum_{i=1}^n \sum_{k=1}^c \exp(\alpha_i^* + \mathbf{x}'_{ik}\beta) \\
&= \sum_{i=1}^n y_i \alpha_i^* + \sum_{i=1}^n \sum_{k=1}^c y_{ik} \mathbf{x}'_{ik}\beta - \sum_{i=1}^n \sum_{k=1}^c \exp(\alpha_i^* + \mathbf{x}'_{ik}\beta).
\end{aligned}$$

Fazendo $m_i = y_i$. para o total da i -ésima linha e a transformação do parâmetro

$$\begin{aligned}
\tau_i &= \sum_{k=1}^c \mu_{ik} = \sum_{k=1}^c \exp(\alpha_i^* + \mathbf{x}'_{ik}\beta) \\
&= \sum_{k=1}^c \exp(\alpha_i^*) \exp(\mathbf{x}'_{ik}\beta) \\
&= \exp(\alpha_i^*) \sum_{k=1}^c \exp(\mathbf{x}'_{ik}\beta),
\end{aligned}$$

e, portanto,

$$\exp(\alpha_i^*) = \frac{\tau_i}{\sum_{k=1}^c \exp(\mathbf{x}'_{ik}\beta)},$$

ou ainda,

$$\alpha_i^* = \log \tau_i - \log \left[\sum_{k=1}^c \exp(\mathbf{x}'_{ik}\beta) \right]$$

Considerando agora o logaritmo da verossimilhança como uma função de (τ, β) , obtém-se

$$\begin{aligned}
L(\tau, \beta) &= \sum_{i=1}^n \left\{ m_i \left(\log \tau_i - \log \left[\sum_{k=1}^c \exp(\mathbf{x}'_{ik} \beta) \right] \right) + \sum_{k=1}^c y_{ik} \mathbf{x}'_{ik} \beta - \tau_i \right\} \\
&= \sum_{i=1}^n (m_i \log \tau_i - \tau_i) + \sum_{i=1}^n \left\{ \sum_{k=1}^c y_{ik} \mathbf{x}'_{ik} \beta - m_i \log \left(\sum_{k=1}^c \exp(\mathbf{x}'_{ik} \beta) \right) \right\} \\
&= L_m(\tau; m) + L_{Y|m}(\beta, y).
\end{aligned}$$

O primeiro termo acima é o logaritmo da verossimilhança de distribuição de Poisson para τ baseado nos totais de linhas $Y_i \sim P(\tau_i)$. O segundo termo é o logaritmo da verossimilhança condicional dado $\{Y_i\}$, o qual depende somente de β e não do parâmetro de perturbação. Todas as informações referentes à β encontram-se no segundo termo. Em particular é evidente que $\hat{\beta}$ e $\text{cov}(\hat{\beta})$ baseados em $L_{Y|m}(\beta, y)$ são idênticos àqueles baseados no logaritmo da verossimilhança completa. Portanto, as probabilidades obtidas de um modelo de resposta multinomial são calculadas da forma

$$\pi_{ik} = \mu_{ik} / \sum_{i,k} \mu_{ik} = \frac{\exp(\mathbf{x}'_{ik} \beta)}{\sum_{k=1}^c \exp(\mathbf{x}'_{ik} \beta)}.$$

A equivalência descrita acima entre os modelos loglineares e o modelo de resposta multinomial depende muito da suposição de que os parâmetros τ_i e daí α_i^* sejam irrestritos exceto pela desigualdade necessária $\tau_i \geq 0$.

2.4.3 Logits para respostas ordinais

Dados provenientes de variáveis respostas ordinais freqüentemente surgem na prática. Em geral os modelos são baseados em probabilidades de respostas acumuladas $[\Pr(Y \leq k)] = \gamma_k(\mathbf{x})$, onde $\gamma_k(\mathbf{x}) = \pi_1(\mathbf{x}) + \dots + \pi_k(\mathbf{x})$, em vez de probabilidades das categorias $\pi_k(\mathbf{x})$. Os dois grupos de probabilidades são equivalentes, porém modelos simples para as probabilidades acumuladas têm propriedades ótimas para respostas de escala ordinal mais do que modelos simples baseados em probabilidades das categorias. Em particular, modelos lineares usando escala logística, $\log \left[\frac{\gamma_k(\mathbf{x})}{1 - \gamma_k(\mathbf{x})} \right]$, ou escala complemento log-log, $\log \{ -\log(1 - \gamma_k(\mathbf{x})) \}$ são encontrados freqüentemente em trabalhos práticos (McCullagh, 1980).

Quando as categorias de respostas tem uma ordenação natural, os modelos logit podem incorporar esta ordenação indiretamente na sua construção (Agresti, 2002). Uma das maneiras de usar categorias de resposta com ordenação é formando logits de probabilidades acumuladas.

Logits acumulados

Os modelos mais populares para resposta ordinal são generalizações multicategorizadas da regressão logística (2.10) e este modelo usa logits de probabilidades acumuladas, chamados *logits acumulados* (Agresti & Natarajan, 2001). Para a variável resposta Y_i sendo agora ordinal c -categórica (isto é, com c categorias), $i = 1, \dots, n$ e $\mathbf{x}_i = (\mathbf{x}_{1i}, \dots, \mathbf{x}_{pi})'$ vetor de variáveis explicativas com o correspondente vetor de parâmetros $\beta = (\beta_1, \dots, \beta_p)'$, o modelo tem a forma

$$\text{logit}[\Pr(Y_i \leq k)] = \text{logit}[\gamma_k(\mathbf{x}_i)] = \alpha_k^* + \mathbf{x}_i' \beta, \quad (2.15)$$

$k = 1, \dots, c - 1$, e denotando $\Pr(Y_i = k) = \pi_k(\mathbf{x}_i)$, tal que $\sum_{k=1}^c \pi_k = 1$ e $[\Pr(Y_i \leq k)] = \gamma_k(\mathbf{x}_i)$, onde $\gamma_k(\mathbf{x}_i) = \pi_1(\mathbf{x}_i) + \dots + \pi_k(\mathbf{x}_i)$, então temos que o logit acumulado é definido como

$$\begin{aligned} L_k &= \text{logit}[\gamma_k(\mathbf{x})] = \log \left[\frac{\gamma_k(\mathbf{x})}{1 - \gamma_k(\mathbf{x})} \right] \\ &= \log \left[\frac{\pi_1(\mathbf{x}) + \dots + \pi_k(\mathbf{x})}{\pi_{k+1}(\mathbf{x}) + \dots + \pi_c(\mathbf{x})} \right], \end{aligned}$$

$k = 1, \dots, c - 1$, cada logit acumulado usa todas as c categorias de respostas.

Notemos que o modelo (2.15) assume efeito idêntico β nos preditores para cada probabilidade acumulada; este caso é referido como *modelo de "odds"* (chances) proporcionais, e que cada logit acumulado tem seu próprio intercepto. O α_k^* aumentam em k , uma vez que $\Pr(Y_i \leq k | \mathbf{x}_i)$ aumenta em k para \mathbf{x}_i fixo.

A razão de chances em $\mathbf{x} = \mathbf{x}_1$ e $\mathbf{x} = \mathbf{x}_2$ é dada por

$$\frac{\gamma_k(\mathbf{x}_1)/(1 - \gamma_k(\mathbf{x}_1))}{\gamma_k(\mathbf{x}_2)/(1 - \gamma_k(\mathbf{x}_2))} = \exp\{\beta'(\mathbf{x}_1 - \mathbf{x}_2)\}, \quad (2.16)$$

o qual não depende da escolha da categoria k . A razão de chances das probabilidades acumuladas é chamada *razão de chances acumuladas*. E, por exemplo, se \mathbf{x} é uma variável indicadora para dois grupos de tratamento, T_1 e T_2 , logo, (2.16) pode ser escrito como

$$\frac{\text{odds}(Y \leq k | T_1)}{\text{odds}(Y \leq k | T_2)} = \exp\{\beta'(\mathbf{x}_1 - \mathbf{x}_2)\}, \quad k = 1, \dots, c - 1,$$

onde $\beta'(\mathbf{x}_1 - \mathbf{x}_2)$ mede o efeito do tratamento. Ou seja, a chance de tomar resposta menor ou igual que k em $\mathbf{x} = \mathbf{x}_1$ são $\exp\{\beta'(\mathbf{x}_1 - \mathbf{x}_2)\}$ vezes a chance de $\mathbf{x} = \mathbf{x}_2$. O logaritmo da razão de chances acumuladas é proporcional à distância entre \mathbf{x}_1 e \mathbf{x}_2

Logits de categorias-adjacentes

Modelos alternativos que utilizam probabilidades categóricas simples em vez de probabilidades acumuladas são os modelos logits de categorias-adjacentes e os de categorias de base.

Sejam $\{\pi_1(\mathbf{x}), \dots, \pi_c(\mathbf{x})\}$ as probabilidades de respostas para um grupo de variáveis explicativas do vetor \mathbf{x} .

Se o modelo ordinal usa os logits de categorias-adjacentes, a expressão é

$$\log[\pi_k(\mathbf{x})/\pi_{k+1}(\mathbf{x})], \quad k = 1, \dots, c - 1.$$

Estes logits formam um grupo básico dos logits de base categorizada.

O modelo que considera como base a última categoria c é,

$$\log[\pi_k(\mathbf{x})/\pi_c(\mathbf{x})], \quad k = 1, \dots, c - 1.$$

Logits de razão-continuação

Outra possibilidade é o modelo logit razão-continuação, o qual usa logits da forma

$$\log\left[\frac{\pi_k(\mathbf{x})}{\pi_{k+1}(\mathbf{x}) + \dots + \pi_c(\mathbf{x})}\right]$$

ou

$$\log\left[\frac{\pi_{k+1}(\mathbf{x})}{\pi_1(\mathbf{x}) + \dots + \pi_k(\mathbf{x})}\right].$$

O modelo logit razão-continuação é útil quando um mecanismo sequencial determina a variável resposta, tais como sobrevivência através de vários períodos de idade.

Capítulo 3

Modelos para dados longitudinais

3.1 Considerações iniciais

Na estatística univariada, uma suposição básica é que cada um dos sujeitos, ou unidade de investigação, ocasione uma única medida de alguma variável de interesse. Na estatística multivariada a medida simples de cada indivíduo é substituída por um vetor de medidas.

Estudos longitudinais, em que indivíduos são medidos repetidamente através do tempo, estão em contraste com o estudo corte-transversal, em que uma simples resposta é medida para cada indivíduo. Dados longitudinais podem ser obtidos tanto prospectivamente, seguindo os sujeitos sobre o tempo, ou retrospectivamente, extraindo as múltiplas medidas de cada indivíduo a partir de registros históricos (Diggle *et al.*, 2001). Sociólogos, economistas e administradores freqüentemente se referem a estudos longitudinais como *estudos em painel*. É importante salientar que o tipo de problema considerado sob a denominação de análise de dados longitudinais difere daquele comumente tratado na literatura estatística por *análise de séries temporais* pelo fato de que, neste caso, em geral, dispomos de uma única unidade amostral com muitas observações ao longo do tempo (por exemplo, 100 ou mais), ao passo que naquele, lidamos com várias unidades amostrais (por exemplo, 5 ou mais) observadas em poucos instantes (por exemplo, 2 a 10) (Singer & Andrade, 1986).

Conjuntos de dados longitudinais são formados por observações ao longo do tempo de uma variável resposta (Y) e um grupo de covariáveis para cada um dos sujeitos, (ou unidade de investigação), como mostrado no Quadro 3.1, em que \mathbf{x}_{it} é o vetor ($p \times 1$) das covariáveis no tempo t , com $t = 1, \dots, T$ do i -ésimo sujeito com $i = 1, \dots, n$.

Quadro 3.1: Estrutura de um conjunto de dados longitudinais.

Unidade Amostral	Variáveis explicativas				Resposta			
	1	2	...	T	1	2	...	T
1	\mathbf{x}_{11}	\mathbf{x}_{12}	...	\mathbf{x}_{1T}	Y_{11}	Y_{12}	...	Y_{1T}
2	\mathbf{x}_{21}	\mathbf{x}_{22}	...	\mathbf{x}_{2T}	Y_{21}	Y_{22}	...	Y_{2T}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	\mathbf{x}_{n1}	\mathbf{x}_{n2}	...	\mathbf{x}_{nT}	Y_{n1}	Y_{n2}	...	Y_{nT}

A principal vantagem de um estudo longitudinal é sua eficácia para estudar a mudança ao

longo do tempo. Outro mérito do estudo longitudinal é sua capacidade para distinguir o grau de variação em Y sobre o tempo para uma pessoa a partir da variação em Y entre as pessoas. A análise de dados longitudinais, como outros métodos estatísticos, tem dois componentes: análise exploratória e a análise confirmatória. A análise exploratória compreende técnicas para visualizar padrões nos dados e a análise confirmatória evidencia a favor ou contra as hipóteses (Diggle *et al.*, 2001). Extensões de modelos lineares generalizados (MLGs) para análise de dados longitudinais discretos ou contínuos podem ser usados. Com modelos não lineares para dados discretos, tais como regressão logística, diferentes suposições sobre a fonte de correlação pode levar a coeficientes de regressão com distintas interpretações. Aqui será discutido MLGs para dados longitudinais com enfoque em modelos *marginais*.

Um possível objetivo da análise de dados longitudinais é descrever a esperança marginal da variável resposta como função das variáveis preditoras (Zeger & Liang, 1986). Tipicamente o interesse é explicar o padrão de mudanças sobre o tempo ou a dependência das respostas em relação às covariáveis. Uma dificuldade com análise de dados longitudinais não-gaussianos é a falta de uma rica classe de modelos tais como o gaussiano multivariado para a distribuição conjunta de y_{it} ($t = 1, \dots, T$). Devido a essa falta de uma distribuição multivariada (conjunta) para dados não gaussianos das medidas repetidas, é conveniente se utilizar o método de quase-verossimilhança, que nesse caso é suficiente poucas suposições sobre a distribuição da variável dependente, que é o conhecimento do relacionamento entre a média e a variância da variável resposta, a função de ligação e o preditor linear da forma usual do MLG.

Liang & Zeger (1986) desenvolveram um método baseado em uma extensão dos modelos lineares generalizados para observações longitudinais. Eles introduziram uma classe de equações de estimação as quais permitem obter estimativas consistentes dos parâmetros de regressão e suas variâncias sob condições regulares. Estes métodos são semi-paramétricos, em que as equações de estimação são derivadas sem a especificação completa da distribuição conjunta das observações dos sujeitos. Em vez disso, eles especificam a verossimilhança para as distribuições marginais (univariadas) e uma matriz de covariância de “trabalho” para os vetores das observações (Ware, Lipsitz & Speizer, 1988).

Seja y_{it} o valor observado da variável resposta e \mathbf{x}_{it} o vetor ($p \times 1$) de covariáveis observadas no tempo t , com $t = 1, \dots, T$ do i -ésimo sujeito com $i = 1, \dots, n$. A média e variância de Y_{it} são representadas, respectivamente, por $E(Y_{it}) = \mu_{it}$ e $\text{Var}(Y_{it}) = a(\phi)V_{it}$.

Seja $\mathbf{Y}_i = (y_{i1}, \dots, y_{iT})'$ um vetor das respostas ($T \times 1$) e $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT})'$ uma matriz ($T \times p$) para o i -ésimo sujeito (cada unidade de investigação, $i = 1, \dots, n$). Então podemos considerar que a densidade marginal de y_{it} tem uma distribuição exponencial da forma

$$f(y_{it}, \theta_{it}, \phi) = \exp\{[y_{it}\theta_{it} - b(\theta_{it})]/a(\phi) + c(y_{it}, \phi)\},$$

onde $\theta_{it} = g(\eta_{it})$, $\eta_{it} = \mathbf{x}'_{it}\beta$. Assim, para esta formulação, os dois primeiros momentos de

y_{it} são dados, respectivamente, por

$$E(\mathbf{Y}_{it}) = \frac{\partial b(\theta_{it})}{\partial \theta_{it}} \quad \text{e} \quad \text{Var}(\mathbf{Y}_{it}) = \frac{\partial^2 b(\theta_{it})}{\partial \theta_{it}^2} a(\phi).$$

Ou seja, é uma extensão dos modelos lineares generalizados apresentados em (2.3).

Com uma observação no tempo para cada unidade experimental, a modelagem é feita em relação à média populacional de Y , chamada de resposta média marginal. Com medidas repetidas existem vários enfoques diferentes que podem ser adotados. Uma simples e freqüente estratégia é: primeiro reduzir os valores repetidos dentro de uma ou duas medidas sumárias em seguida, analisar cada variável sumária como uma função das covariáveis.

Em vez de reduzir as respostas repetidas em estatísticas sumárias, pode-se modelar Y_{it} em termos de \mathbf{x}_{it} . O modelo se refere a média marginal, visto que medidas repetidas não são provavelmente independentes, esta análise marginal pode incluir suposições sobre a forma da correlação. O enfoque do modelo marginal tem a vantagem de separadamente modelar a média e covariância e tem-se ainda inferências válidas sobre β que podem ser feitas mesmo quando uma forma incorreta de variância é assumida.

Harville (1977) e Laird & Ware (1982) desenvolveram os chamados modelos de efeitos aleatórios no qual observações repetidas para um sujeito são assumidos compartilharem de um componente aleatório comum. Azzalini (1984) discutiu modelos em que a estrutura autoregressiva do erro foi assumida.

Poucas técnicas têm sido disponíveis quando a resposta não é gaussiana. Modelos de efeitos aleatórios para respostas binárias são a exceção. Koch *et al.* (1977), os chamados modelos log-lineares e Ochi & Prentice (1984), um modelo probit; Stiratelli, Laird & Ware (1984) e Anderson & Aitkin (1985) desenvolveram um modelo logístico admitindo covariáveis dependentes do tempo.

Com uma simples observação no tempo ($t = 1$) para cada sujeito ($i = 1, \dots, n$), um modelo linear generalizado pode ser aplicado para obter uma descrição para uma variedade de variáveis respostas contínuas ou discretas. Com observações repetidas ($t > 1$), contudo, a correlação entre valores observados para um dado sujeito pode ser levado em consideração.

3.2 Método de quase-verossimilhança

O método de quase-verossimilhança foi proposto por Wedderburn (1974) e posteriormente examinado extensivamente por McCullagh & Nelder (1989). Este é um método usado em regressão que requer poucas suposições sobre a distribuição da variável dependente e daí pode ser usado com uma variedade de variáveis respostas. Na análise de verossimilhança devemos especificar a exata forma da distribuição, na quase-verossimilhança, especificamos

somente o relacionamento entre a média da resposta como função das covariáveis e a covariância, isto é, especifica-se somente a estrutura da média e covariância, ou equivalentemente, especifica-se o modelo pela média e uma função de variância que expresse a dependência da variância com a média. É possível desenvolver métodos que são aplicáveis para diversos tipos de variáveis respostas. A distribuição da variável resposta ficará determinada quando a função de variância escolhida coincidir com a função de variância de alguma distribuição da família exponencial (Paula, 2003).

Considerando uma única observação no tempo, para todo i e nesse caso retirando o índice t , podemos tratar cada unidade de investigação como um escalar, dessa forma a quase-verossimilhança pode ser aplicada no enfoque de regressão.

Seja μ_i a esperança de Y_i e suponha que

$$\mu_i = h(\eta_i), \quad (3.1)$$

com

$$\eta_i = \mathbf{X}_i' \beta, \quad (3.2)$$

onde β é um vetor ($p \times 1$) de parâmetros e h é a função inversa de g que é a função de ligação (McCullagh & Nelder, 1983b). Na quase-verossimilhança, a função de variância, $V(\mu)$, de Y_i é expressada como uma função conhecida da média.

Se Y é a variável aleatória de interesse, assumimos que

$$E(Y) = \mu(\beta) \quad \text{e} \quad \text{Var}(Y) = \sigma^2 V(\mu),$$

onde $V(\mu)$ é uma função conhecida da média μ e σ^2 é o parâmetro de dispersão. O logaritmo da função de quase-verossimilhança é definido por

$$Q(\mu, y) = \frac{1}{\sigma^2} \int_y^\mu \frac{y-t}{V(t)} dt, \quad (3.3)$$

e tem propriedades semelhantes ao logaritmo da função de verossimilhança usual (Paula, 2003), tais como

$$E\left\{ \frac{\partial Q(\mu; Y)}{\partial \mu} \right\} = 0 \quad (3.4)$$

e

$$E\left[\left\{ \frac{\partial Q(\mu; Y)}{\partial \mu} \right\}^2 \right] = -E\left\{ \frac{\partial^2 Q(\mu; Y)}{\partial \mu^2} \right\} \quad (3.5).$$

Ou ainda, o estimador de quase-verossimilhança é a solução do sistema de equação quase-score

$$Q(U_j) = \sum_{i=1}^n \frac{\partial \mu_i}{\partial \beta_j} V_i^{-1}(y_i - \mu_i) = 0, \quad j = 1, \dots, p. \quad (3.6)$$

As equações (3.6) são de fato equações escores para β quando Y_i tem distribuição a partir da família exponencial (Zeger & Liang, 1986).

3.3 Equações de estimação para respostas independentes

Considerando inicialmente a suposição básica de que observações repetidas a partir de um dado sujeito são independentes de um outro sujeito, surge então um estimador, $\hat{\beta}_I$, de β , cuja função escore da análise de verossimilhança é da forma

$$U_I(\beta) = \sum_{i=1}^n \mathbf{X}'_i \Delta_i C_i = 0, \quad (3.7)$$

onde $\Delta_i = \text{diag}(\partial\theta_{it}/\partial\eta_{it})$ é uma matriz $(T \times T)$ e $C_i = Y_i - b'_i(\theta) = Y_i - \mu_i$ é de ordem $(T \times 1)$ para o i -ésimo sujeito. O estimador $\hat{\beta}_I$ é definido como solução da equação (3.7).

Sob condições usuais de regularidade, Liang & Zeger (1986) mostraram que quando $n \rightarrow \infty$, $\hat{\beta}_I$ é um estimador consistente de β e que $n^{1/2}(\hat{\beta}_I - \beta)$ é assintoticamente gaussiano multivariado com matriz de covariância V_I dada por

$$\begin{aligned} \mathbf{V}_I &= \lim_{n \rightarrow \infty} n \left(\sum_{i=1}^n \mathbf{X}'_i \Delta_i A_i \Delta_i \mathbf{X}_i \right)^{-1} \left[\sum_{i=1}^n \mathbf{X}'_i \Delta_i \text{cov}(Y_i) \Delta_i \mathbf{X}_i \right] \left(\sum_{i=1}^n \mathbf{X}'_i \Delta_i A_i \Delta_i \mathbf{X}_i \right)^{-1} \\ &= \lim_{n \rightarrow \infty} n \{H_1(\beta)\}^{-1} H_2(\beta) \{H_1(\beta)\}^{-1}, \end{aligned}$$

com $H_1(\beta) = \sum_{i=1}^n \mathbf{X}'_i \Delta_i A_i \Delta_i \mathbf{X}_i$ e $H_2(\beta) = \sum_{i=1}^n \mathbf{X}'_i \Delta_i \text{cov}(Y_i) \Delta_i \mathbf{X}_i$, onde os cálculos dos momentos para os Y_i 's são tomados com respeito ao verdadeiro modelo subjacente (básico) e $A_i = \text{diag}\{b''(\theta_{it})\} = \text{diag } V_i$ a matriz diagonal $(T \times T)$ para cada i . Portanto, a variância de $\hat{\beta}_I$ pode ser estimada consistentemente por

$$\text{Var}(\hat{\beta}_I) = \{H_1(\hat{\beta}_I)\}^{-1} \left(\left[\sum_{i=1}^n \mathbf{X}'_i \Delta_i C_i C'_i \Delta_i \mathbf{X}_i \right]_{\hat{\beta}_I} \right) \{H_1(\hat{\beta}_I)\}^{-1}.$$

Os estimadores $\hat{\beta}_I$ e $\text{Var}(\hat{\beta}_I)$ são consistentes no caso em que seja oferecido a especificação correta do modelo de regressão. Note que isto requer que os dados faltosos tenham desaparecido segundo um mecanismo completamente aleatório no sentido de Rubin (1976). O estimador $\hat{\beta}_I$ não tem alta eficiência nos casos onde a autocorrelação é grande. Portanto é proposto uma equação de estimação generalizada que leva a obter estimadores com alta eficiência que serão apresentados na próxima seção.

3.4 Equações de estimação generalizadas

Para a obtenção das estimativas dos parâmetros, quando os dados não são gaussianos, são disponíveis poucas técnicas. As equações de estimação generalizadas fornecem um razoável enfoque para obter as estimativas dos parâmetros sob poucas suposições sobre a dependência dentro do *cluster* para as respostas gaussianas e não-gaussianas. A análise pode ser escolhida a partir de três diferentes métodos para se estimar os parâmetros de regressão β e os associados com parâmetros de correlação α . GEEs de primeira ordem (GEE1 - Liang & Zeger (1986)) trata α como um parâmetro de perturbação (“nuisance”) e o foco primário é obter β . GEEs de segunda ordem (GEE2 - Prentice & Zhao (1991)) estima α e β conjuntamente. Regressão logística alternativa estendida (ALR - Carey *et al.* (1993)) substitui a equação de estimação para α no método GEE1 por uma equação de estimação não linear não viesada e oferece alta eficiência na estimação de ambos α e β . Os erros-padrão para todos os três métodos são calculados usando os estimadores da variância robusta ‘sandwich’. O método GEE2 estima os parâmetros de associação α mais precisamente, contudo ele tem a desvantagem de que a consistência de β depende de ter especificado o modelo correto para a razão de chances global, e que ele é computacionalmente pesado tornando-se inviável quando os dados dos grupos tornam-se grande. Assim, em situações onde a inferência referente a β é primária ou quando a estimação usando o método GEE2 é intratável, o GEE1 ou ALR podem ser mais apropriados (Huang *et al.*, 2002). O método de equações de estimação generalizadas (GEE) é assim referido pois as estimativas dos parâmetros são obtidas pela solução das equações de estimação que são equações escores.

Segundo Hardin & Hilbe (2002) todos os modelos GEE consideram uma equação de estimação que é escrita em duas partes. A primeira parte estima os parâmetros de regressão e a segunda estima os parâmetros de associação ou os parâmetros de distribuição da variância de segunda ordem.

Qualquer modelo GEE que assume ortogonalidade das equações de estimação para os parâmetros de regressão e parâmetros de associação é denominado GEE1. Os modelos GEE2 são uma generalização dos modelos GEE1.

3.4.1 Estimação de α e ϕ

Definimos

$$\mathbf{V}_i = A_i^{\frac{1}{2}} R(\alpha) A_i^{\frac{1}{2}} \phi, \quad (3.8)$$

sendo \mathbf{V}_i igual a $\text{cov}(Y_i)$ se $R(\alpha)$ é de fato a verdadeira matriz de correlação para os Y_i 's. Aqui, n_i é o número de observações feitas no tempo, para o i -ésimo indivíduo, assim $n_i = T$ onde $t = (1, \dots, T)$. $R(\alpha)$ é uma matriz simétrica ($n_i \times n_i$) a qual satisfaz o requerimento de ser uma matriz de correlação, sendo α um vetor de parâmetros ($q \times 1$) o qual caracteriza

completamente $R(\alpha)$, que é chamada de matriz de correlação de “trabalho”.

O vetor de parâmetros de correlação α , e o parâmetro de escala, ϕ , podem ser estimados a partir dos resíduos de Pearson.

Seja \hat{r}_{it} o resíduo de Pearson definido por

$$\hat{r}_{it} = \frac{(y_{it} - \hat{\mu}_{it})}{\sqrt{V(\hat{\mu}_{it})}}.$$

A estimação de ϕ pode ser definida por

$$\hat{\phi} = \sum_{i=1}^n \sum_{t=1}^{n_i} \hat{r}_{it}^2 / (N - p),$$

onde $N = \sum n_i = n \times T$.

O estimador de α depende da escolha de $R(\alpha)$. Em geral uma estimativa consistente é dada pela função

$$\hat{R}_{uv} = \sum_{i=1}^n \hat{r}_{iu} \hat{r}_{iv} / (N - p).$$

Tipos de $R(\alpha)$ diferentes, levam cada uma a distintas análises. O estimador de α varia de caso a caso. Existe uma variedade de estruturas simples que podem ser apropriadas para o uso do modelo da matriz de correlação de trabalho, embora que para qualquer $R(\alpha)$, $\hat{\beta}_G$ e \hat{V}_G serão consistentes. Contudo a escolha de $R(\alpha)$, próxima da verdadeira correlação, aumenta a eficiência. Segundo Horton & Lipsitz (1999) quando o número de observações no tempo por sujeito (unidade de investigação) ou “cluster” é pequeno, em um desenho balanceado e completo, então a matriz não estruturada é recomendada. Para dados com medidas mistas, pode ser razoável considerar o modelo onde a correlação é uma função do tempo entre observações (isto é, M-dependente ou auto-regressivo). Para dados com observações agrupadas de alguma maneira, pode não ser lógico ordenar as observações dentro de um “cluster” (grupo) e uma estrutura permutável pode ser mais apropriada.

As principais matrizes de estruturas de correlação de “trabalho” utilizadas são:

1. Independente (“Independence”) - Considera as respostas independentes numa mesma unidade. Isto implica que quando $R(\alpha) = I$, a matriz identidade e nesse caso temos as equações de estimação para respostas independentes. E portanto, $R(\alpha)$ assume a forma

$$R(\alpha) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}_{n_i \times n_i}. \quad (3.9)$$

Nesta situação, o modelo é conhecido como de regressão múltipla.

2. Auto-regressiva (“AR1”) - Se a distribuição da variável resposta é normal, esta estrutura de correlação corresponde ao modelo auto regressivo de 1^a ordem, $\text{corr}(Y_{it}, Y_{it'}) = \alpha^{|t-t'|}$. Como $E(\hat{r}_{it}\hat{r}_{it'}) \equiv \alpha^{|t-t'|}$, a estimação de α é feita pela inclinação da regressão de $\log(\hat{r}_{it}\hat{r}_{it'})$ em $\log(|t-t'|)$ e o parâmetro ϕ deve ser estimado para a determinação de $\hat{\beta}_G$ e \hat{V}_G .

$$R(\alpha) = \begin{pmatrix} 1 & \alpha & 0 & \cdots & 0 \\ \alpha & 1 & \alpha & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & \alpha & 1 \end{pmatrix}_{n_i \times n_i} . \quad (3.10)$$

3. Não estruturada (“Unstructured”) - Todos os valores de correlação são diferentes e portanto temos $\frac{1}{2}n_i(n_i - 1)$ diferentes parâmetros para estimar, e então

$$R(\alpha) = \begin{pmatrix} 1 & \alpha_{12} & \cdots & \alpha_{1n_i} \\ \alpha_{12} & 1 & \cdots & \alpha_{2n_i} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1n_i} & \alpha_{2n_i} & \cdots & 1 \end{pmatrix}_{n_i \times n_i} . \quad (3.11)$$

Nesse caso $R(\alpha)$ pode ser estimada por

$$(\phi n)^{-1} \sum_{i=1}^n \hat{A}_i^{-1/2} \hat{C}_i \hat{C}_i' \hat{A}_i^{-1/2} .$$

4. Uniforme ou permutável (“Exchangeable”) - A correlação entre quaisquer variáveis é a mesma, $\text{corr}(Y_{it}, Y_{it'}) = \alpha$ para todo $t \neq t'$, ou seja, todas as variâncias são iguais e todas as covariâncias são iguais, isto é,

$$R(\alpha) = \begin{pmatrix} 1 & \alpha & \cdots & \alpha \\ \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \cdots & 1 \end{pmatrix}_{n_i \times n_i} . \quad (3.12)$$

Dado ϕ , α pode ser estimado por

$$\hat{\alpha} = \phi^{-1} \sum_{i=1}^n \sum_{t>t'} \hat{r}_{it}\hat{r}_{it'} / \left\{ \sum_{i=1}^n \frac{1}{2}n_i(n_i - 1) - p \right\} .$$

5. M -dependente (“ M -dependence”) - $\text{corr}(Y_{it}, Y_{i,t+l}) = \alpha_l$, $l = 1, \dots, M$, depende das M observações anteriores. Por exemplo, se $M = n_i - 1$, então

$$R(\alpha) = \begin{pmatrix} 1 & \alpha_1 & \cdots & \alpha_{n_i-1} \\ \alpha_1 & 1 & \cdots & \alpha_{n_i-2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n_i-1} & \alpha_{n_i-2} & \cdots & 1 \end{pmatrix}_{n_i \times n_i} . \quad (3.13)$$

Um estimador natural de α_l , dado β e ϕ , é

$$\hat{\alpha}_t = \phi^{-1} \sum_{i=1}^n \hat{r}_{it} \hat{r}_{i,t+1} / (n - p).$$

Suponha $q = 1$ e $\alpha_t = \alpha$ ($t = 1, \dots, n_i - 1$). Um comum estimador de α pode ser

$$\hat{\alpha} = \sum_{i=1}^{n_i-1} \hat{\alpha}_t / (n_i - 1).$$

Uma extensão fácil é obtida para M -dependente.

6. Fixa (“Fixed”)- é especificada pelo pesquisador, ou seja tem a forma geral

$$R(\alpha) = \begin{pmatrix} 1 & r_{12}^* & \cdots & r_{1n_i}^* \\ r_{12}^* & 1 & \cdots & r_{2n_i}^* \\ \vdots & \vdots & \ddots & \vdots \\ r_{1n_i}^* & r_{2n_i}^* & \cdots & 1 \end{pmatrix}_{n_i \times n_i}, \quad (3.14)$$

onde r_{ts}^* é a correlação especificada pelo pesquisador.

A escolha da estrutura de correlação mais adequada em cada situação é a critério do pesquisador. Também em determinadas situações não existem grandes diferenças entre as estimativas obtidas segundo as diferentes estruturas de correlação. Horton & Lipsitz (1999) afirmam que comparações de estimativas e erros-padrão a partir de várias estruturas de correlação podem indicar sensibilidade na não especificação correta da estrutura de variância. Para a estrutura de trabalho de independência e a estrutura de trabalho fixa, não é preciso a estimação de α .

As equações de estimação gerais são

$$\begin{aligned} S_\beta(\beta, \alpha) &= \sum_{i=1}^n D_i^T \mathbf{V}_i^{-1} C_i \\ &= \sum_{i=1}^n U_i(\beta, \alpha) = 0, \end{aligned} \quad (3.15)$$

onde $D_i = \partial\{b'_i(\theta)\}/\partial\beta = \partial\mu_i/\partial\beta = A_i \Delta_i \mathbf{X}_i$. O sistema de equações (3.15) se reduz a um sistema de equações de respostas independentes (3.7) se especificamos $R(\alpha)$ como a matriz identidade. Notemos que $S_\beta(\beta, \alpha)$ depende também de α e ϕ que são estimados separadamente de β . A dependência de α pode ser resolvida substituindo α na GEE acima com uma estimativa $n^{\frac{1}{2}}$ -consistente, $\hat{\alpha}(\hat{\beta})$.

Para obter um estimador de β consideremos o mesmo enfoque dado por Nelder & Wedderburn (1972), onde para cada i , $U_i(\beta, \alpha) = D_i^T \mathbf{V}_i^{-1} C_i$ é similar para a função derivada

da quase-verossimilhança (3.7), enfoque defendido por Wedderburn (1974) e McCullagh & Nelder (1983a), embora os V_i 's aqui não sejam somente função de β mas também de α , desde que $\text{Var}(Y_i) = \text{Var}(Y_i; \beta, \alpha)$. A equação (3.15) pode também ser escrita substituindo α em (3.8) e (3.15) por $\hat{\alpha}(Y, \beta, \phi)$, um estimador $n^{\frac{1}{2}}$ -consistente de α quando β e ϕ são conhecidos, ou seja, $\hat{\alpha}$ satisfaz $n^{\frac{1}{2}}(\hat{\alpha} - \alpha) \rightarrow O_p(1)$. Para completar o processo, substituímos ϕ por $\hat{\phi}(Y, \beta)$, um estimador $n^{\frac{1}{2}}$ -consistente quando β é conhecido. Conseqüentemente, (3.15) tem a forma

$$\sum_{i=1}^n U_i[\beta, \hat{\alpha}\{\beta, \hat{\phi}(\beta)\}] = 0, \quad (3.16)$$

e $\hat{\beta}_G$ é definido como a solução da equação (3.16).

Sob condições de regularidade, Liang & Zeger (1986) mostraram que quando $n \rightarrow \infty$, $\hat{\beta}_G$ é um estimador consistente de β e que $n^{\frac{1}{2}}(\hat{\beta}_G - \beta)$ é assintoticamente gaussiano multivariado com matriz de covariância V_G dada por

$$V_G = \lim_{n \rightarrow \infty} n \left(\sum_{i=1}^n D_i' V_i^{-1} D_i \right)^{-1} \left[\sum_{i=1}^n D_i' V_i^{-1} \text{cov}(Y_i) V_i^{-1} D_i \right] \left(\sum_{i=1}^n D_i' V_i^{-1} D_i \right)^{-1}.$$

A variância estimada $\text{var}(\hat{\beta}_G)$ ou \hat{V}_G de $\hat{\beta}_G$ pode ser obtida substituindo $\text{cov}(Y_i)$ por sua covariância amostral ($C_i C_i'$) e β, ϕ, α por suas respectivas estimativas na expressão de V_G , ou seja,

$$\hat{V}_G = \lim_{n \rightarrow \infty} n \left(\sum_{i=1}^n \hat{D}_i' \hat{V}_i^{-1} \hat{D}_i \right)^{-1} \left[\sum_{i=1}^n \hat{D}_i' \hat{V}_i^{-1} C_i C_i' \hat{V}_i^{-1} \hat{D}_i \right] \left(\sum_{i=1}^n \hat{D}_i' \hat{V}_i^{-1} \hat{D}_i \right)^{-1}. \quad (3.17)$$

Como no caso de independência, a consistência de $\hat{\beta}_G$ e \hat{V}_G depende somente da especificação correta da média, não a especificação correta da escolha de $R(\alpha)$. Note que a variância assintótica de $\hat{\beta}_G$ não depende da escolha do estimador de α e

ϕ entre aqueles que são $n^{\frac{1}{2}}$ -consistente. Portanto, se a matriz de covariância for especificada corretamente, então $\hat{\beta}_G$ é assintoticamente não viesado e eficiente. Também se a matriz de covariância não for especificada corretamente, $\hat{\beta}_G$ é assintoticamente não viesado e eficiente.

O erro-padrão “ingênuo” (Naive) para $\hat{\beta}_G$, obtido a partir da estimativa “Naive” (equação 2.6) pode ser muito enganoso. Porém erros-padrão consistentes para $\hat{\beta}_G$ são ainda possíveis usando o estimador ‘sandwich’ (algumas vezes chamado de estimador “robusto” ou “empírico”) que são as estimativas obtidas a partir dos estimadores propostos por Liang & Zeger (1986) aplicado para dados longitudinais, equação (3.17).

De acordo com Prentice (1988), podemos estimar os parâmetros de associação considerando um segundo grupo de equações de estimação $S_\alpha(\beta, \alpha) = 0$ e simultaneamente resolvendo as equações expandidas para $\hat{\beta}$ e $\hat{\alpha}$.

Hardin & Hilbe (2002) formalmente escrevem o estimador do parâmetro de associação α como as equações da forma

$$S_\alpha(\beta, \alpha) = \sum_{i=1}^n \left(\frac{\partial \xi_i}{\partial \alpha} \right)' \mathbf{H}^*_{i^{-1}} (\mathbf{W}^*_i - \xi_i) = [0]_{q \times 1},$$

onde

$$\begin{aligned} \mathbf{W}^*_i &= (r_{i1}r_{i2}, r_{i1}r_{i3}, \dots, r_{iT-1}r_{iT})'_{q \times 1} \\ \mathbf{H}^*_i &= \text{diag}(\text{Var}(\mathbf{W}^*_{it}))_{q \times q} \\ \xi_i &= \text{E}(\mathbf{W}^*_i)_{q \times 1} \end{aligned} \quad (3.18)$$

tal que r_{it} é o resíduo de Pearson, \mathbf{H}^*_i é a matriz diagonal e $q = \binom{T}{2}$. A partir desta equação de estimação, é nítido que a parametrização de matriz de correlação ingressa através da equação (3.18). No ajuste desta equação de estimação substitui-se \hat{r}_{it} obtido a partir da estimativa de $\hat{\beta}$.

A obtenção dos estimadores do parâmetro α podem ser obtidos a partir da solução de equação de estimação acima.

3.5 Estimação para GEE2

Uma representação geral loglinear da verossimilhança para um único grupo (*cluster*) com T observações ordinais pode ser escrito como

$$l(\Theta) = \theta_0 + \sum_{t=1}^T \theta'_{1t} \mathbf{y}_t + \sum_{t < j} \theta'_{2tj} \mathbf{y}_t \otimes \mathbf{y}_j + \sum_{t < j < k} \theta'_{3tjk} \mathbf{y}_t \otimes \mathbf{y}_j \otimes \mathbf{y}_k + \dots + \theta'_{t1:T} \mathbf{y}_1 \otimes \dots \otimes \mathbf{y}_T,$$

onde \mathbf{y}_t é o vetor C de variáveis indicadoras acumuladas de ordem $p \times 1$ e \otimes representa o produto Kronecker. Seja $\mathbf{y}'_t = (c_{1t}, c_{2t}, \dots, c_{pt})'$ e $\mathbf{y}'_j = (c_{1j}, c_{2j}, \dots, c_{pj})'$ vetores de ordem $p \times 1$, então

$$\mathbf{y}_t \otimes \mathbf{y}_j = \begin{pmatrix} c_{1t} \\ c_{2t} \\ \vdots \\ c_{pt} \end{pmatrix} (c_{1j}, c_{2j}, \dots, c_{pj})' = \begin{bmatrix} c_{1t}c_{1j} & c_{1t}c_{2j} & \dots & c_{1t}c_{pj} \\ c_{2t}c_{1j} & c_{2t}c_{2j} & \dots & c_{2t}c_{pj} \\ \vdots & \vdots & \vdots & \vdots \\ c_{pt}c_{1j} & c_{pt}c_{2j} & \dots & c_{pt}c_{pj} \end{bmatrix},$$

ou seja, $\mathbf{y}_t \otimes \mathbf{y}_j$ é a notação do produto Kronecker entre \mathbf{y}_t e \mathbf{y}_j . Para o produto Kronecker os vetores ou matrizes não precisam necessariamente ter a mesma ordem (Rao & Rao, 1998).

Devido o interesse nas médias marginais e na correlação, considere uma partição dos parâmetros canônicos $(\theta_1, \theta_2, \theta_3, \dots, \theta_t)$ com (θ_1, θ_2) os parâmetros de primeira e segunda ordem, denominados de baixa ordem e seja $\gamma^{**} = (\theta_3, \dots, \theta_t)$ representando os parâmetros

canônicos de alta ordem, transforme estes parâmetros canônicos de baixa ordem para os parâmetros de regressão da média marginal β e os parâmetros de correlação α , embora qualquer nível de partição seja possível. Marginalmente os dois primeiros momentos fornecem reprodutibilidade e interpretabilidade dos parâmetros marginais em comparação com modelos que possuem estimativas para os parâmetros que utilizam momentos de ordens maiores. O parâmetro γ^{**} é conhecido a partir da propriedade geral da família exponencial de que parâmetros de terceira e mais ordens, γ^{**} é ortogonal para os parâmetros marginais (β, α) . Temos uma equação de estimação que é para estimar β e outra equação para estimar α .

No GEE2 não existe a suposição de que as duas equações são ortogonais. Inicialmente é na equação de estimação de β tratando α como ancilar.

Considere Ψ a equação de estimação generalizada, similar à equação (3.6), isto é,

$$\Psi = \left[\sum_{i=1}^n \Psi_i(\mathbf{x}_i, \beta) \right]_{p \times 1} = [0]_{p \times 1},$$

onde $\Psi_i(\mathbf{x}_i, \beta)$ é a equação de estimação para o i -ésimo indivíduo. Portanto,

$$\begin{aligned} \Psi(\beta) &= \sum_{i=1}^n \frac{\partial \mu_i}{\partial \beta} V_i^{-1} \left(\frac{\mathbf{y}_i - \mu_i}{\phi} \right) = \sum_{i=1}^n \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta} V_i^{-1} \left(\frac{\mathbf{y}_i - \mu_i}{\phi} \right) \\ &= \sum_{i=1}^n \frac{\partial \mu_i}{\partial \eta_i} \mathbf{x}_i V_i^{-1} \left(\frac{\mathbf{y}_i - \mu_i}{\phi} \right) = [0]_{p \times 1}, \end{aligned}$$

com $j = 1, \dots, p$ e V_i a função de variância.

Como definido em Hardin & Hilbe (2002), para a estimação do modelo marginal considere as equações de estimação combinando os parâmetros de regressão com os parâmetros de ancilaridade dadas por:

$$\begin{aligned} \Psi(\beta, \alpha) &= (\Psi_\beta(\beta, \alpha), \Psi_\alpha(\beta, \alpha)) \\ &= \begin{pmatrix} \sum_{i=1}^n \mathbf{x}_i \text{diag} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) V_i^{-1} \left(\frac{\mathbf{y}_i - \mu_i}{\phi} \right) \\ \sum_{i=1}^n \left(\frac{\partial \xi_i}{\partial \alpha} \right)' \mathbf{H}_i^{*-1} (\mathbf{W}_i^* - \xi_i) \end{pmatrix} \end{aligned}$$

Na estimação supõe-se que a equação de estimação para a correlação é ortogonal a equação de estimação para β . Em cada passo no algoritmo MLG, estimamos primeiro $R(\alpha)$, isto é α , e depois usamos isto para estimar β . A convergência é declarada quando as mudanças nas estimativas dos parâmetros é menor que um certo critério, ou a mudança na soma dos quadrados dos desvios é menor que um dado critério. O foco do procedimento para resolver as equações de estimação é a iteração do algoritmo dos mínimos quadrados ponderados, este algoritmo é uma modificação do algoritmo de Newton-Raphson no qual a

matriz Hessiana esperada é substituída pela Hessiana observada. A modificação é conhecida como o método de escore de Fisher.

Assim para um γ^{**} fixo a contribuição do i -ésimo cluster para as equações escores é

$$\begin{bmatrix} \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \alpha} \end{bmatrix} = \mathbf{U}_i(\beta, \alpha) = \begin{bmatrix} \frac{\partial \mu_i}{\partial \beta} & 0 \\ \frac{\partial \nu_i}{\partial \beta} & \frac{\partial \nu_i}{\partial \alpha} \end{bmatrix}' \begin{bmatrix} \mathbf{V}_{i11} & \mathbf{V}_{i12} \\ \mathbf{V}_{i21} & \mathbf{V}_{i22} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Y}_i - \mu_i(\beta) \\ \mathbf{W}_i - \nu_i(\beta, \alpha) \end{bmatrix},$$

onde

$$\mathbf{Y}_i = (\mathbf{Y}'_{i1}, \mathbf{Y}'_{i2}, \dots, \mathbf{Y}'_{it})',$$

$$\mu_i = \mathbf{E}(\mathbf{Y}_i),$$

$$\mathbf{W}_i = [(\mathbf{Y}_{i1} \otimes \mathbf{Y}_{i2})', (\mathbf{Y}_{i1} \otimes \mathbf{Y}_{i3})', \dots, (\mathbf{Y}_{i(t-1)} \otimes \mathbf{Y}_{it})']',$$

$$\nu_i = \mathbf{E}(\mathbf{W}_i),$$

$$\mathbf{V}_{i11} = \text{var}(\mathbf{Y}_i),$$

$$\mathbf{V}_{i12} = \text{cov}(\mathbf{Y}_i, \mathbf{W}_i),$$

e

$$\mathbf{V}_{i22} = \text{var}(\mathbf{W}_i).$$

Note que a solução de (β, α) das equações escores depende de γ^{**} somente através das matrizes de peso \mathbf{V}_{22} e \mathbf{V}_{12} . Portanto, pela teoria da função de estimação padrão, o estimador $(\hat{\beta}, \hat{\alpha})$ possuirá certa robustez para a mal especificação de γ^{**} . Como γ^{**} representa os parâmetros que são geralmente de pouco interesse, frequentemente utilizamos γ^{**} fixo.

Usando $\gamma^{**} = 0$ como um modelo de trabalho para dados ordinais temos as chamadas equações GEE2 que permitem a modelagem de cluster de tamanho desigual. Modelagem completa de γ^{**} , embora mais flexível, geralmente restringe a atenção para situações onde os dados são balanceados sobre os clusters, porque o parâmetro canônico γ^{**} depende do tamanho do cluster.

Para n clusters independentes de respostas ordinais, a verossimilhança conjunta tem a forma

$$\begin{aligned} L(\beta, \alpha | \gamma^{**}) &= \prod_{i=1}^n \Pr(\mathbf{Y}_{i1} = \mathbf{y}_{i1}, \dots, \mathbf{Y}_{it} = \mathbf{y}_{it}) \\ &= \prod_{i=1}^n \exp \left(\sum_{t=1}^T \theta'_{it}(\beta, \alpha, \gamma^{**}) \mathbf{y}_{it} + \sum_{i < t} \theta'_{itn}(\beta, \alpha, \gamma^{**}) \mathbf{y}_{it} \otimes \mathbf{y}_{in} + A_i(\Theta) + B_i(\mathbf{y}_i) \right) \end{aligned}$$

Estimativas de máxima verossimilhança dos valores conhecidos para γ^{**} fixo são a solução das equações escores, da forma (3.15), e pode ser obtido como segue.

3.5.1 Algoritmo de estimação

A estimação GEE2 alterna entre escore de Fisher para atualização de (β, α) e ajuste proporcional iterativo (IPF - interactive proportional fitting) para obter a probabilidade conjunta completa para cada cluster a partir do qual as matrizes necessárias de pesos são formadas.

O algoritmo de estimação pode ser sumarizado como segue:

- Obter valores iniciais $(\beta^{(0)}, \alpha^{(0)})$. Usar $\beta^{(0)}$ como o ajuste de uma regressão assumindo independência e $\alpha^{(0)} = 0$ ou, no esforço para acelerar a convergência, pode usar as estimativas obtidas a partir de GEE1 como descrito mais adiante.

- Usar IPF para obter a probabilidade conjunta para cada cluster baseada em uma estimativa corrente $(\beta^{(t)}, \alpha^{(t)})$ e o valor fixo de γ^{**} . Usar as estimativas $(\beta^{(t)}, \alpha^{(t)})$ para obter $\mathbf{V}_{i11}^{(t)}$ e o vetor de probabilidade conjunta para obter $\mathbf{V}_{i12}^{(t)}$ e $\mathbf{V}_{i22}^{(t)}$.

- Tomar o escore de Fisher a um passo para (β, α) como segue:

$$\begin{pmatrix} \beta^{(t+1)} \\ \alpha^{(t+1)} \end{pmatrix} = \begin{pmatrix} \beta^{(t)} \\ \alpha^{(t)} \end{pmatrix} + \left(\sum_{i=1}^n \tilde{D}'_i \tilde{V}_i^{-1} \tilde{D}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{U}_i(\beta^{(t)}, \alpha^{(t)}) \right),$$

onde \tilde{D}_i representa a matriz de derivadas para o i -ésimo cluster $\partial(\mu_i, \nu_i)/\partial(\beta, \alpha)$ e $\tilde{V}_i = \text{cov}(Y_i, W_i)$.

- Reiterar até que a convergência seja alcançada, como monitorado através da mudança relativa das estimativas ou em termos da verossimilhança maximizada.

3.5.2 Propriedades assintóticas de $(\hat{\beta}_{gee2}, \hat{\alpha}_{gee2})$

É sabido que a solução das equações de estimação GEE2 permanece consistente somente sob as suposições de existência do primeiro e segundo momentos. Se realmente $\gamma_i^{**} = \gamma^{**}$ (tipicamente $\gamma^{**} = 0$) então $(\hat{\beta}, \hat{\alpha})$ são estimadores de máxima verossimilhança (MV) e daí são assintoticamente gaussianos com variância assintótica de $(\hat{\beta}, \hat{\alpha})$ dada pela inversa da matriz de informação:

$$\mathcal{I}(\beta, \alpha) = \sum_{i=1}^n \tilde{D}'_i \tilde{V}_i^{-1} \tilde{D}_i.$$

Contudo, sob a mal especificação do terceiro ou maior momentos, a informação inversa pode não ter variância assintótica como solução das equações escores. Seja $\tilde{V}_i^{(V)}$ a verdadeira covariância de Y_i e W_i e seja $\tilde{V}_i^{(M)}$ a covariância do modelo. A solução para as equações do GEE2 são assintoticamente gaussianas com matriz de covariância

$$V^* = \left[\sum_{i=1}^n \tilde{D}'_i (\tilde{V}_i^{(M)})^{-1} \tilde{D}_i \right]^{-1} \left[\sum_{i=1}^n \tilde{D}'_i (\tilde{V}_i^{(M)})^{-1} \tilde{V}_i^{(V)} (\tilde{V}_i^{(M)})^{-1} \tilde{D}_i \right] \left[\sum_{i=1}^n \tilde{D}'_i (\tilde{V}_i^{(M)})^{-1} \tilde{D}_i \right]^{-1},$$

onde o estimador robusto do termo médio é dado por $\sum_{i=1}^n U_i U'_i$.

Sob as suposições do modelo da família exponencial com parâmetros $\delta = (\beta, \alpha)$, temos que, quando $n \rightarrow \infty$

$$(\hat{\mathcal{I}})^{-1/2}(\hat{\delta} - \delta) \rightarrow N(0, \mathbf{I}_{p+q}).$$

Porém apenas sob a suposição de independência do primeiro e segundo momentos sobre todos os clusters, temos

$$(\hat{V}^*)^{-1/2}(\hat{\delta} - \delta) \rightarrow N(0, \mathbf{I}_{p+q}).$$

Note que embora as equações de estimação GEE2 sejam equações escores para β e α sob o modelo da família exponencial quadrática (para γ^{**} fixo), eles conservam propriedades desejáveis quando este modelo não é seguro. Primeiro as equações de estimação são não viesadas, e a solução $(\hat{\beta}, \hat{\alpha})$ permanece consistente somente sob a suposição da correta especificação do modelo. Segundo, as equações de estimação permanecem equações escores para um modelo de parâmetros mistos onde os parâmetros canônicos de primeira e segunda ordem são transformados para os parâmetros marginais e as restantes associações de alta ordem são modeladas via parâmetros canônicos.

A desvantagem do método GEE2, contudo, é que as estimativas dos parâmetros no modelo marginal não são eficientes se não é especificado o modelo de associação (Agresti & Natarajan, 2001).

3.6 Estimação para GEE1

Temos que as equações escores desenvolvidas anteriormente podem equivalentemente ser representadas como segue:

$$\begin{bmatrix} \mathbf{U}_1(\beta, \alpha) \\ \mathbf{U}_2(\beta, \alpha) \end{bmatrix} = \sum_{i=1}^n \begin{bmatrix} \frac{\partial \mu_i}{\partial \beta} & 0 \\ \frac{\partial \sigma_i}{\partial \beta} & \frac{\partial \sigma_i}{\partial \alpha} \end{bmatrix}' \begin{bmatrix} \mathbf{V}_{i11} & \mathbf{V}_{i12}^c \\ \mathbf{V}_{i21}^c & \mathbf{V}_{i22}^c \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Y}_i - \mu_i(\beta) \\ \mathbf{S}_i - \sigma_i(\beta, \alpha) \end{bmatrix}.$$

O sobre-escrito c denota as matrizes de covariância correspondentes para o produto centrado $\mathbf{S}_{i(t,n)} = (\mathbf{Y}_{it} - \mu_{it}) \otimes (\mathbf{Y}_{in} - \mu_{in})$ e $\sigma_i = E(\mathbf{S}_i)$, que possui as propriedades de invariância e eficiência.

O enfoque GEE1 considera os parâmetros de associação como uma perturbação (“nuisance”) (Liang & Zeger 1986; Zeger & Liang 1986) e seu foco principal é fazer inferência para o parâmetro da regressão marginal da média β . As equações de estimação GEE2 são modificadas usando $\mathbf{V}_{i12}^c = 0$, $\partial \sigma_i / \partial \beta = 0$ e usando uma aproximação para \mathbf{V}_{i22}^c . O resultado produz equações de estimação separadas para os parâmetros (β) , da média marginal, e os parâmetros do segundo momento marginal (α) :

$$\mathbf{U}_1^*(\beta, \alpha) = \sum_{i=1}^n \begin{bmatrix} \frac{\partial \mu_i}{\partial \beta} \end{bmatrix}' \mathbf{V}_{i11}^{-1} (\mathbf{Y}_i - \mu_i(\beta))$$

e

$$\mathbf{U}_2^*(\beta, \alpha) = \sum_{i=1}^n \left[\frac{\partial \sigma_i}{\partial \alpha} \right]' (\hat{\mathbf{V}}_{i22}^c)^{-1} (\mathbf{S}_i - \sigma_i(\beta, \alpha))$$

Note que usando produtos centrados ou não centrados são equivalentes somente quando se usa a matriz de covariância completa. Se estabelecemos $\mathbf{V}_{i12}^c = 0$ então isto leva a um estimador diferente do que se estabelecemos $\mathbf{V}_{i12} = 0$.

A GEE2 pode ser escrita como

$$\sum_{i=1}^n \begin{pmatrix} \frac{\partial \mu_i}{\partial \beta} & \frac{\partial \mu_i}{\partial \alpha} \\ \frac{\partial \sigma_i}{\partial \beta} & \frac{\partial \sigma_i}{\partial \alpha} \end{pmatrix}' \begin{pmatrix} V(\mathbf{y}_i, \mathbf{y}_i) & V(\mathbf{y}_i, \mathbf{s}_i) \\ V(\mathbf{s}_i, \mathbf{y}_i) & V(\mathbf{s}_i, \mathbf{s}_i) \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{y}_i - \mu_i \\ \mathbf{s}_i - \sigma_i \end{pmatrix} = [0].$$

O que difere GEE1 do GEE2 é assumir que os dois primeiros termos na equação de estimação são bloco diagonal, assume matrizes de zero na posição fora da diagonal. Portanto, GEE2 é uma generalização do GEE1. Em muitas aplicações a média é definida somente pelos parâmetros de regressão, mas a suposição de que $\partial \mu_i / \partial \alpha \neq 0$ implica que a correlação é uma função dos parâmetros de regressão β .

3.6.1 Algoritmo de estimação

Estimação de (β, α) para dados ordinais tem a diferença de que a covariância do vetor de respostas Y_i tem uma estrutura de bloco-diagonal de cada vetor de indicadores Y_{ij} determinado pela média μ_{ij} . Qualquer estrutura de associação de “trabalho” deverá preservar a estrutura de bloco-diagonal da matriz de peso. Em adição, para a equação de estimação α , usamos os produtos centrados e usamos uma matriz de peso de bloco-diagonal. A seguir o algoritmo de estimação para GEE1:

- Obter estimativas iniciais $(\beta^{(0)}, \alpha^{(0)})$. Usar $\beta^{(0)}$ com o ajuste obtido de uma regressão de chances proporcionais assumindo independência e $\alpha^{(0)} = 0$.
- Usando a iteração Gauss-Seidel toma um passo Newton próximo a solução para as equações de estimação \mathbf{U}_1^* e \mathbf{U}_2^* como segue:

$$\beta^{(t+1)} = \beta^{(t)} + \left(\sum_{i=1}^n D'_{i11} V_{i11}^{-1} D_{i11} \right)^{-1} \sum_{i=1}^n \mathbf{U}_1^*(\beta^{(t)}, \alpha^{(t)}),$$

$$\alpha^{(t+1)} = \alpha^{(t)} + \left(\sum_{i=1}^n D'_{i22} \hat{V}_{i22}^{-1} D_{i22} \right)^{-1} \sum_{i=1}^n \mathbf{U}_2^*(\beta^{(t)}, \alpha^{(t)}),$$

com $D_{i11} = \partial \mu_i / \partial \beta$ e $D_{i22} = \partial \sigma_i / \partial \alpha$.

- Iteração até que a mudança relativa nas estimativas dos parâmetros sejam pequenas.

O algoritmo de estimação claramente realça a maior distinção entre a estimação GEE2 e GEE1. Temos que as equações de estimação para β e α são resolvidas separadamente. Isto produz estimação consistente para β sob a mal especificação do modelo de associação (Zeger & Liang, 1986).

3.6.2 Propriedades assintóticas de $(\hat{\beta}_{gee1}, \hat{\alpha}_{gee1})$

A prova da normalidade assintótica do estimador $(\hat{\beta}_{gee1}, \hat{\alpha}_{gee1})$ é dado por Liang & Zeger (1986). A variância assintótica de $(\hat{\beta}_{gee1}, \hat{\alpha}_{gee1})$ é dado por

$$AV_{gee1} = \left(\sum_{i=1}^n C'_{1i} B_{1i} C_{2i} \right)^{-1} \left(\sum_{i=1}^n C'_{1i} B_{1i} V^*_i B'_{1i} C_{1i} \right) \left(\sum_{i=1}^n C'_{2i} B_{1i} C_{1i} \right)^{-1},$$

onde

$$C_{1i} = \begin{bmatrix} \partial\mu_i/\partial\beta & 0 \\ 0 & \partial\sigma_i/\partial\alpha \end{bmatrix}, \quad C_{2i} = \begin{bmatrix} \partial\mu_i/\partial\beta & 0 \\ \partial\sigma_i/\partial\beta & \partial\sigma_i/\partial\alpha \end{bmatrix}$$

$$B_{1i} = \begin{bmatrix} V_{i11} & 0 \\ 0 & \hat{V}_{i22} \end{bmatrix}^{-1},$$

e $V^*_i = \text{cov}(Y_i, \mathbf{S}_i)$. Isto pode ser estimado usando o estimador robusto substituindo o termo intermediário com

$$\sum_{i=1}^n \begin{pmatrix} U^*_{1i} \\ U^*_{2i} \end{pmatrix} \begin{pmatrix} U^*_{1i} \\ U^*_{2i} \end{pmatrix}'.$$

Na estatística existe uma extensiva literatura para a seleção de modelos para a regressão linear clássica para dados independentes. Um poderoso e amplamente utilizado critério de seleção do modelo é o Critério de Informação Akaike (AIC) (Akaike, 1973) que é baseado na verossimilhança e nas propriedades assintóticas do estimador de máxima verossimilhança (EMV). Assim como não é assumido nenhum tipo de distribuição quando se considera a abordagem de quase-verossimilhança, então a verossimilhança não é definida; e nesse caso o AIC não pode ser usado diretamente. Pan (2001) propõe uma extensão do AIC para GEE, que envolve a quase-verossimilhança e que está baseado na informação Kullback-Leibler (Kullback & Leibler, 1951).

Capítulo 4

Modelos marginais

4.1 Considerações iniciais

Modelo marginal é um método que caracteriza a esperança marginal de uma variável resposta contínua ou discreta como uma função de variáveis explicativas, este tipo de modelo é conveniente quando o interesse principal são inferências sobre a média populacional. Esses modelos são úteis em estudos com aplicações nas ciências sociais, biológicas e da saúde, em especial, nos estudos epidemiológicos. Pela esperança marginal, queremos obter a resposta “média” sobre a sub-população que compartilha um valor comum de \mathbf{x} (Diggle *et al.*, 2001).

Um modelo “média-populacional”, também conhecido como um modelo marginal, é obtido através da introdução da parametrização da covariância a nível painel (ou correlação), em que essa correlação é estimada pela média sobre a informação de todos os painéis. Na modelagem dos efeitos das covariáveis nas esperanças marginais pode-se especificar também um modelo de associação entre observações de cada sujeito.

Modelos marginais para a estrutura da média, tem recebido crescente popularidade desde que eles têm sido implementado em diversos “softwares” padrão (Ziegler *et al.*, 2000). Com o intuito de obter estimativas consistentes para a média e a estrutura de associação serão utilizados os enfoques GEE1 e GEE2, apresentados no capítulo anterior.

A análise de um modelo marginal depara-se com dificuldades de três tipos: primeiro, as observações no tempo não são independentes; segundo, indivíduos apresentam variações, assim as distribuições marginais têm grande dispersão; terceiro, hipóteses marginais não determinam as frequências esperadas (Wickens, 1998).

O modelo marginal tem as seguintes suposições:

- 1 - A esperança marginal da variável resposta, $E(Y_{it}) = \mu_{it}$, depende das variáveis explicativas, \mathbf{x}_{it} , por $g(\mu_{it}) = \mathbf{x}'_{it}\beta$ onde g é uma função de ligação conhecida;
- 2 - A variância marginal depende da média marginal da seguinte forma $\text{Var}(Y_{it}) = V(\mu_{it})\phi$, onde V é uma função conhecida da variância e ϕ é o parâmetro de escala;
- 3 - A correlação entre Y_{ij} e Y_{it} é uma função das médias marginais e em alguns casos é necessário estimar o parâmetro adicional α , visto que a $\text{Corr}(Y_{ij}, Y_{it}) = \rho(\mu_{ij}, \mu_{it}; \alpha)$ onde $\rho(\cdot)$ é uma função conhecida.

No modelo de regressão logística, especificamos que as esperanças marginais $E(Y_{it}) = \mu_{it}$, satisfaz em $\text{logit}(\mu_{it}) = \mathbf{x}'_{it}\beta$. Assim a função estimada GEE é a equação score para β

quando os dados seguem uma distribuição de probabilidade loglinear e se especificamos corretamente $\text{Var}(Y_{it})$.

O modelo de regressão logística generalizada assume que as respostas de cada covariável tem uma distribuição multinomial, onde as respostas multinomiais para diferentes covariáveis são independentes, devido a essa suposição este modelo é comumente chamado de modelo logit multinomial.

Os coeficientes de regressão do modelo marginal, β , têm interpretação similar aos parâmetros de um estudo transversal. Modelos do tipo marginal são análogos para dados correlacionados aos MLGs para dados independentes.

Para exemplificar um modelo marginal logístico, considere o problema de avaliar, ao longo do tempo, a dependência da infecção respiratória em crianças, em função da situação da vitamina A. Seja x_{it} indicando se a i -ésima criança é ou não deficiente de vitamina A (1 - sim; 0 - não) na t -ésima visita. Seja Y_{it} denotando se a criança tem infecção respiratória (1-sim; 0- não) e seja $\mu_{it} = E(Y_{it})$. Um modelo marginal é dado pelas seguintes suposições:

- $\text{logit}(\mu_{it}) = \log \frac{\mu_{it}}{1-\mu_{it}} = \log \frac{\Pr(Y_{it}=1)}{\Pr(Y_{it}=0)} = \beta_0 + \beta_1 x_{it}$,
- $\text{Var}(Y_{it}) = \mu_{it}(1 - \mu_{it})$,
- $\text{Corr}(Y_{it}, Y_{ij}) = \alpha$.

O coeficiente de regressão transformado $\exp(\beta_0)$ é a razão entre a probabilidade de uma criança infectada e a probabilidade de uma criança não infectada na sub-população que não é deficiente de vitamina A. O parâmetro $\exp(\beta_1)$ é a razão de chance de infecção entre as crianças deficientes de vitamina A e as crianças providas com vitamina A.

Com dados binários ou outra forma de dados categorizados, dois tipos de modelos para respostas ordinal diferem em termos de se eles têm efeito de média populacional ou se têm efeito de grupo-específico ou sujeito-específico.

O modelo marginal se refere às distribuições marginais das médias sobre os grupos ou *clusters* na população. Os grupos são de alguma forma similares, segundo se refere a distribuições condicionais ao nível de *cluster* (por exemplo, o sujeito). Interpretações a nível populacional são mais relevantes em estudos epidemiológicos.

A seguir serão discutidos modelos marginais para dados longitudinais categorizados com algum tipo de ordenação. No capítulo 2 foram abordados alguns modelos para variáveis respostas binárias e, na seção 2.4, respostas multinomiais, e na sub-seção 2.4.3, modelos para respostas ordinais. Portanto, a ênfase agora para este tipo de modelo é para quando os dados são longitudinais.

Seja $\pi_k(\mathbf{x}; i)$ a probabilidade de que a resposta seja a categoria k ($k = 1, \dots, c$), para uma observação no tempo no i -ésimo *cluster* ($i = 1, \dots, n$) com valores \mathbf{x} do vetor ($p \times 1$) de variáveis explicativas. Seja $\pi_k(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \pi_k(\mathbf{x}; i)$, a probabilidade correspondente a “média populacional”, isto é a média das probabilidades $\pi_k(\mathbf{x}; i)$ para todos os “clusters” na

população de interesse.

Seja $\gamma_k(\mathbf{x}; i) = \pi_1(\mathbf{x}; i) + \cdots + \pi_k(\mathbf{x}; i)$ e $\gamma_k(\mathbf{x}) = \pi_1(\mathbf{x}) + \cdots + \pi_k(\mathbf{x})$, com $\gamma_0(\mathbf{x}; i) = \gamma_0(\mathbf{x}) = 0$.

Para um grupo de valores preditos $\mathbf{x}^* = (\mathbf{x}_1, \dots, \mathbf{x}_t)$ se forem realizadas t observações no tempo teremos $\pi_{k_1, \dots, k_t}(\mathbf{x}^*)$ como sendo a probabilidade marginal da seqüência de respostas (k_1, \dots, k_t) . Se as observações no tempo dentro dos grupos são independentes, a probabilidade da seqüência de resposta para o i -ésimo “cluster” será o produto $\pi_{k_1}(\mathbf{x}_1; i) \cdots \pi_{k_t}(\mathbf{x}_t; i)$, então

$$\pi_{k_1, \dots, k_t}(\mathbf{x}^*) = \frac{1}{n} \sum_{i=1}^n \pi_{k_1}(\mathbf{x}_1; i) \cdots \pi_{k_t}(\mathbf{x}_t; i).$$

Assim, $\pi_{k_1, \dots, k_t}(\mathbf{x}^*)$ é a probabilidade marginal em termos da média sobre os grupos, e $\pi_k(\mathbf{x})$ é também a distribuição marginal de primeira ordem de uma distribuição t -dimensional.

Podemos considerar modelos logits acumulados do tipo (2.11) para as probabilidades marginais de primeira ordem. Para cada tempo, seja t_i o número de observações no tempo do *cluster* i , e seja $\mathbf{x}_i^* = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{it_i})$, $i = 1, \dots, n$ e β o vetor de parâmetros $(p \times 1)$. Os modelos têm a forma

$$\text{logit}[\gamma_k(\mathbf{x}_{it})] = \alpha_k^* + \mathbf{x}_{it}'\beta, \quad k = 1, \dots, c-1 \text{ e } t = 1, \dots, t_i. \quad (4.1)$$

ou, reescrevendo

$$\begin{aligned} \text{logit}[\gamma_k(\mathbf{x}_{it})] &= \log \left[\frac{\gamma_k(\mathbf{x}_{it})}{1 - \gamma_k(\mathbf{x}_{it})} \right] \\ &= \log \left[\frac{\pi_1(\mathbf{x}_{it}) + \cdots + \pi_k(\mathbf{x}_{it})}{1 - \pi_1(\mathbf{x}_{it}) + \cdots + \pi_k(\mathbf{x}_{it})} \right], \end{aligned}$$

e portanto,

$$\left[\frac{\pi_1(\mathbf{x}_{it}) + \cdots + \pi_k(\mathbf{x}_{it})}{1 - \pi_1(\mathbf{x}_{it}) + \cdots + \pi_k(\mathbf{x}_{it})} \right] = \exp(\alpha_k^* + \mathbf{x}_{it}'\beta),$$

ou ainda,

$$\pi_1(\mathbf{x}_{it}) + \cdots + \pi_k(\mathbf{x}_{it}) = \frac{\exp(\alpha_k^* + \mathbf{x}_{it}'\beta)}{1 + \exp(\alpha_k^* + \mathbf{x}_{it}'\beta)}.$$

O vetor β ($p \times 1$) têm efeito fixo, ou seja, o modelo assume efeito idêntico β nos preditores, para todos os clusters, em relação às probabilidades acumuladas.

Seja $Y_{k_1, \dots, k_{t_i}}(i) = 1$, se o *cluster* i assume o grupo de respostas k_1, \dots, k_{t_i} e seja $Y_{k_1, \dots, k_{t_i}}(i) = 0$, caso contrário. Marginalmente assumimos que estes indicadores são ensaios de uma distribuição multinomial com parâmetros $\{\pi_{k_1, \dots, k_{t_i}}(\mathbf{x}_i^*)\}$. Então, o logaritmo da

verossimilhança é proporcional a

$$\sum_{i=1}^n \left(\sum_{k_1=1}^c \sum_{k_2=1}^c \cdots \sum_{k_{t_i}=1}^c y_{k_1, \dots, k_{t_i}}(i) \log[\pi_{k_1, \dots, k_{t_i}}(\mathbf{x}_i^*)] \right),$$

onde $y_{k_1, \dots, k_{t_i}}(i)$ denota a realização de $Y_{k_1, \dots, k_{t_i}}(i)$. É difícil maximizar este logaritmo da verossimilhança porque o modelo (4.1) refere às probabilidades marginais $\{\pi_k(\mathbf{x}_{it})\}$ e $\{\pi_{k_1, \dots, k_{t_i}}(\mathbf{x}_i^*)\}$ são probabilidades multinomiais no logaritmo da verossimilhança. Daí não se poder substituir diretamente na função de verossimilhança e maximizar usando métodos padrões. Esta dificuldade reflete a necessidade de uma família multivariada de distribuições para respostas categorizadas. Em vez de tentar especificar completamente a distribuição conjunta, se pode aplicar a metodologia baseada em uma generalização multivariada da quase-verossimilhança e considerar as equações de estimação generalizadas e suas estratégias de estimação.

Como já mencionado no capítulo anterior, o enfoque GEE2 difere do GEE1 nas equações de estimação para os parâmetros de associação. No caso GEE2 os parâmetros são assumidos não serem ortogonais para a equação estimada para os coeficientes de regressão. Quando a estrutura de associação é de interesse, um enfoque GEE2 é disponível para modelar a associação usando a razão de chances global (Heagerty & Zeger, 1996). Uma desvantagem do método GEE2, contudo, é que as estimativas dos parâmetros no modelo marginal não são muito consistentes se o modelo de associação é mal especificado.

Para o modelo marginal usando o R, foi utilizado o objeto “ORDGEE” do pacote “GEESE”, o qual produz um ajuste para as equações de estimação generalizadas para dados agrupados (“clustered”) ordinais. Este objeto foi implementado por Yan (2003). As estimativas oferecidas pelo R, são baseadas no enfoque dado por Heagerty & Zeger (1996), estes autores consideram que dado um vetor de medidas ordinais, \mathbf{O}_i , para o i -ésimo “cluster” (sujeito) e O_{it} , a t -ésima observação no tempo do i -ésimo sujeito, e \mathbf{x}_{it} as covariáveis associadas com O_{it} . Podemos representar a medida ordinal $O_{it} = k$, onde $k \in [0, 1, 2, \dots, c]$ através das variáveis indicadoras acumulativas

$$Y_{itk} = I(O_{it} > k),$$

onde $k \in [0, 1, 2, \dots, c - 1]$. E, portanto, para o modelo de *odds proportional* para as médias marginais é assumido que

$$\text{logit}[E(Y_{itk})] = \alpha_k^* + \mathbf{x}_{it}' \boldsymbol{\gamma}^*,$$

ou ainda

$$\text{logit}[\Upsilon_{itk}] = \alpha_k^* + \mathbf{x}_{it}' \boldsymbol{\gamma}^*,$$

com $t = (1, \dots, t_i)$ e $k = 1, \dots, c - 1$ e $t = 1, \dots, t_i$.

Na utilização do pacote R é possível considerar dois tipos de modelos, que são modelos com intercepto constante e modelos com intercepto não constante, em relação ao tempo, além da possibilidade de mudanças das estruturas das matrizes de correlação. Para o modelo com o intercepto constante, considere a forma

$$\text{logit}[\Upsilon_{itk}] = \alpha_k^* + \mathbf{x}'_{it}\gamma^*.$$

E para o intercepto não constante, considere a forma

$$\text{logit}[\Upsilon_{itk}] = \alpha_{kt}^* + \mathbf{x}'_{it}\gamma^*,$$

com $i = (1 \dots n)$, $k = 1, \dots, c - 1$ e $t = 1, \dots, t_i$.

Utilizar modelo com intercepto não constante pode ser mais útil em situações onde se deseja obter estimativas mais sensíveis a mudanças no tempo, embora as estimativas considerando os dois tipos de interceptos são similares.

Em muitas aplicações, cada grupo (*cluster*) é considerado como o conjunto de medidas repetidas de cada sujeito. Em outras aplicações, cada *cluster* é o conjunto de unidades amostrais que são consideradas mais similares de acordo com uma(s) determinada(s) característica(s) segundo algum critério.

Capítulo 5

Aplicação

5.1 Organização dos dados e análise preliminar

A maioria dos estudos sobre a situação de saúde e nutrição de populações utiliza a abordagem transversal, porém estas pesquisas são estáticas em relação ao tempo, ao espaço e aos grupos populacionais atingidos, isto implica numa análise limitada na descrição dos processos dinâmicos como, por exemplo, os eventos relacionados ao crescimento. No caso do estudo de “coorte”, grande parte dessa limitação é superada pois possibilita o acompanhamento nas mudanças ligadas às situações de interesse. Embora este tipo de estudo tenha um custo elevado, o estudo de “coorte” possibilita obter um banco de dados para análises mais complexas com grande riqueza descritiva e analítica. É mediante essa perspectiva que a Universidade Federal de Pernambuco (Departamento de Nutrição, Materno-Infantil e de Fisiologia e Farmacologia) em colaboração com a Universidade de Londres (LSHTM) e Universidade de Montpellier-França, se dispuseram a desenvolver um projeto de pesquisa na Zona da Mata Meridional de Pernambuco.

O estudo longitudinal foi constituído inicialmente de uma “coorte” de 652 crianças de famílias residentes em áreas urbanas da zona da mata meridional do estado de Pernambuco, das quais apenas 477 delas tinham informações completas desde o nascimento até os 2, 4, 6, 9, 12, 15 e 18 meses de vida.

A coleta dos dados sobre as crianças foi realizada em 6 maternidades, localizadas nos municípios da zona da mata meridional de Pernambuco, no período de setembro de 1997 a agosto de 1998. A avaliação nutricional foi realizada nas primeiras 24 horas do nascimento. O acompanhamento da morbidade e do aleitamento materno foi realizado através de visitas domiciliares. As medidas antropométricas foram realizadas aos 2, 4, 6, 9, 12, 15 e 18 meses de vida. As informações sobre alimentação e morbidade eram prestadas pelas mães e incluíam informações diárias sobre consumo de leite materno, água, chá, suco, leite artificial e outros alimentos e a ocorrência de diarreia.

As variáveis que foram acompanhadas ao longo dos 18 meses como aleitamento materno, diarreia e outro tipo de alimentação foram organizadas de modo a fornecer apenas informações relevantes aos meses citados acima, juntamente com as medidas antropométricas. Para as variáveis relacionadas à alimentação da criança foram contados os dias, por exemplo, quantidades de dias que uma criança mamou quando tinha 2 meses, 4 meses e, assim por

diante. A ocorrência de diarreia foi calculada pelo total de dias com a doença, aos 2 meses, 4 meses e, até os 18 meses.

Para a avaliação da condição nutricional sabe-se que o peso é considerado o melhor indicador de crescimento da criança por representar todas as mudanças que ocorrem no corpo e é muito sensível pois modifica-se muito rapidamente em situações de desnutrição. Consideramos neste trabalho as medidas antropométricas avaliadas através do escore Z: WAZ (peso por idade) que reflete o estado nutricional atual, porém não permite que se diferencie o déficit atual crônico ou passado, a nível individual ou populacional e a HAZ (altura por idade) que é utilizado no diagnóstico de déficit linear e reflete a desnutrição passada ou crônica. As análises preliminares revelaram que com a variável WAZ os modelos apresentaram maior poder explicativo do que com a variável HAZ e, por esta razão a modelagem aqui apresentada se refere ao comportamento nutricional da criança avaliada pelo indicador WAZ.

Atualmente a OMS (Organização Mundial de Saúde) e o UNICEF (Fundo das Nações Unidas para a Criança - United Nations International Children's Emergency Fund) recomendam a população norte-americana como padrão de referência mundial. Assim, o padrão do National Center for Health Statistics (NCHS), elaborado a partir de dados obtidos de crianças brancas, americanas, de classe média, através de estudos longitudinais realizados nos Estados Unidos de 1929 a 1975, tem sido adotado também pelo Ministério da Saúde como referência da população brasileira. Embora exista um novo padrão de referência para os gráficos de crescimento CDC 2000 (Centers for Disease Control and Prevention), desenvolvido pelo NCHS, que é uma versão revisada do gráficos de crescimento de 1977 da população dos Estados Unidos desde 1960, contudo, a OMS ainda não recomenda o uso como padrão internacional, (<http://www.cdc.gov/nchs/about/major/nhanes/growthcharts>).

Os procedimentos mais simples para comparar os dados obtidos com o padrão de referência são de dois tipos: a classificação pela medição do percentil e o sistema de desvio-padrão (DP). O sistema de desvio-padrão ou escore Z, que foi aqui utilizado, localiza o peso/idade da criança avaliada em números de desvios-padrão do valor mediano da população de referência. Um escore Z negativo indica que a medida da criança encontra-se abaixo do padrão de referência.

A classificação de referência para o escore Z, segundo Sawaya (1996), é como segue:

1. Desnutrição grave: $Z < -3$ DP;
2. Desnutrição moderada: $-3 \leq Z < -2$ DP;
3. Desnutrição leve: $-2 \leq Z < -1$ DP;
4. Eutrófica: $Z \geq -1$ DP.

A descrição e as respectivas distribuições de frequências das variáveis sócio-econômicas, disponíveis no banco de dados, consideradas como explicativas da variável WAZ são apresentadas no Quadro 1. É importante salientar que das 477 crianças incluídas no estudo temos que 32,5% de suas mães tinham idade menor que 20 anos, 37,1% das mães tinham

até 4 anos de escolaridade, 7,3% das crianças nasceram com baixo peso. Um diferencial sócio-econômico observado foi a existência de geladeira em casa, que através do Quadro 1 verificamos que 44,4% das famílias, nesse estudo, não possuem geladeira em casa e quanto a renda *per capita*, cerca de 70,0% das famílias têm essa renda abaixo de 0,50 do salário mínimo, que na época da coleta dos dados era de R\$180,00 e ainda 34,7% das mães informam morarem em domicílios com 6 pessoas ou mais, excluindo a criança pertencente ao estudo.

Quadro 1: Distribuição de Frequência das variáveis sócio-econômicas

Variáveis	Categoria	Codificação	Frequência (%)
Idade da mãe	≥ 20 anos	0	322 (67,5)
	< 20 anos	1	155 (32,5)
Anos de escolaridade da mãe	≥ 5 anos	0	300 (62,9)
	≤ 4 anos	1	177 (37,1)
Peso ao nascer da criança	≥ 3.000 (sobrepeso)	0	277 (58,1)
	2.500 + 3.000 (peso adequado)	1	165 (34,6)
	< 2.500 (baixo peso/peso insuficiente)	2	35 (7,3)
Sexo da criança	feminino	0	236 (49,5)
	masculino	1	241 (50,5)
Existência de geladeira	sim	0	265 (55,6)
	não	1	212 (44,4)
Renda familiar <i>per capita</i> (em S.M.)	$\geq 0,50$	0	132 (29,5)
	0,25 + 0,50	1	139 (31,1)
	$< 0,25$	2	176 (39,4)
Total de moradores	≤ 5	0	311 (65,3)
	≥ 6	1	165 (34,7)

Fonte dos dados básicos: arquivos da Pesquisa.

As variáveis explicativas consideradas e que variam ao longo do tempo são: dias em aleitamento materno (mamada), dias com diarreia (diarréia), tipo de alimentação (alimentação) e outros tipos de alimentos (outros), onde a variável “alimentação” corresponde a água, chá, suco ou leite artificial e a variável “outros”, que são os outros tipos de alimentos, além dos citados acima. As medidas descritivas dessas variáveis são apresentadas no Quadro 2, de onde se pode destacar que para a variável dias em aleitamento ou “mamada” notamos que a mediana das quantidades acumuladas, aos dois meses, foi de 56 dias, ou seja, em quase 60 dias de vida 50% das crianças mamaram até 56 dias. Aos 4 meses, cerca de 120 dias, o número mediano de dias que mamaram foi de 101, isto é, 50% das crianças mamaram até 101 dias, o que indica que algumas mães já estavam introduzindo outro tipo de alimentação. Essa defasagem do número de dias com os dias que mamaram fica mais acentuada ao longo dos meses pois é possível verificar a partir dos 6 meses que a mediana se mantém em 105

dias.

Em relação à variável dias com diarreia ou “diarreia” temos que aos 2 meses 50% das crianças ainda não tinham apresentado nenhum episódio enquanto que a média foi de mais que 1 dia. Já aos 6 meses a mediana foi de 2 dias, enquanto que a média (4,86) foi mais que o dobro da mediana. Tanto as variáveis dias de alimentação, quanto outros tipos de alimentação tiveram suas médias bem próximas das medianas e também próximas ao número de dias, ao longo dos meses, o que indica que as mães, nos primeiros dias de vida das crianças, já utilizam outros tipos de alimentos além do aleitamento materno na alimentação dessas crianças.

A medida antropométrica WAZ (escore Z - peso por idade), utilizada como variável dependente, foi categorizada em três níveis:

0 - Se $WAZ < -2 DP$ (desnutrição moderada a grave);

1 - Se WAZ estiver entre $[-2, -1] DP$ (desnutrição leve);

2 - Se $WAZ \geq -1 DP$ (eutrófico).

A medida WAZ (peso por idade), calculada em cada tempo, cujas distribuições de frequências encontram-se no Quadro 3, revela que a prevalência da desnutrição grave é crescente até os 12 meses, variando desde 1,5% aos 2 meses atingindo 6,9% aos 12 meses quando apresenta um leve declínio atingindo 5,9% aos 18 meses. Quanto à desnutrição leve, a menor prevalência observada foi aos 4 meses (6,9%) com evolução crescente, atingindo 22,9% aos 15 meses, com um pequeno declínio aos 18 meses.

Foi então ajustado um modelo marginal para dados longitudinais, considerando a classificação antropométrica WAZ como variável dependente (categorizada em três níveis) e como variáveis independentes: peso ao nascer da criança, existência de geladeira, anos de estudo da mãe, sexo da criança e total de moradores, categorizadas da forma apresentada no Quadro 1.

Inicialmente foram ajustados modelos de regressão logística para variável resposta ordinal WAZ, em cada tempo, considerando todas as variáveis disponíveis no banco de dados. A variável sexo não apareceu significativa estatisticamente em nenhum dos modelos, porém, como ficou comprovado que existe diferença significativa entre as médias do peso ao nascer segundo o sexo ($p\text{-valor}=0,003$) optou-se por deixar essas duas variáveis no modelo. Como se pretendia analisar a influência do tamanho da família na classificação nutricional foi incluída no modelo o total de moradores, embora nas análises preliminares não tenha sido significativa estatisticamente para explicar WAZ, em nenhum dos momentos, já que estudos sócio-demográficos na área nutricional, evidenciam que a quantidade de moradores da residência afeta o estado nutricional dos residentes, principalmente das crianças. Tendo em vista que no modelo ajustado, com os dados aos 6 meses, apareceram significantes as variáveis “mamada”, “peso ao nascer”, “geladeira” e “renda familiar”, foi calculada a correlação entre “geladeira” e “renda familiar”, já que a variável “geladeira” também reflete o diferencial

Quadro 2: Medidas descritivas das variáveis observadas ao longo dos meses (valores em dias).

Variáveis Dependentes do tempo	Medidas	Tempos (meses)								
		2	4	6	9	12	15	18		
Dias em aleitamento	Mediana Média	56 49,48	101 86,49	105 110,81	105 136,49	105 155,51	105 170,3	105 183,30		
Dias com diarreia	Mediana Média	0 1,46	1 3,18	2 4,86	4 7,49	7 11,01	10 14,44	12 17,35		
Dias de consumo de alimentos ¹	Mediana Média	54 47,46	116 106,94	179 168,64	269 259,04	360 349,44	450 440,07	540 530,74		
Dias de consumo de outros ² alimentos	Mediana Média	32 27,88	93 77,57	155 134,33	245 223,19	335 313,09	426 403,51	516 494,04		

Fonte dos dados básicos: arquivos da Pesquisa.

1 - Alimentação: água, chá, suco, leite artificial.

2 - Outros alimentos.

Quadro 3: Distribuição de frequência de frequência da medida antropométrica WAZ ao longo dos meses* .

WAZ Categoria	Classificação da categoria	Tempos (meses)								
		2	4	6	9	12	15	18		
0	desnutrição grave	7 (1,5)	8 (1,7)	15 (3,1)	25 (5,2)	33 (6,9)	29 (6,1)	28 (5,9)		
1	desnutrição leve	45 (9,4)	33 (6,9)	56 (11,7)	97 (20,3)	107 (22,4)	109 (22,9)	97 (20,3)		
2	normal (eutrófico)	425 (89,1)	436 (91,4)	406 (85,1)	355 (74,4)	337 (70,6)	339 (71,1)	352 (73,8)		
	Total	477 (100,0)	477 (100,0)	477 (100,0)	477 (100,0)	477 (100,0)	477 (100,0)	477 (100,0)		

Fonte dos dados básicos: arquivos da Pesquisa.

* Valores entre parenteses representam porcentagens.

sócio-econômico das famílias aqui estudadas e, foi encontrada uma correlação de Pearson positiva de 0,350 com significância estatística (p-valor=0,000). Por ser a “geladeira” uma variável que aparece significativa, um número maior de vezes, ao longo dos meses, foi preferido construir o modelo com a variável “geladeira” e não com a variável “renda familiar”.

Para as variáveis adotadas para a construção dos modelos, considere a matriz $\mathbf{X}_t = (\mathbf{x}_{1t}, \mathbf{x}_{2t}, \mathbf{x}_{3t}, \mathbf{x}_{4t}, \mathbf{x}_{5t}, \mathbf{x}_{6t})$, onde \mathbf{x}_{1t} é a quantidade acumulada de dias em aleitamento materno (mamadas) no tempo t , isto é, a variável “mamada” é dependente do tempo, \mathbf{x}_{2t} é a variável peso ao nascer, \mathbf{x}_{3t} é a variável existência de geladeira, \mathbf{x}_{4t} é a variável relacionada com a quantidade de anos de escolaridade da mãe, \mathbf{x}_{5t} é a variável sexo da criança e \mathbf{x}_{6t} é a variável total de moradores. Todas essas variáveis foram categorizadas como já mostrado no Quadro 1. É importante observar que todas as variáveis, exceto “mamada”, são constantes no tempo, logo, para simplificar a notação serão denotadas por \mathbf{x}_{1t} , \mathbf{x}_2 , \mathbf{x}_3 , \mathbf{x}_4 , \mathbf{x}_5 e \mathbf{x}_6 .

5.2 Modelo ajustado usando SAS

O modelo aqui ajustado foi obtido utilizando o pacote computacional SAS com a estrutura de correlação de independência com intercepto constante ao longo do tempo. Os erros-padrão foram calculados usando os estimadores da variância robusta.

Considerando o modelo logit para dados longitudinais, equação (4.1)

$$\text{logit} [\gamma_k(\mathbf{x}_{it})] = \alpha_k^* + \mathbf{x}'_{it}\beta, \quad (5.1)$$

onde os índices t iguais a t_1, t_2, \dots, t_7 referem-se aos tempos 2, 4, \dots , 18, respectivamente, em meses de vida da criança; e k correspondente ao nível $k = 1$ se WAZ=0 e $k = 2$ se WAZ=1, e $\pi_k(\mathbf{x}_t)$ a probabilidade correspondente a média das probabilidades de todos os grupos, para todas as crianças na população de interesse e que a resposta seja a categoria (ou nível) k e $\gamma_k(\mathbf{x}_t) = \pi_1(\mathbf{x}_t) + \dots + \pi_k(\mathbf{x}_t)$. Assim o modelo (4.1) é dado por

$$\text{logit} [\gamma_k(\mathbf{x}_t)] = \alpha_k^* + \beta_{1t}\mathbf{x}_{1t} + \beta_2\mathbf{x}_2 + \beta_3\mathbf{x}_3 + \beta_4\mathbf{x}_4 + \beta_5\mathbf{x}_5 + \beta_6\mathbf{x}_6. \quad (5.2)$$

Os parâmetros estimados pelo modelo, supondo a metodologia GEE e usando o procedimento GENMOD, rotina do SAS (Ver Anexo), apresentados no Quadro 4, revelam que a variável “mamada” deixa de ser significativa aos 9 meses ($t = t_4$), o que indica a existência de um outro fator associado com a variável WAZ. Verificou-se, como já era previsível pela análise preliminar, que as estimativas dos parâmetros das variáveis sexo e total de moradores não são significantes, mas, optou-se por deixá-las no modelo pela sua importância relativa.

Para esse conjunto de dados os modelos ajustados para cada nível k são

$$\text{logit}[\hat{\gamma}_1(\mathbf{x}_t)] = -4.1943 + \hat{\beta}_{1t}x_{1t} + \hat{\beta}_2x_2 + \hat{\beta}_3x_3 + \hat{\beta}_4x_4 + \hat{\beta}_5x_5 + \hat{\beta}_6x_6,$$

e

$$\text{logit}[\hat{\gamma}_2(\mathbf{x}_t)] = -2.2246 + \hat{\beta}_{1t}x_{1t} + \hat{\beta}_2x_2 + \hat{\beta}_3x_3 + \hat{\beta}_4x_4 + \hat{\beta}_5x_5 + \hat{\beta}_6x_6,$$

onde $\hat{\beta}_j$ são as estimativas dos parâmetros apresentadas no Quadro 4. Assim as probabilidades acumuladas são obtidas da forma

$$\hat{\gamma}_k(\mathbf{x}_t) = \frac{\exp(\hat{\alpha}_k^* + \mathbf{x}_t' \hat{\beta})}{1 + \exp(\hat{\alpha}_k^* + \mathbf{x}_t' \hat{\beta})}, \quad k = 1, 2. \quad (5.3)$$

Analisando os valores de cada uma das razões de chance (exponencial das estimativas dos parâmetros, apresentadas no Quadro 4), pode-se destacar que as chances de uma criança ter qualquer tipo de desnutrição ao longo do tempo é 3 vezes maior quando nasce com peso adequado em relação à criança que nasce com sobrepeso e 10 vezes maior quando nasce com baixo peso em relação às que nascem com sobrepeso. Quando a casa onde mora a criança não tem geladeira, faz a criança ter quase 2 vezes mais chance de ter qualquer tipo de desnutrição em relação à criança que mora na casa com geladeira.

Quadro 4: Estimativas dos parâmetros para o modelo marginal usando o pacote SAS.

Parâmetros ⁽¹⁾	Estimativas	Erro padrão	p-valor	Exponencial das estimativas
α_1^*	-4,1943	0,3177	<,0001	0,0151
α_2^*	-2,2246	0,2521	<,0001	0,1081
β_{1t_1}	-0,0260	0,0038	<,0001	0,9743
β_{1t_2}	-0,0168	0,0024	<,0001	0,9833
β_{1t_3}	-0,0078	0,0016	<,0001	0,9922
β_{1t_4}	-0,0011	0,0010	0,2664	0,9989
β_{1t_5}	0,0001	0,0008	0,8760	0,8278
β_{1t_6}	0,0004	0,0007	0,5859	1,0001
β_{1t_7}	-0,0002	0,0007	0,7145	0,9998
β_2	1,1590	0,1476	<,0001	3,1867 ⁽²⁾ /10,1553 ⁽³⁾
β_3	0,5534	0,2003	0,0057	1,7392
β_4	0,4242	0,2004	0,0343	1,5284
β_5	0,1919	0,1930	0,3201	1,2115
β_6	0,0924	0,1952	0,6359	1,0968

Fonte dos dados básicos: arquivos da Pesquisa;

(1) - Os índices t_1, t_2, \dots, t_7 referem-se aos valores da variável mamada aos 2, 4, \dots , 18 meses de vida da criança;

(2) - Estimativa referente à variável “peso” quando a resposta é categoria 1;

(3) - Estimativa referente à variável “peso” quando a resposta é categoria 2.

As probabilidades ajustadas pelo modelo, para cada sexo, encontram-se apresentadas nos Quadros 5 e 6 para meninas e meninos, respectivamente, adiante inseridos. Com o objetivo de generalizar as conclusões, foram feitas estimativas em duas situações distintas considerando combinações de algumas categorias das variáveis “mamada” e condição geral

ou “cenário”. A condição geral foi criada a partir de combinações das variáveis explicativas, exceto a variável sexo. Foram considerados como tendo dois “cenários”:

1 - Pior cenário: as crianças que nasceram com baixo peso ($x_2 = 2$), nas casas onde moram não têm geladeira ($x_3 = 1$), as mães das crianças estudaram menos de 5 anos ($x_4 = 1$) e na residência moram mais que 6 pessoas, excluindo a criança ($x_6 = 1$);

2 - Melhor cenário: as crianças que nasceram com sobrepeso ($x_2 = 0$), nas casas onde moram têm geladeira ($x_3 = 0$), as mães das crianças estudaram mais de 4 anos ($x_4 = 0$) e na residência moram no máximo 5 pessoas, excluindo a criança ($x_6 = 0$).

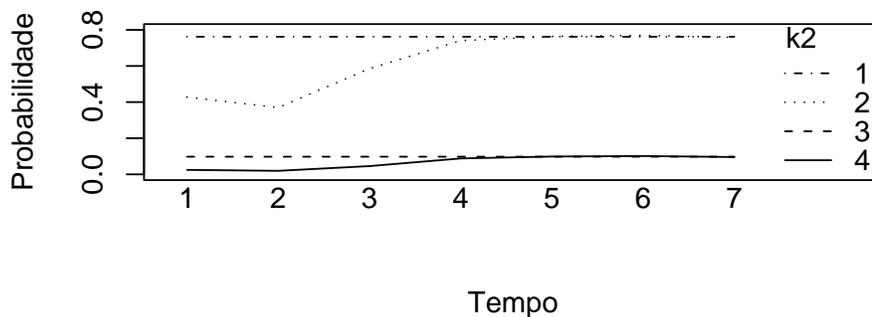
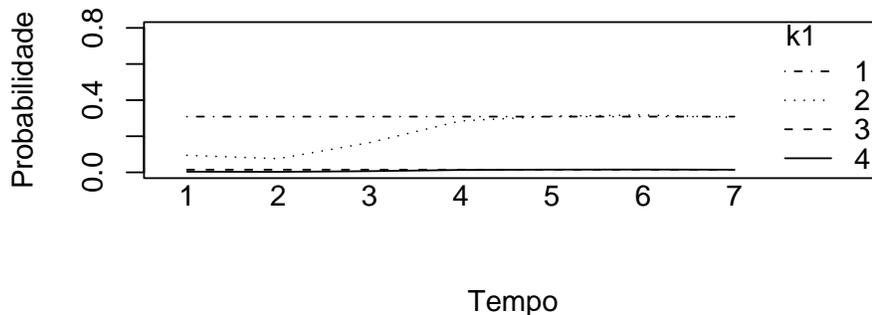
Também foi considerado a condição de aleitamento da criança em relação à quantidade de dias de mamada tendo duas classes: nunca mamou e mamou um número mediano de dias. Os Quadros 5 e 6 apresentam as probabilidades acumuladas calculadas a partir de (5.3) segundo situação geral e condição de aleitamento para crianças do sexo feminino e masculino, respectivamente.

A partir do Quadro 5 constatou-se que uma menina, aos 2 meses, que nunca mamou, no pior cenário, tem probabilidade de ter desnutrição grave de 0,3087 e de ter qualquer tipo de desnutrição (leve ou grave) de 0,7619. No melhor cenário, tem-se que a probabilidade de ter desnutrição grave é de 0,0149 e ter qualquer tipo de desnutrição é de 0,0976, ou seja, se uma criança do sexo feminino nunca mamou, a probabilidade de ter qualquer tipo de desnutrição se pertence ao pior “cenário” da condição geral é quase 8 vezes maior do que se a criança tem a condição classificada como no melhor “cenário”.

Considerando que a criança mamou uma quantidade mediana de dias tem-se que a probabilidade de ter desnutrição grave é 0,0943 e ter qualquer tipo de desnutrição é de 0,4274 aos dois meses de vida da criança, no pior “cenário”. Aos 15 meses, a probabilidade de ter desnutrição grave é 0,3177, cerca de três vezes maior que probabilidade aos 2 meses mantendo a mesma condição geral e, de ter algum tipo de desnutrição a probabilidade é 0,7695, cerca de mais de 80% que aos dois meses. Na melhor condição geral, para o número mediano de mamadas, temos que a probabilidade de ter desnutrição grave é de 0,0035 e ter qualquer tipo de desnutrição de 0,0246 nos dois primeiros meses. Aos 15 meses, tem-se que a probabilidade de ter desnutrição grave é 0,0155, enquanto que para a desnutrição leve ou grave a probabilidade é de 0,1013. Portanto, tem-se que no melhor cenário as probabilidades de ter desnutrição grave são bastante baixas. Estas probabilidades podem ser observadas através da Figura 5.1, na qual o primeiro gráfico em que $k1$ se refere às probabilidades de desnutrição moderada à grave e o segundo gráfico em que $k2$ se refere às probabilidades de qualquer tipo de desnutrição. Nesta figura é facilmente notado que a partir do tempo $t = 4$, ou seja, aos 9 meses, as probabilidades se estabilizam em relação à mesma condição geral, independente do estado nutricional da criança, passando a não existir diferença de nunca ter mamado ou ter mamado um número mediano de dias. Em resumo, pode-se concluir que

até 9 meses de vida além da condição sócio-econômica a condição nutricional é explicada pela quantidade de dias em aleitamento mas, a partir de 9 meses apenas a condição sócio-econômica explica a condição nutricional.

Figura 5.1: Probabilidades acumuladas de desnutrição (modelo marginal) na “condição geral”, das crianças do sexo feminino, acompanhadas no tempo $t = 1, \dots, 7$.



- k1 - Probabilidades referentes à desnutrição moderada a grave;
- k2 - Probabilidades referentes a qualquer tipo de desnutrição;
- 1 - Referente ao pior cenário quando nunca mamou;
- 2 - Referente ao pior cenário quando mamou número mediano de dias;
- 3 - Referente ao melhor cenário quando nunca mamou;
- 4 - Referente ao melhor cenário quando mamou número mediano de dias;

Com relação aos meninos, conforme mostram os dados do Quadro 6, tem-se que um menino que não mamou ao longo dos meses, no pior cenário tem a probabilidade de ter desnutrição grave de 0,3511 e ter qualquer tipo de desnutrição (leve ou grave) de 0,7949. No melhor cenário, tem-se que a probabilidade de ter desnutrição grave é de 0,0179 e ter qualquer tipo de desnutrição é de 0,1158.

Considerando agora que o menino mamou um número acumulado mediano de vezes em

Quadro 5: Probabilidades acumuladas de desnutrição (modelo marginal), segundo condição geral - Crianças do sexo feminino.

Situação do aleitamento	Condição Geral	Situação Nutricional (k)*	Tempos (meses)								
			2	4	6	9	12	15	18		
Nunca mamou	Pior cenário	1	0,3087	0,3087	0,3087	0,3087	0,3087	0,3087	0,3087	0,3087	0,3087
		2	0,7619	0,7619	0,7619	0,7619	0,7619	0,7619	0,7619	0,7619	0,7619
	Melhor cenário	1	0,0149	0,0149	0,0149	0,0149	0,0149	0,0149	0,0149	0,0149	0,0149
		2	0,0976	0,0976	0,0976	0,0976	0,0976	0,0976	0,0976	0,0976	0,0976
Mamou número mediano de dias	Pior cenário	1	0,0943	0,0756	0,1645	0,2846	0,3109	0,3177	0,3109	0,3177	0,3042
		2	0,4274	0,3697	0,5853	0,7404	0,7638	0,7695	0,7638	0,7695	0,7581
	Melhor cenário	1	0,0035	0,0028	0,0066	0,0133	0,0150	0,0133	0,0150	0,0155	0,0146
		2	0,0246	0,0194	0,0455	0,0879	0,0985	0,1013	0,0879	0,0985	0,1013

Fonte dos dados básicos: arquivos da Pesquisa

* - São os níveis da variável resposta. $\{k = 1$ se $waz < -2$ Desvios-padrão, e $k = 2$ se waz estiver entre $[-2, 1]$ }.

Quadro 6: Probabilidades acumuladas de desnutrição (modelo marginal), segundo condição geral - Crianças do sexo masculino.

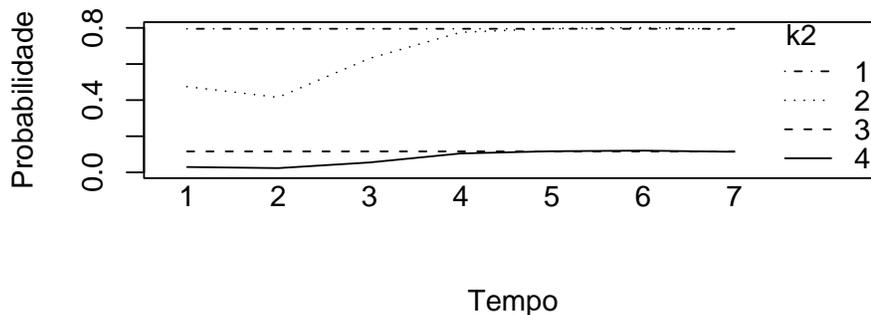
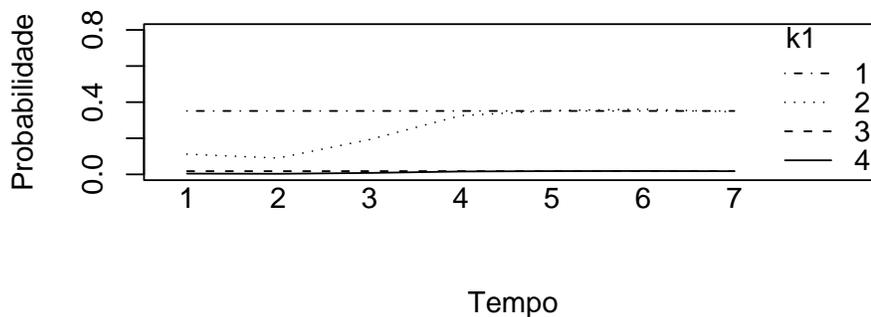
Situação do aleitamento	Condição Geral	Situação Nutricional (k)*	Tempos (meses)								
			2	4	6	9	12	15	18		
Nunca mamou	Pior cenário	1	0,3511	0,3511	0,3511	0,3511	0,3511	0,3511	0,3511	0,3511	0,3511
		2	0,7949	0,7949	0,7949	0,7949	0,7949	0,7949	0,7949	0,7949	0,7949
Mamou número mediano de dias	Melhor cenário	1	0,0179	0,0179	0,0179	0,0179	0,0179	0,0179	0,0179	0,0179	0,0179
		2	0,1158	0,1158	0,1158	0,1158	0,1158	0,1158	0,1158	0,1158	0,1158
	Pior cenário	1	0,1120	0,0902	0,1926	0,3252	0,3535	0,3607	0,3607	0,3607	0,3463
		2	0,4748	0,4154	0,6310	0,7755	0,7967	0,8018	0,8018	0,8018	0,7915
	Melhor cenário	1	0,0042	0,0033	0,0080	0,0160	0,0181	0,0187	0,0187	0,0187	0,0176
		2	0,0296	0,0234	0,0546	0,1045	0,1169	0,1202	0,1202	0,1202	0,1136

Fonte dos dados básicos: arquivos da Pesquisa

* - São os níveis da variável resposta. { $k = 1$ se $waz < -2$ Desvios-padrão, e $k = 2$ se waz estiver entre $[-2, 1)$ }.

cada mês, tem-se que a probabilidade de ter desnutrição grave é 0,1120 e ter qualquer tipo de desnutrição é de 0,4748 aos dois meses de vida da criança no pior cenário. Aos 15 meses, a probabilidade de ter desnutrição grave é 0,3607, cerca de três vezes maior que probabilidade aos 2 meses, mantendo as mesmas condições sociais e, para as crianças que apresentaram qualquer tipo de desnutrição a probabilidade é 0,8018, quase 70% maior que aos dois meses. Na melhor situação social, para o número mediano de mamadas, tem-se que a probabilidade de ter desnutrição grave é de 0,0042 e ter qualquer tipo de desnutrição de 0,0296 nos dois primeiros meses. Aos 15 meses, tem-se que a probabilidade de ter desnutrição grave é 0,0187, enquanto que para a desnutrição leve ou grave a probabilidade é de 0,1202.

Figura 5.2: Probabilidades acumuladas de desnutrição (modelo marginal) na “condição geral”, das crianças do sexo masculino, acompanhadas no tempo $t = 1, \dots, 7$.

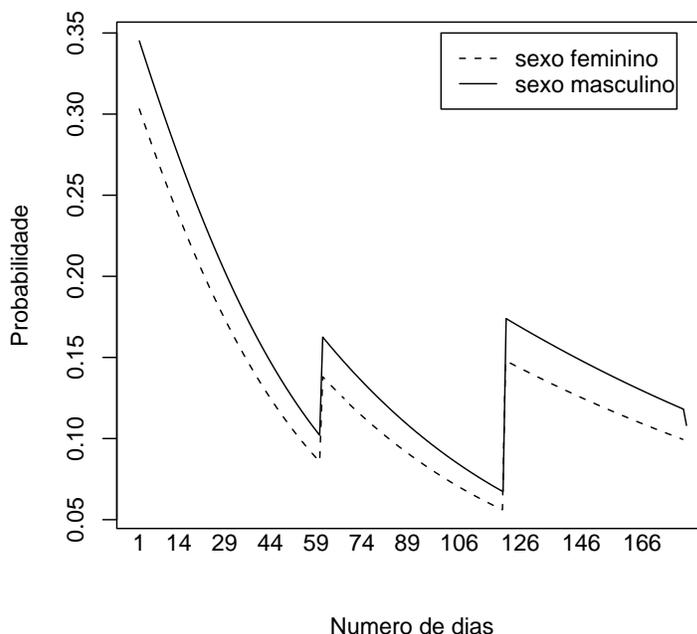


- k1 - Probabilidades referentes à desnutrição moderada a grave;
- k2 - Probabilidades referentes a qualquer tipo de desnutrição;
- 1 - Referente ao pior cenário quando nunca mamou;
- 2 - Referente ao pior cenário quando mamou número mediano de dias;
- 3 - Referente ao melhor cenário quando nunca mamou;
- 4 - Referente ao melhor cenário quando mamou número mediano de dias;

Através da Figura 5.2, observamos que as probabilidades de desnutrição das crianças do sexo masculino têm comportamento similar ao das crianças do sexo feminino, o qual a partir do tempo $t = 4$, referente aos 9 meses, essas probabilidades tendem a ser iguais dentro de um mesmo cenário, quando a criança nunca mamou ou mamou um número mediano de vezes.

No pior cenário, se a criança mama um número mediano de vezes, as probabilidades de ter desnutrição ao longo dos meses são menores do que para as crianças que nunca mamaram, sendo que essas probabilidades, para os meninos, são levemente maiores, o que mostra a maior dificuldade dos meninos em obter ganho de peso nos primeiros meses de vida, como pode ser visto através da Figura 5.3, embora possamos considerar esta diferença desprezível pois no modelo a variável sexo não apresenta coeficiente significativo.

Figura 5.3: Probabilidades de desnutrição moderada a grave, das crianças, de acordo com a quantidade de dias em aleitamento, no pior cenário, por sexo.



Em resumo concluímos que as probabilidades de desnutrição no pior cenário tanto para meninos quanto para meninas, ao longo dos meses, são consideravelmente maiores quando a criança não mama em relação às crianças que mamam um número mediano de vezes. Em geral podemos concluir que as probabilidades de desnutrição são as mais altas durante todo o tempo quando a criança nunca mamou e vive no pior “cenário” e, aumentam ao longo do tempo para as crianças que mamaram um número mediano de vezes neste mesmo “cenário”.

5.3 Modelo ajustado usando R considerando intercepto constante no tempo

As estimativas para os parâmetros do modelo marginal com o ajuste obtido pelo R, considerando o intercepto constante no tempo são valores próximos das estimativas obtidas pelo SAS, considerando a estrutura de correlação para respostas independentes, isto pode ser visto comparando as estimativas através dos Quadros 4 e 7. É importante destacar também que a significância das estimativas para os parâmetros do Quadro 7 concordam com as encontradas no Quadro 4, isso nos indica que os dois pacotes, SAS e R, utilizam métodos similares de estimação para a obtenção das estimativas.

Quadro 7: Estimativas dos parâmetros para o modelo marginal usando o pacote R com intercepto constante no tempo.

Parâmetros ⁽¹⁾	Estimativas	Erro padrão	p-valor	Exponencial das estimativas
α_1^*	-3,9615	0,1952	0,0001	0,0189
α_2^*	-1,9861	0,1781	0,0001	0,1372
β_{1t_1}	-0,0274	0,0068	<,0001	0,9729
β_{1t_2}	-0,0178	0,0041	<,0001	0,9823
β_{1t_3}	-0,0102	0,0020	<,0001	0,9898
β_{1t_4}	-0,0027	0,0012	0,0238	0,9973
β_{1t_5}	0,0001	0,0009	0,9884	1,0001
β_{1t_6}	-0,0001	0,0008	0,9879	0,9999
β_{1t_7}	-0,0006	0,0008	0,5067	0,9994
β_2	1,0105	0,1044	0,0000	$2,7469^{(2)}/7,5458^{(3)}$
β_3	0,5318	0,1894	0,0049	1,7019
β_4	0,6365	0,1616	<,0001	1,8898
β_5	-0,0259	0,1609	0,8721	0,9744
β_6	-0,0484	0,1628	0,7662	0,9527

Fonte dos dados básicos: arquivos da Pesquisa;

(1) - Os índices t_1, t_2, \dots, t_7 referem-se aos valores da variável mamada aos 2, 4, \dots , 18 meses de vida da criança;

(2) - Estimativa referente à variável “peso” quando a resposta é categoria 1;

(3) - Estimativa referente à variável “peso” quando a resposta é categoria 2.

Como mencionado anteriormente, que as estimativas são próximas, para os pacotes SAS e R, temos que as probabilidades de desnutrição das crianças do sexo feminino obtidas para o modelo ajustado pelo R, Quadro 8, tem seu valores similares às probabilidades obtidas no Quadro 5.

Em relação ao Quadro 9, verificou-se que também as probabilidades calculadas são bem

Quadro 8: Probabilidades acumuladas de desnutrição, obtidas do modelo com intercepto constante, segundo condição geral - Crianças do sexo feminino.

Situação do aleitamento	Condição Geral	Situação Nutricional (k)*	Tempos (meses)								
			2	4	6	9	12	15	18		
Nunca mamou	Pior cenário	1	0,3056	0,3056	0,3056	0,3056	0,3056	0,3056	0,3056	0,3056	0,3056
		2	0,7604	0,7604	0,7604	0,7604	0,7604	0,7604	0,7604	0,7604	0,7604
Mamou número mediano de dias	Melhor cenário	1	0,0187	0,0187	0,0187	0,0187	0,0187	0,0187	0,0187	0,0187	0,0187
		2	0,1207	0,1207	0,1207	0,1207	0,1207	0,1207	0,1207	0,1207	0,1207
	Pior cenário	1	0,0871	0,0680	0,1311	0,2490	0,3058	0,3058	0,3058	0,3054	0,2924
		2	0,4076	0,3445	0,5209	0,7050	0,7606	0,7606	0,7602	0,7602	0,7487
	Melhor cenário	1	0,0041	0,0031	0,0065	0,0141	0,0187	0,0187	0,0187	0,0187	0,0176
		2	0,0289	0,0222	0,0449	0,0937	0,1208	0,1208	0,1206	0,1206	0,1141

Fonte dos dados básicos: arquivos da Pesquisa

* - São os níveis da variável resposta. { $k = 1$ se $waz < -2$ Desvios-padrão, e $k = 2$ se waz estiver entre $[-2, 1)$ }.

Quadro 9: Probabilidades acumuladas de desnutrição, obtidas do modelo com intercepto constante, segundo condição geral - Crianças do sexo masculino.

Situação do aleitamento	Condição Geral	Situação Nutricional (k)*	Tempos (meses)								
			2	4	6	9	12	15	18		
Nunca mamou	Pior cenário	1	0,3001	0,3001	0,3001	0,3001	0,3001	0,3001	0,3001	0,3001	0,3001
	Melhor cenário	2	0,7556	0,7556	0,7556	0,7556	0,7556	0,7556	0,7556	0,7556	0,7556
Mamou número mediano de dias	Pior cenário	1	0,0182	0,0182	0,0182	0,0182	0,0182	0,0182	0,0182	0,0182	0,0182
	Melhor cenário	2	0,1179	0,1179	0,1179	0,1179	0,1179	0,1179	0,1179	0,1179	0,1179
	Pior cenário	1	0,0851	0,0663	0,1281	0,2441	0,3004	0,2999	0,2871	0,2871	0,2871
	Melhor cenário	2	0,4013	0,3387	0,5144	0,6996	0,7558	0,7554	0,7438	0,7438	0,7438
	Pior cenário	1	0,0040	0,0031	0,0063	0,0138	0,0182	0,0171	0,0171	0,0171	0,0171
	Melhor cenário	2	0,0282	0,0217	0,0438	0,0915	0,1181	0,1178	0,1178	0,1178	0,1178

Fonte dos dados básicos: arquivos da Pesquisa

* - São os níveis da variável resposta. { $k = 1$ se $waz < -2$ Desvios-padrão, e $k = 2$ se waz estiver entre $[-2, 1)$ }.

próximas das probabilidades encontradas para desnutrição em crianças do sexo masculino (Quadro 6).

5.4 Modelo ajustado usando R considerando intercepto não constante no tempo

Considerando que o intercepto do modelo logit não é constante, utilizando o pacote R, foi possível definir dois tipos de estruturas de correlação: independência e permutável.

Se define modelos logits para $\Upsilon_k(\mathbf{x}_t)$, a média das probabilidades de todos os grupos em que a resposta seja a categoria k como:

$$\text{logit}[\hat{\Upsilon}_1(\mathbf{x}_t)] = \hat{\alpha}_{1t}^* + \hat{\gamma}_{1t}^*x_{1t} + \hat{\gamma}_2^*x_2 + \hat{\gamma}_3^*x_3 + \hat{\gamma}_4^*x_4 + \hat{\gamma}_5^*x_5 + \hat{\gamma}_6^*x_6,$$

e

$$\text{logit}[\hat{\Upsilon}_2(\mathbf{x}_t)] = \hat{\alpha}_{2t}^* + \hat{\gamma}_{1t}^*x_{1t} + \hat{\gamma}_2^*x_2 + \hat{\gamma}_3^*x_3 + \hat{\gamma}_4^*x_4 + \hat{\gamma}_5^*x_5 + \hat{\gamma}_6^*x_6.$$

Assim as probabilidades estimadas são obtidas da forma

$$\hat{\Upsilon}_k(\mathbf{x}_t) = \frac{\exp(\hat{\alpha}_{kt}^* + \mathbf{x}_t' \hat{\gamma}^*)}{1 + \exp(\hat{\alpha}_{kt}^* + \mathbf{x}_t' \hat{\gamma}^*)}, \quad k = 1, 2. \quad (5.4)$$

Com o objetivo de comparar as estimativas do SAS e do R para os modelos marginais, foi necessário trocar o sinal das estimativas do R, já que os modelos propostos por Liang & Zeger (1986), utilizam as covariáveis com as respostas acumuladas até a categoria k , ver modelo (5.2).

Os parâmetros estimados pelo modelo, supondo a metodologia GEE2 e usando a rotina do R, apresentados no Quadro 10, considerando a estrutura de correlação de independência, e, no Quadro 11 considerando a estrutura de correlação permutável, revelam que a variável “mamada” deixa de ser significativa a partir dos 12 meses ($t = t5$), o que confirma a indicação da existência de um outro fator associado com a variável WAZ a partir desse momento. Temos ainda, como já notado pela análise preliminar, que as estimativas dos parâmetros das variáveis sexo e total de moradores, em ambos os modelos, também não são significantes. As estimativas dos coeficientes do modelo marginal considerado, ($\hat{\gamma}_{1t}^*$, $\hat{\gamma}_2^*$, $\hat{\gamma}_3^*$, $\hat{\gamma}_4^*$, $\hat{\gamma}_5^*$ e $\hat{\gamma}_6^*$), tem interpretações similares aos já descritos para o modelo calculado pelo pacote SAS.

Quadro 10: Estimativas dos parâmetros para o modelo usando o R, considerando estrutura de correlação de independência.

Parâmetros ⁽¹⁾	Estimativas	Erro-padrão		p-valor
		Naive	Robusto	
α_{1t1}^*	-4,3146	0,3864	0,3612	0,0000
α_{2t1}^*	-1,9810	0,2831	0,2299	0,0000
α_{1t2}^*	-3,7221	0,3395	0,2868	0,0000
α_{2t2}^*	-2,0050	0,2809	0,2236	0,0000
α_{1t3}^*	-3,7134	0,3360	0,2784	0,0000
α_{2t3}^*	-2,0051	0,2790	0,2183	0,0000
α_{1t4}^*	-3,7816	0,3396	0,2799	0,0000
α_{2t4}^*	-2,1181	0,2810	0,2197	0,0000
α_{1t5}^*	-4,0688	0,3597	0,3001	0,0000
α_{2t5}^*	-1,9764	0,2772	0,2166	0,0000
α_{1t6}^*	-4,0914	0,3617	0,3145	0,0000
α_{2t6}^*	-1,9548	0,2761	0,2202	0,0000
α_{1t7}^*	-4,2245	0,3761	0,3430	0,0000
α_{2t7}^*	-1,8666	0,2749	0,2160	0,0000
γ_{1t1}^*	-0,0270	0,0080	0,0068	<,0001
γ_{1t2}^*	-0,0175	0,0052	0,0042	<,0001
γ_{1t3}^*	-0,0100	0,0034	0,0021	<,0001
γ_{1t4}^*	-0,0026	0,0018	0,0012	0,0289
γ_{1t5}^*	-0,0000	0,0013	0,0009	0,9859
γ_{1t6}^*	-0,0001	0,0011	0,0008	0,9567
γ_{1t7}^*	-0,0006	0,0010	0,0009	0,5194
γ_2^*	1,0175	0,1388	0,1055	0,0000
γ_3^*	0,5237	0,2131	0,1901	0,0059
γ_4^*	0,6302	0,2071	0,1624	0,0001
γ_5^*	-0,0241	0,1975	0,1629	0,8821
γ_6^*	-0,0513	0,2070	0,1639	0,7542

Fonte dos dados básicos: arquivos da Pesquisa;

(1) - Os índices $t1, t2, \dots, t7$ referem-se aos valores da variável mamada aos 2, 4, \dots , 18 meses de vida da criança.

Como mostrado nos Quadros 10 e 11, os erros-padrão (EP) “Naive” e Robusto não são muito diferentes, contudo, foram considerados na análise apenas os erros-padrão robusto. É verificado também que as estimativas dos parâmetros de regressão, nas duas diferentes estruturas de correlação, são estimativas bastante similares.

Quadro 11: Estimativas dos parâmetros para o modelo usando o R, considerando estrutura de correlação de permutável.

Parâmetros ⁽¹⁾	Estimativas	Erro-padrão		p-valor
		Naive	Robusto	
α_{1t1}^*	-4,3088	0,3917	0,3626	0,0000
α_{2t1}^*	-1,9678	0,2896	0,2308	0,0000
α_{1t2}^*	-3,7344	0,3510	0,2632	0,0000
α_{2t2}^*	-1,9814	0,3513	0,2630	<,0001
α_{1t3}^*	-3,6958	0,3455	0,2523	0,0000
α_{2t3}^*	-2,0284	0,3453	0,2520	<,0001
α_{1t4}^*	-3,9699	0,3709	0,2690	0,0000
α_{2t4}^*	-1,6122	0,3707	0,2691	<,0001
α_{1t5}^*	-4,1664	0,3861	0,2744	0,0000
α_{2t5}^*	-1,7335	0,3857	0,2738	<,0001
α_{1t6}^*	-4,1305	0,3808	0,2798	0,0000
α_{2t6}^*	-1,8359	0,3806	0,2795	<,0001
α_{1t7}^*	-4,2202	0,3807	0,3416	0,0000
α_{2t7}^*	-1,8655	0,2819	0,2191	0,0000
γ_{1t1}^*	-0,0261	0,0080	0,0069	0,0001
γ_{1t2}^*	-0,0171	0,0053	0,0042	<,0001
γ_{1t3}^*	-0,0098	0,0035	0,0021	0,0004
γ_{1t4}^*	-0,0026	0,0019	0,0012	0,0323
γ_{1t5}^*	-0,0000	0,0013	0,0009	0,9947
γ_{1t6}^*	0,0001	0,0011	0,0008	0,8717
γ_{1t7}^*	-0,0006	0,0011	0,0009	0,4805
γ_2^*	1,0064	0,1439	0,1100	0,0000
γ_3^*	0,5180	0,2194	0,1937	0,0075
γ_4^*	0,6310	0,2134	0,1656	0,0001
γ_5^*	-0,0277	0,2035	0,1661	0,8675
γ_6^*	-0,0434	0,2135	0,1681	0,7961

Fonte dos dados básicos: arquivos da Pesquisa;

(1) - Os índices $t1, t2, \dots, t7$ referem-se aos valores da variável mamada aos 2, 4, \dots , 18 meses de vida da criança.

Para a validade da estrutura da matriz de correlação, temos que o parâmetro de correlação estimado, $\hat{\alpha}$, foi de -0,0369 e erro-padrão igual a 0,1509, o qual produziu uma estatística de Wald igual a -0.2448 (p-valor=0.5967), que sob a hipótese de que $\alpha = 0$ o teste não é rejeitado, indicando que o parâmetro da estrutura de correlação permutável não é significativo, ou seja, as estimativas aqui obtidas não são estatisticamente diferentes das estimativas obtidas com a estrutura de correlação supondo independência.

Os erros-padrão considerados foram os robustos que são estimados consistentemente através do estimador “sandwich” da matriz de covariância, contudo as estimativas dos erros-padrão, mostrados nos Quadros 10 e 11, não são muito diferentes daqueles que derivam do estimador Naive da matriz de variância.

Quadro 12: Comparação das estimativas dos parâmetros para o modelo marginal de diferentes estruturas de correlação, considerando a estatística $\hat{\gamma}^*/EP(\hat{\gamma}^*)$.

Parâmetros ⁽¹⁾	Independência	Permutável
α_{1t1}^*	11,9441	11,8831
α_{2t1}^*	8,6181	8,5250
α_{1t2}^*	12,9764	14,1897
α_{2t2}^*	8,9685	7,5339
α_{1t3}^*	13,3361	14,6463
α_{2t3}^*	9,1835	8,0488
α_{1t4}^*	13,5067	14,7543
α_{2t4}^*	9,6397	5,9908
α_{1t5}^*	13,5548	15,1805
α_{2t5}^*	9,1232	6,3303
α_{1t6}^*	13,0095	14,7638
α_{2t6}^*	8,8772	6,5675
α_{1t7}^*	12,3158	12,3552
α_{2t7}^*	8,6402	8,5134
γ_{1t1}^*	3,9821	3,7784
γ_{1t2}^*	4,1728	4,0783
γ_{1t3}^*	4,8024	4,6880
γ_{1t4}^*	2,1839	2,1403
γ_{1t5}^*	0,0177	0,0064
γ_{1t6}^*	0,05431	0,1615
γ_{1t7}^*	0,6443	0,7055
γ_2^*	-9,6447	-9,1441
γ_3^*	-2,7553	-2,6747
γ_4^*	-3,8810	-3,8092
γ_5^*	0,1482	0,1668
γ_6^*	0,3130	0,2583

Fonte dos dados básicos: arquivos da Pesquisa;

(1) - Os índices $t1, t2, \dots, t7$ referem-se aos valores da variável mamada aos 2, 4, \dots , 18 meses de vida da criança.

A nível de comparação das diferentes estruturas da matriz de correlação: a estrutura de independência da forma (3.9) e a estrutura permutável da forma (3.12), foi calculado

a estatística $\hat{\gamma}^*/EP(\hat{\gamma}^*)$, onde o $EP(\hat{\gamma}^*)$ considerado foi o robusto, também chamado de consistente.

É interessante notar que os coeficientes padronizados ($\hat{\gamma}^*/EP(\hat{\gamma}^*)$) são bem próximos considerando as diferentes estruturas da matriz de covariância. Assim por exemplo, Testes do tipo Wald para efeito das covariáveis produzirão níveis de significância muito similares, isso é também confirmado pelo p-valor mostrado nos Quadros 10 e 13.

Como notado no Quadro 12 para diferentes estruturas de correlação, a estatística considerada, produziu valores similares, sendo um pouco diferente para as estimativas de α_{2t} nos tempos $t = t_4$, $t = t_5$, e $t = t_6$, ou seja, para os meses 9, 12 e 15, respectivamente. Portanto as estimativas são parecidas, e isso é confirmado com o teste da significância do parâmetro α para a estrutura de correlação permutável. Podemos concluir que para esse conjunto de dados e usando o pacote R, as estimativas dos parâmetros do modelo podem ser consideradas as obtidas utilizando a estrutura de correlação de independência.

Através do Quadro 13, verificou-se que as probabilidades, em geral, tem crescimento e decréscimo similares às probabilidades mostradas no Quadro 6, em relação ao tempo. Contudo tem-se que as maiores diferenças detectadas encontram-se aos 2 meses e aos 18 meses, no pior cenário. Por exemplo, para uma criança do sexo masculino que não mamou e, no pior cenário, a probabilidade estimada de desnutrição grave foi de 0,3511 (Quadro 6) e, com esse novo enfoque foi estimada a probabilidade em 0,2313. Uma outra probabilidade estimada que apresentou maior diferença foi aos 18 meses na situação da criança ter desnutrição grave, no pior cenário e que mamou um número mediano de dias, foi de 0,3463 (Quadro 6) e de 0,2362 (Quadro 13). Embora os Quadros 6, 9 e 13 apresentem algumas diferenças nas probabilidades estimadas de desnutrição esses valores são bem próximos, como mostrado na Figura 5.4. Isso mostra que apesar do pacote SAS, utilizar o método GEE1, e o R, utilizar os dois métodos GEE1 e GEE2, para modelos com intercepto constante ou intercepto não constante no tempo, as estimativas das probabilidades de desnutrição são similares.

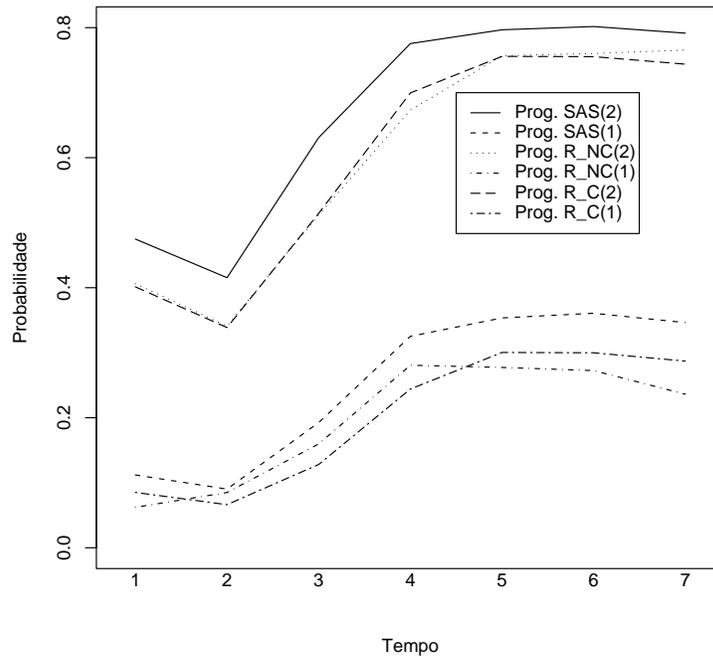
Quadro 13: Probabilidades estimadas do modelo com intercepto não constante de desnutrição em meninos, segundo condição geral e situação de aleitamento, usando o pacote R, com estrutura de independência.

Situação do aleitamento	Condição Geral	Situação Nutricional (k)*	Tempos (meses)						
			2	4	6	9	12	15	18
Nunca mamou	Pior cenário	1	0,2313	0,3524	0,3544	0,3389	0,2778	0,2733	0,2477
		2	0,7563	0,7518	0,7518	0,7302	0,7571	0,7611	0,7768
Mamou número mediano de dias	Melhor cenário	1	0,0129	0,0231	0,0233	0,0218	0,0164	0,0161	0,0141
		2	0,1187	0,1162	0,1162	0,1051	0,1192	0,1214	0,1312
	Pior cenário	1	0,0622	0,0850	0,1597	0,2807	0,2774	0,2725	0,2362
		2	0,4062	0,3410	0,5120	0,6731	0,7568	0,7603	0,7656
	Melhor cenário	1	0,0029	0,0044	0,0082	0,0166	0,0164	0,0160	0,0132
		2	0,0288	0,0220	0,0435	0,0820	0,1189	0,1220	0,1242

Fonte dos dados básicos: arquivos da Pesquisa

* - São os níveis da variável resposta $\{k = 1 \text{ se } waz < -2 \text{ Desvios-padrão, e } k = 2 \text{ se } waz \text{ estiver entre } [-2, 1) \}$

Figura 5.4: Probabilidades estimadas pelo SAS e pelo R, com intercepto constante (C) e não constante (NC), no tempo, dos tipos de desnutrição (1 e 2) no pior cenário para crianças do sexo masculino, quando mamam um número mediano de dias.



Por fim, pode-se concluir que as estimativas das probabilidades de desnutrição aumentam ao longo do tempo, no pior cenário. E que, após os 6 meses de vida, as crianças que mamaram, no pior cenário, têm as probabilidades de desnutrição bem próximas das probabilidades de desnutrição daquelas crianças que nunca mamaram no melhor cenário. Esta constatação mostra a extrema importância da lactação materna nos primeiros meses de vida especialmente para as crianças que se encontram em precária situação sócio-econômica. Portanto, dentro da política de incentivo ao aleitamento, as campanhas publicitárias devem estar direcionadas à população de mais baixa renda.

Capítulo 6

Conclusões

Os modelos marginais são uma ferramenta muito útil para realizar análises de dados longitudinais categorizados.

A estimação destes modelos pode ser realizada através da abordagem das equações de estimação generalizadas (GEE).

As vantagens de se trabalhar com GEE é não requerer a distribuição multivariada e a consistência da estimação mesmo com uma estrutura de correlação mal especificada.

A utilização das equações de estimação permitem obter estimativas consistentes dos parâmetros de regressão e suas variâncias sob condições regulares. Tais equações são obtidas sem a especificação completa da distribuição conjunta das observações dos sujeitos.

Todos os modelos GEE que assumem ortogonalidade para as equações de estimação para os parâmetros de regressão, assim como os parâmetros de associação, são chamados “GEE de ordem 1” ou “GEE1”. Uma generalização destes modelos são os chamados GEE2.

As equações de estimação (GEE2) fornecem inferências consistentes sobre os parâmetros do modelo marginal de baixa ordem e podem ser aplicada a clusters de tamanhos diferentes. Uma limitação é que inferências sobre as médias marginais são consistentes somente quando a especificação do modelo de associação é correta. Assim GEE2 é preferível quando a associação é o foco principal.

É conveniente utilizar o GEE1 quando o tamanho do cluster é grande. A consistência de $\hat{\beta}$ obtido a partir do GEE1 não depende da especificação correta do modelo de associação.

Se os dados não são binários e o foco da análise inclui interpretação da correlação, então o modelo GEE2 é preferível sobre um modelo GEE1.

As probabilidades estimadas utilizando os dois pacotes computacionais são bastantes próximas, apesar do pacote SAS utilizar o método GEE1, e o R utilizar o método GEE2. Em nosso estudo sobre estado nutricional de crianças as estimativas oferecidas pelo R resultaram sempre um pouco menores que as oferecidas pelo SAS.

Em relação o modelo ajustado utilizando o pacote SAS, podemos concluir que as crianças do sexo masculino tem probabilidades de serem desnutridas levemente maiores que as meninas. Já utilizando o pacote R, ocorre o contrário, as meninas tem as probabilidades de desnutrição levemente maiores em relação aos meninos. Essa pequena diferença se justifica pelo fato de que em todos os modelos ajustados, a variável sexo aparece não significante estatisticamente.

Em relação ao R, modelos com intercepto constante ou não, produz probabilidades esti-

madras similares, sendo o modelo com intercepto não constante propício para situações onde a modificação do intercepto ao longo do tempo tenha grande contribuição para a variável resposta.

Em forma geral podemos concluir que mesmo que a criança tenha uma “condição geral” desfavorável, se esta mamou um número mediano de vezes, as probabilidades de desnutrição são menores em todos os tempos, em relação a uma criança que nunca mamou destacando a importância da lactação materna.

O enfoque de modelos mistos é algumas vezes conhecido como modelagem *condicional* ou sujeito-específico: os coeficientes de regressão se aplicam para cada indivíduo mas não necessariamente para a população ao contrário do *marginal*, onde os coeficientes de regressão se aplicam para a população. Modelos *cluster*-específico usualmente representam o efeito do grupo (*cluster*) por um termo de efeito aleatório no modelo. Estes efeitos aleatórios podem ajudar a descrever a variabilidade extra que ocorre em várias situações, devido a fatores tais como heterogeneidade do sujeito ou covariáveis não observadas.

Os modelos marginais permitiram estimar as probabilidades de desnutrição ao longo do tempo de maneira relativamente simples, incorporando a correlação entre as observações.

Propostas de trabalhos futuros:

- Obtenção do critério modificado de Akaike, do pacote computacional R, para a seleção de um modelo.
- Obtenção do modelo ajustado pelo R, para diferentes estruturas de correlação, quando o intercepto é constante no tempo.
- Realização de análise de diagnóstico e de resíduos para estes tipos de modelos.
- Implementação de modelos mistos com efeitos aleatórios para o mesmo conjunto de dados.

Apêndice

Apêndice 1 - Programa SAS para obtencao do modelo marginal

```
/******  
* Leitura do banco de dados e convercao do arquivo do SPSS *  
* para SAS *  
*****/  
  
filename Z 'c:\userroot\alunos\Silvia Patricia\bancotrab.por';  
proc convert SPSS=Z;  
run;  
  
proc fsvview;  
run;  
  
/******  
* mamada - Quantidade de dias em aleitamento; *  
* numero - Registro da cianca (alfanumerico); *  
* caso3 - Registro da cianca (numerico); *  
* peso_cod - Peso da crianca (codificado); *  
* gela_cod - Existencia de geladeira; *  
* estmae_c - Anos de estudo da mae; *  
* t1, t2, t3, t4, t5, t6 e t7 - Variaveis dummies; *  
* tempo_id - Tempo em meses; *  
* sexo_cod - Sexo da crianca; *  
* waz_ord - Medida antropometrica WAZ (codificada); *  
* total_co - Total de moradores; *  
* mamad - Quantidade de dias em aleitamento (modificada); *  
*****/  
  
data banco1;  
set work.data1;  
mamad=mamada/100;  
keep numero count1 caso3 mamada peso_cod gela_cod estmae_c  
t1 t2 t3 t4 t5 t6 t7 tempo_id sexo_cod waz_ord total_co  
mamad;  
run;  
  
/******  
* Procedimento para obtencao do modelo marginal *  
*****/  
  
proc genmod data=banco1;  
class numero;  
output out=probs predicted=phat xbeta=x;  
model waz_ord= mamad*t1 mamad*t2 mamad*t3 mamad*t4 mamad*t5  
mamad*t6 mamad*t7 peso_cod gela_cod estmae_c  
sexo_cod total_co  
/dist = multinomial link=cumlogit;  
repeated subject=numero / type=indep corrw;  
run;
```

```

/*****
* Estimativas dos parametros de regressao para o modelo marginal *
*****/

```

The GENMOD Procedure
Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr	> Z
Intercept1	-4.1943	0.3177	-4.8170	-3.5717	-13.20	<.0001	
Intercept2	-2.2246	0.2521	-2.7187	-1.7306	-8.83	<.0001	
MAMADA*T1	-0.0260	0.0038	-0.0334	-0.0187	-6.91	<.0001	
MAMADA*T2	-0.0168	0.0024	-0.0216	-0.0120	-6.91	<.0001	
MAMADA*T3	-0.0078	0.0016	-0.0108	-0.0047	-5.00	<.0001	
MAMADA*T4	-0.0011	0.0010	-0.0030	0.0008	-1.11	0.2664	
MAMADA*T5	0.0001	0.0008	-0.0015	0.0017	0.16	0.8760	
MAMADA*T6	0.0004	0.0007	-0.0010	0.0018	0.54	0.5859	
MAMADA*T7	-0.0002	0.0007	-0.0015	0.0010	-0.37	0.7145	
PESO_COD	1.1590	0.1476	0.8698	1.4482	7.85	<.0001	
GELA_COD	0.5534	0.2003	0.1609	0.9459	2.76	0.0057	
ESTMAE_C	0.4242	0.2004	0.0314	0.8169	2.12	0.0343	
SEXO_COD	0.1919	0.1930	-0.1864	0.5703	0.99	0.3201	
TOTAL_CO	0.0924	0.1952	-0.2901	0.4749	0.47	0.6359	

```

/*****
* Procedimento para obtencao das probabilidades do modelo *
* marginal *
*****/

```

```

proc print data=probs;
run;

```

Apêndice 2 - Programa R para obtencao dos modelos marginais

```

/*****
* Procedimento para obtencao do modelo marginal *
* sob a estrutura de independencia *
*****/
mam2<-read.table("c:/userroot/alunos/Silvia Patricia/bancotrar2.dat",
header = T)
pesocod<-mam2$pesocod
estmaec<-mam2$estmaec
gelacod<-mam2$gelacod
sexocod<-mam2$sexocod
caso3<-mam2$caso3
tempoid<-mam2$tempoid
numero<-mam2$numero
mam2<-mam2[order(mam2$grupo, mam2$tempoid),]

```

```

totalco<-mam2$totalco
totalco[totalco==9999]<-NA
mam2$wazord<-factor(mam2$wazord)
wazord<-mam2$wazord
mamada<-mam2$mamada
t1<-mam2$t1
t2<-mam2$t2
t3<-mam2$t3
t4<-mam2$t4
t5<-mam2$t5
t6<-mam2$t6
t7<-mam2$t7
mamad1<- mamada*t1
mamad2<- mamada*t2
mamad3<- mamada*t3
mamad4<- mamada*t4
mamad5<- mamada*t5
mamad6<- mamada*t6
mamad7<- mamada*t7
fit<- ordgee(wazord~mamad1+mamad2+mamad3+mamad4+mamad5+mamad6+mamad7
+pesocod+gelacod+estmaec+sexocod+totalco, id=caso3, data=mam2,
mean.link="logit",corstr="independence",int.const=FALSE)
summary(fit)

/*****
* Estimativas dos parametros de regressao para o modelo marginal *
* sob a estrutura de covariancia de independencia com intercepto *
* nao constante *
*****/

Call:
ordgee(formula = wazord ~ mamad1 + mamad2 + mamad3 + mamad4 +
  mamad5 + mamad6 + mamad7 + pesocod + gelacod + estmaec +
  sexocod + totalco, id = caso3, data = mam2, mean.link = "logit",
  corstr = "independence", int.const = FALSE)

Mean Model:
Mean Link:          logit
Variance to Mean Relation: binomial

Coefficients:
      estimate   Naive.se  robust.se   z.robust    p.value
Inter:1:0  4.314599e+00  0.386448925  0.3612325727  11.94410274  0.000000e+00
Inter:1:1  1.981004e+00  0.283159937  0.2298651400   8.61811245  0.000000e+00
Inter:2:0  3.722152e+00  0.339468715  0.2868408520  12.97636659  0.000000e+00
Inter:2:1  2.005064e+00  0.280899479  0.2235677251   8.96848660  0.000000e+00
Inter:3:0  3.713410e+00  0.336018565  0.2784469793  13.33614846  0.000000e+00
Inter:3:1  2.005147e+00  0.278996293  0.2183431004   9.18346717  0.000000e+00
Inter:4:0  3.781696e+00  0.339558365  0.2799861608  13.50672449  0.000000e+00
Inter:4:1  2.118182e+00  0.281070249  0.2197342506   9.63974615  0.000000e+00
Inter:5:0  4.068817e+00  0.359701616  0.3001751867  13.55480916  0.000000e+00
Inter:5:1  1.976383e+00  0.277204127  0.2166326928   9.12319968  0.000000e+00
Inter:6:0  4.091434e+00  0.361684260  0.3144961736  13.00948814  0.000000e+00

```

```

Inter:6:1  1.954842e+00  0.276193407  0.2202093855  8.87719492  0.000000e+00
Inter:7:0  4.224486e+00  0.376103317  0.3430132179  12.31581158  0.000000e+00
Inter:7:1  1.866619e+00  0.274871482  0.2160380438  8.64023506  0.000000e+00
mamad1    2.702037e-02  0.007972801  0.0067854858  3.98208285  6.831396e-05
mamad2    1.747089e-02  0.005184854  0.0041868725  4.17277825  3.009078e-05
mamad3    1.007561e-02  0.003440516  0.0020980477  4.80237641  1.567936e-06
mamad4    2.636164e-03  0.001854752  0.0012070661  2.18394353  2.896640e-02
mamad5    1.643631e-05  0.001269482  0.0009302504  0.01766869  9.859032e-01
mamad6    4.506740e-05  0.001116209  0.0008303589  0.05427460  9.567164e-01
mamad7    5.643657e-04  0.001044683  0.0008759800  0.64426787  5.194017e-01
pesocod   -1.017523e+00  0.138784837  0.1055009113  -9.64468472  0.000000e+00
gelacod   -5.237514e-01  0.213095305  0.1900908706  -2.75526835  5.864402e-03
estmaec   -6.302058e-01  0.207148830  0.1623815722  -3.88101813  1.040201e-04
sexocod    2.415273e-02  0.197538090  0.1629196279  0.14824933  8.821460e-01
totalco    5.129670e-02  0.206974824  0.1638726098  0.31302789  7.542595e-01

```

Scale is fixed.

Correlation Model:

Correlation Structure: independence

Returned Error Value: 0

Number of clusters: 477 Maximum cluster size: 7

```

/*****
* Estimativas dos parametros de regressao para o modelo marginal *
* sob a estrutura de covariancia permutavel com intercepto      *
* nao constante                                                *
*****/

```

```

fit<- ordgee(wazord~mamad1+mamad2+mamad3+mamad4+mamad5+mamad6+mamad7
+pesocod+gelacod+estmaec+sexocod+totalco, id=caso3, data=mam2,
mean.link="logit",corstr="exchangeable",int.const=FALSE)
summary(fit)

```

Call:

```

ordgee(formula = wazord ~ mamad1 + mamad2 + mamad3 + mamad4 +
mamad5 + mamad6 + mamad7 + pesocod + gelacod + estmaec +
sexocod + totalco, id = caso3, data = mam2, mean.link = "logit",
corstr = "exchangeable", int.const = FALSE)

```

Mean Model:

Mean Link: logit
Variance to Mean Relation: binomial

Coefficients:

	estimate	naive.se	robust.se	z.robust	p.value
Inter:1:0	4.308797e+00	0.391742953	0.3625972788	11.883146856	0.000000e+00
Inter:1:1	1.967847e+00	0.289598981	0.2308328967	8.524984558	0.000000e+00
Inter:2:0	3.734374e+00	0.351070018	0.2631751535	14.189692236	0.000000e+00
Inter:2:1	1.981418e+00	0.351329713	0.2630010627	7.533878138	4.929390e-14
Inter:3:0	3.695808e+00	0.345492581	0.2523367419	14.646333181	0.000000e+00
Inter:3:1	2.028375e+00	0.345322911	0.2520090841	8.048818708	8.881784e-16

```

Inter:4:0  3.969987e+00  0.370986323  0.2690736452  14.754277375  0.000000e+00
Inter:4:1  1.612210e+00  0.370686526  0.2691121789   5.990847006  2.087509e-09
Inter:5:0  4.166415e+00  0.386154854  0.2744588885  15.180471367  0.000000e+00
Inter:5:1  1.733552e+00  0.385666414  0.2738472859   6.330359359  2.445908e-10
Inter:6:0  4.130521e+00  0.380843009  0.2797742692  14.763763981  0.000000e+00
Inter:6:1  1.835964e+00  0.380593990  0.2795513020   6.567537178  5.115419e-11
Inter:7:0  4.220210e+00  0.380737681  0.3415741380  12.355179627  0.000000e+00
Inter:7:1  1.865574e+00  0.281876518  0.2191332263   8.513421493  0.000000e+00
mamad1    2.611195e-02  0.008067166  0.0069108079   3.778422836  1.578247e-04
mamad2    1.714682e-02  0.005271072  0.0042043788   4.078324765  4.536137e-05
mamad3    9.887052e-03  0.003494566  0.0021090285   4.687965118  2.759350e-06
mamad4    2.659532e-03  0.001899635  0.0012425664   2.140353986  3.232617e-02
mamad5    6.059154e-06  0.001309771  0.0009504010   0.006375365  9.949132e-01
mamad6    1.379853e-04  0.001159771  0.0008545101   0.161478790  8.717163e-01
mamad7    6.293972e-04  0.001104826  0.0008921759   0.705463112  4.805221e-01
pesocod   -1.006450e+00  0.143984755  0.1100652636  -9.144122890  0.000000e+00
gelacod   -5.180105e-01  0.219392582  0.1936664744  -2.674755712  7.478372e-03
estmaec   -6.310254e-01  0.213449071  0.1656588489  -3.809185917  1.394251e-04
sexocod    2.772596e-02  0.203513840  0.1661833881   0.166839544  8.674963e-01
totalco    4.343633e-02  0.213463744  0.1681357737   0.258340803  7.961439e-01

```

Scale is fixed.

```

Correlation Model:
Correlation Structure:  exchangeable
Correlation Link:      log

```

```

Estimated Correlation Parameters:
      estimate robust.se    wald.z      p
alpha -0.03695405  0.1509220 -0.2448553  0.5967158

```

```

Returned Error Value: 1
Number of clusters: 477 Maximum cluster size: 7

```

Apêndice 3 - Programa R para obtencao dos modelos marginais

```

/*****
* Estimativas dos parametros de regressao para o modelo marginal *
* sob a estrutura de covariancia de independencia com intercepto *
* constante *
*****/

fit<- ordgee(wazord~mamad1+mamad2+mamad3+mamad4+mamad5+mamad6+mamad7+
pesocod+gelacod+estmaec+sexocod+totalco, id=caso3, data=mam2,
mean.link="logit", corstr="independence", int.const=TRUE)

summary(fit)

```

Call:
ordgee(formula = wazord ~ mamad1 + mamad2 + mamad3 + mamad4 +
mamad5 + mamad6 + mamad7 + pesocod + gelacod + estmaec +
sexocod + totalco, id = caso3, data = mam2, mean.link = "logit",
corstr = "independence", int.const = TRUE)

Mean Model:
Mean Link: logit
Variance to Mean Relation: binomial

Coefficients:

	estimate	san.se	wald	p
Inter:0	3.961550e+00	0.1951663435	4.120220e+02	0.000000e+00
Inter:1	1.986148e+00	0.1781466658	1.242990e+02	0.000000e+00
mamad1	2.737248e-02	0.0068081893	1.616461e+01	5.806900e-05
mamad2	1.780768e-02	0.0041495168	1.841703e+01	1.774652e-05
mamad3	1.024884e-02	0.0020744901	2.440771e+01	7.795627e-07
mamad4	2.725356e-03	0.0012058910	5.107759e+00	2.381907e-02
mamad5	-1.757346e-06	0.0009153174	3.686136e-06	9.984681e-01
mamad6	1.202552e-05	0.0007971471	2.275782e-04	9.879638e-01
mamad7	5.787011e-04	0.0008715806	4.408526e-01	5.067112e-01
pesocod	-1.010536e+00	0.1044608430	9.358286e+01	0.000000e+00
gelacod	-5.317980e-01	0.1894099265	7.882936e+00	4.990340e-03
estmaec	-6.364634e-01	0.1616566520	1.550100e+01	8.246161e-05
sexocod	2.588267e-02	0.1608783590	2.588349e-02	8.721852e-01
totalco	4.840126e-02	0.1628159440	8.837297e-02	7.662559e-01

Scale is fixed.

Correlation Model:
Correlation Structure: independence

Returned Error Value: 0
Number of clusters: 477 Maximum cluster size: 7

Referências bibliográficas

- [1] Agresti, A. (2002) *Categorical data analysis*. New York: Wiley
- [2] Agresti, A. & Natarajan, R. (2001) “Modeling Clustered Categorical Data: A Survey”. *International Statistical Review*, **69**, 3, 345-371.
- [3] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proceedings of the Second International Symposium on Information Theory*, B. N. Petrov and F. Csaki (eds), 267-281. Budapest: Akademiai Kiado.
- [4] Anderson, D. A. & Aitkin, M. (1985) “Variance component models with binary response: Interviewer variability”. *Journal of the royal Statistical Society. Series B.* **47**, 203-210.
- [5] Andrade, D. F. & Singer, J. M. (1986) “Análise de dados Longitudinais”. VII Simpósio Nacional de Probabilidade e Estatística. Campinas, SP.
- [6] Azzalini, A. (1984) “Estimations and hypothesis testing for collections of autoregressive time series”. *Biometrika*, **71**, 85-90.
- [7] Carey, V. J., Zeger, S. L., & Dieggle, P. (1993) “Modelling Multivariate Binary Data With Logistic Regressions”. *Biometrika*, **80**, 517-526.
- [8] Clayton, D. (1992) “Repeated Ordinal Measurements : A Generalized Estimating Equation Approach”, Technical Report. Medical Research Council Biostatistics Unit. Cambridge, U.K.
- [9] Demétrio, C. G. B. (2002) “Modelos Lineares Generalizados em Experimentação Agrônômica”. ESALQ/USP - Piracicaba, SP.
- [10] Diggle, P. J., Heagerty, P. J., Liang, K. Y. & Zeger, S. L. (2001) *Analysis of longitudinal data*. Statistical Science Series. Oxford University Press, Oxford.
- [11] Gange, S. J., Linton, K.L.P., Scott, A. J., DeMets, D. L., & Klein, R. (1993) “Analysis of Correlates Ordinal Measures With Ophthalmic Applications”, Technical Report 71, University of Wisconsin Dept. of Biostatistics.
- [12] Hardin, J. W. & Hilbe, J. M. (2002) *Generalized Estimating Equations*. Chapman & Hall/CRC.
- [13] Harville, D. A. (1977) “Maximum likelihood approaches to variance component estimation and to relate problems”. *Journal of the American Statistical Association* **72**, 320-338.
- [14] Heagerty, P. J. & Zeger S. L. (1996) “Marginal Regression Models for Clustered Ordinal Measurements”. *Journal of the American Statistical Association*, **91**, 1024-1036.
- [15] Horton, N. J. & Lipsitz S.R. (1999) “Review of software to fit generalized estimating equation regression models”. *Journal of the American Statistical Association*, **53**, 160-169.

- [16] Huang, G. H., Bandeen-Roche, Karen & Rubin, G. S. (2002) "Building marginal models for multiple ordinal measurements". *Appl. Statist.* **51**, parte 1, 37-57.
- [17] Kim, K. M., Williamson, J. M., & Lipsitz, S. L. (1993) "Regression Analysis for Bivariate Ordered Categorical Data From Ophthalmic Studies" (submitted for publication).
- [18] Koch, G. G., Landis, J. R., Freeman, J. L Freeman, D. H & Lehnen, R. G. (1977) "A general methodology for the analysis of experiments with repeated measurement of categorical data". *Biometrics*, *33*, 133-158.
- [19] Kullback, S. & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79-86.
- [20] Laird, N. M. & Ware, J. H. (1982) "Random-effects models for longitudinal data". *Biometrics*, **38**, 963-974.
- [21] Liang, K. Y. & Zeger, S. L. (1986) "Longitudinal data analysis using generalized linear models". *Biometrika*, **73**, 13-22.
- [22] McCullagh, P. (1980) "Regression models for ordinal data". *Journal of the Royal Statistical Society, Series B, Methodological*, **42**, 109-142.
- [23] McCullagh, P. & Nelder, J. A. (1983a) *Generalized Linear Models*. London: Chapman and Hall.
- [24] McCullagh, P. & Nelder, J. A. (1983b) "Quasi-likelihood functions". *Annals of Statistics*, **11**, 59-67.
- [25] McCullagh, P. & Nelder, J. A. (1989) *Generalized Linear Models*. London: Chapman and Hall.
- [26] Miller, M. E., Davis, C. S. & Landis, J. R. (1993) "The analysis of longitudinal polytomous data: Generalized estimating equations and connections with weighted least squares". *Biometrics*, **49**, 1033-1044.
- [27] Nelder, J. A. & Wedderburn, R. W. M. (1972) "Generalized Linear Models". *J. R. Statist. Soc. ,A* **135**, 370-84.
- [28] Ochi, Y. & Prentice, R. L. (1984) "Likelihood inference in correlated probit regression". *Biometrika* **71**, 531-543.
- [29] Pan, W. (2001) "Akaike's information criterion in generalized estimating equations". *Biometrics*, **57**, 120-125.
- [30] Paula, G. A. (2003) "Modelos de regressão com apoio computacional". São Paulo: IME/USP.
- [31] Paulino, C. D. M & Singer, J. M. (1997) "Análise de dados categorizados" (Versão preliminar). São Paulo: IME/USP.
- [32] Prentice, R. L. (1988) "Correlated binary regression with covariates specific to each binary observation". *Biometrics*, **44**, 1033-48.
- [33] Prentice, R. L., & Zhao, L. P. (1991) "Estimating Equations for Parameters in Means and Covariances of Multivariate Discrete and Continuous Responses". *Biometrics*, **47**,

825-839.

- [34] Rao, C. R. & Rao, M. B. (1998) "Matrix Algebra and Its Applications to Statistics and Econometrics". World Scientific.
- [35] Rubin, D. B. (1976) "Inference and missing data". *Biometrika*, **63**, 81-92.
- [36] Sawaya, A.L. (1996) "Desnutrio energtico-protica. In: SAWAYA, A. L. (org). Desnutrio urbana no Brasil em um perodo de transio". So Paulo, Cortez. Cap1, pp. 19-33.
- [37] Singer, J. L. & Andrade, D. F. (1986) "Análise de dados longitudinais". São Paulo: Associação Brasileira de estatística.
- [38] Stiratelli, R., Laird, N. & Ware, J. H. (1984) "Random-effects models for serial observations with binary response". *Biometrics* **40**, 961-971.
- [39] Wald, A. (1943) "Tests of statistical hypotheses concerning several parameters when the number off observations is large". *Trans. Amer. Math. Soc.*, **54**, 426-482.
- [40] Ware, J. H., Lipsitz, S. & Speizer, F. E (1988) "Issues in the analysis of repeated categorical outcomes". *Statistics in Medicine*, **7**, 95-107.
- [41] Wedderburn, R. W. M. (1974) "Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method". it *Biometrika*, **61**, 439-447.
- [42] Wickens, T. D. (1998) *Categorical data analysis*. Annu. Rev. Psychol. **48**, 537-558.
- [43] Yan, J. (2003) Equations estimating generalized for clustered ordinal responses. R documentation. <http://www.r-project.org>.
- [44] Zeger, S. L., M & Liang, K. Y. (1986) "Longitudinal data analysis for discrete and continuous outcomes". *Biometrics*, **42**, 121-130.
- [45] Ziegler, A., Kastner, C., Brunner, D. & Blettner, M. (2000) "Familial associations of lipid profiles: a generalized estimating equations approach". *Statistics in Medicine*, **19**, 3345-3357.