Luís Fernando Segala

Análise de Experimentos com Medidas Repetidas com Respostas Categóricas Incompletas

ANÁLISE DE EXPERIMENTOS COM MEDIDAS REPETIDAS COM RESPOSTAS CATEGÓRICAS INCOMPLETAS

Luís Fernando Segala

Dissertação de Mestrado
Pós-Graduação em Matemática Aplicada e Computacional
MAC-015

Análise de Experimentos com Medidas Repetidas com Respostas Categóricas Incompletas

Dissertação apresentada ao Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho" - UNESP - para obtenção do grau de Mestre em Ciências Matèmáticas, Área de Concentração: Matemática Aplicada e Computacional.

Luís Fernando Segala

sob orientação do Prof. Dr. Fernando Ferrari

São José do Rio Preto - S.P. - 20 de Janeiro de 1997

À minha filha Camila,
à minha esposa Deise e
aos meus pais Antonio e Luiza
dedico.

AGRADECIMENTOS

Li em algum lugar que sempre existem mais pessoas que merecem agradecimentos do que costumamos reconhecer, assim sei que muitas pessoas importantes para este trabalho não constam na minha pequena lista de agradecimentos. A elas meu muito obrigado.

Ao Prof. Dr. Fernando Ferrari pela orientação.

A todos os professores do DCCE e principalmente à Profª Adriana Barbosa Santos e à Profª Eliana Xavier Linhares de Andrade pela ajuda constante.

Aos amigos pós-graduandos e funcionários do DCCE.

À minha esposa Deise e à minha filha Camila pela compreensão e carinho nos momentos mais difíceis.

A todos os trabalhadores que patrocinaram este trabalho através da CAPES - PROEX.

A Deus.

RESUMO

Vários experimentos estatísticos são classificados como experimentos com Medidas Repetidas. Em tais estudos, as variáveis resposta são observadas na mesma unidade amostral em mais de uma condição de avaliação. Por vezes, algumas observações sobre as unidades amostrais estão faltando, produzindo o que chamamos de "observações incompletas". Neste trabalho apresentamos três métodos para análise de experimentos com observações incompletas com variável resposta categórica. Estes métodos são comparados através de exemplos numéricos com dados reais.

ABSTRACT

Several experimental situations can be classified as repeated measures experiments. In this cases, the response variable are observed in the same sample unit in more than two evaluation conditions. Sometimes some observation about the sample unit are missing, producing the called "missing data". In this work we presented three methods derived to analyse missing data experiments with categorical response variable. These methods are compared across numerical examples with real data.

Índice

Capítulo 1

Introdução

Um grande número de experimentos estatísticos são classificados como experimentos com *Medidas Repetidas* (M.R.). A característica que os define é que cada variável é observada em uma mesma unidade experimental, ou unidade amostral, duas ou mais vezes, ou seja, em duas ou mais *condições de avaliação*, como por exemplo, tempo, dosagem de certa droga, distância de uma fonte de radiação, etc... São exemplos de experimentos com Medidas Repetidas:

1. Os experimentos *"Split-plot"*, em que tratamentos são aleatoriamente aplicados a grupos contendo diferentes tipos de unidades experimentais aleatoriamente distribuídas. Tais estudos são muito usuais em agricultura. Por exemplo, quando se quer saber quais as diferenças entre fertilizantes aplicados a várias variedades de hortaliças, podemos plantar canteiros com as hortaliças aleatoriamente distribuidas e aplicar, também aleatoriamente, um certo tipo de fertilizante a cada canteiro.

2. Os *Estudos Longitudinais*, em que unidades experimentais pertencentes a uma mesma população são alocadas em subpopulações segundo covariáveis, sendo que uma ou mais variáveis denominadas *respostas* são observadas em diversas condições de avaliação. Como exemplo, podemos citar pessoas separadas em estratos segundo o sexo, das quais se observa o nível de colesterol em exames realizados três sábados consecutivos. Tais estudos são frequentes nas áreas médica, social, agrícola, entre outras.

3. Os estudos *"Change-over"*, onde as unidades experimentais, também divididas em subpopulações, são aleatoriamente submetidas a uma de diversas sequências de tratamentos. Por exemplo, num ensaio clínico, um certo grupo recebe os tratamentos A e B, nesta ordem, e outro grupo, B e A.

Existem outros tipos de experimentos com Medidas Repetidas, mas neste trabalho, daremos ênfase aos estudos do tipo 2 e 3. A extensão aos demais casos pode ser feita de maneira natural, bastando para isto, alguns acertos de notação.

As vantagens dos experimentos com Medidas Repetidas sobre os experimentos mais comumente encontrados, chamados de *"cross-sectional"*, nos quais cada unidade experimental é observada numa única ocasião, estão em que os experimentos M.R. fornecem melhores condições para o controle de fatores secundários que possam influenciar a resposta, permitem que se observe o comportamento individual em relação à condição de avaliação, além de que possibilitam o acompanhamento da resposta em diversas condições de avaliação. Temos ainda que notar que os M.R. requerem um número menor de unidades experimentais que os *"cross-sectional"*, tornando-se mais adequados em situações onde o custo na obtenção de unidades é alto, por exemplo, quando estamos estudando indivíduos com certa doença rara.

Quanto às respostas, podemos tê-las na forma de medidas (como por exemplo, peso de animais) ou contagens (por exemplo, número de filhos). Se as respostas são do primeiro tipo, dizemos que são respostas contínuas e temos diversas técnicas abundantemente descritas e trabalhadas na literatura estatística, como em Bryant et al. [4] e Andreoni [2]. Tais técnicas utilizam-se, na maioria das vezes, da "força" da distribuição normal multivariada. Já quando as respostas são do segundo tipo, chamamos de *respostas categóricas* ou *discretas*, podendo incluir aqui, respostas nominais, como raça, cor, etc... Ainda dentre as respostas categóricas, podemos ter respostas dicotômicas (como sucesso/fracasso, masculino/feminino) ou politômicas (com possibilidade de mais de dois níveis de resposta). Neste trabalho estamos interessados nas respostas categóricas politômicas. A razão de termos escolhido as variáveis categóricas está em que existem muitos resultados que já foram obtidos para as contínuas que ainda não estão adaptados às categóricas. Visamos assim com este trabalho, iniciar uma série de estudos nesta área que poderão render bons resultados futuramente.

Um fato comum em M.R. é a perda de dados. Às vezes esta perda ocorre por motivos que fogem ao domínio do pesquisador, mas às vezes se deve à própria maneira que o experimento foi planejado, no entanto esta discussão foge ao escopo deste trabalho. De qualquer forma, é bastante razoável que o pesquisador não queira perder informações sobre uma condição de avaliação. Neste trabalho são apresentados três procedimentos para análise de dados com estrutura incompleta, ou seja, com observações incompletas: o Método KO, descrito em Koch et al. [14], o Método WC, em Woolson e Clarke [28] e o Método LLH, em Lipsitz et al. [16]. Todos estes métodos buscam utilizar além das informações provindas das observações completas, também as informações das observações incompletas para prever quais seriam os parâmetros da distribuição (multinomial para cada subpopulação) esperada caso não ocorresse falta de dados.

Desejamos salientar que enfatizamos neste trabalho situações experimentais em que se pode escrever as respostas das unidades em função do conjunto de covariáveis que as dividiu em subpopulações, das condições de avaliação, bem como da interação entre subpopulações e condições de avaliação, utilizando modelos lineares univariados.

No Capítulo 2, temos a apresentação das estruturas dos dados que trabalhamos bem como as hipóteses que frequentemente são de interesse neste tipo de experimento, com uma pequena explanação sobre testes de tais hipóteses, utilizando modelos lineares univariados como uma maneira simples e eficaz de efetuar tais testes. Ainda neste capítulo, discorremos sobre os motivos que conduzem à falta de dados (aparecimento de observações incompletas). Para finalizar, apresentamos a metodologia de análise dos casos onde todas as respostas foram observadas, que utiliza-se do já bastante conhecido *Método dos Mínimos Quadrados Ponderados* (MMQP).

O Capítulo 3 destina-se a apresentar os métodos que visam incorporar as informações das unidades amostrais que apresentaram respostas incompletas na análise. Temos assim, definidos em tal capítulo, o Método KO, o Método WC e o Método LLH. O Método KO considera as observações incompletas em estratos diferentes das observações completas e aplica o MMQP para estimar os parâmetros da distribuição esperada se não houvesse falta de dados. O Método WC considera o fato da observação estar faltando como mais um nível

de resposta e utiliza uma outra função (razão de probabilidades) ao invés das probabilidades marginais no modelo linear. Já o Método LLH, utiliza-se do Método da Máxima Verossimilhança para obter as estimativas dos parâmetros da distribuição esperada, mas para isto, exige a utilização de processos iterativos. Sugerimos também neste capítulo, dois algorítmos que servem aos propósitos de maximização do Método LLH: o Algorítmo EM, cujos trabalhos de Dempster et al. [6] e Fuchs [9] descrevem sua aplicação a estimação de parâmetros de distribuições multinomiais e o Algorítmo de Newton Raphson (NR), bastante detalhado para o caso de respostas categóricas incompletas por Hocking e Oxspring [11]. A grande vantagem de se utilizar o NR ao invés do EM está em que a estrutura da matriz de variâncias e covariâncias de suas estimativas é facilmente obtida. Já o EM ainda não apresenta este aspecto bem desenvolvido. Louis [19] sugere a utilização da matriz de informação das respostas esperadas condicionada às respostas observadas. Nós preferimos adaptar a estrutura sugerida por Hocking e Oxspring [11], que originalmente serve para o algorítmo NR, também para o algorítmo EM.

A ilustração dos métodos do Capítulo 3 com dados reais está no Capítulo 4, onde apresentamos três problemas bastante conhecidos na literatura da área: um estudo sobre diferença entre os dois olhos quanto à acuidade em mulheres de 30 a 39 anos, um planejamento *"change-over"* com sequências de drogas (A, B e P) em duas condições de avaliação e um estudo sobre obesidade em crianças em idade escolar.

Por fim, comparamos os métodos quanto à dificuldade computacional, eficácia e outros aspectos no Capítulo 5.

Capítulo 2

Definições iniciais

2.1 Estrutura dos dados

Como foi mencionado anteriormente nosso intuito é enfatizar os estudos longitudinais e os estudos "change-over". Assim, a estrutura que vamos definir deverá ser ligeiramente modificada quando da necessidade da utilização das metodologias propostas aqui para outros casos de experimentos com Medidas Repetidas.

Sejam então n unidades experimentais, que são alocadas em S subpopulações segundo um conjunto de covariáveis. Para ilustrar tal situação, podemos pensar nas covariáveis como sendo sexo (masculino (M) e feminino (F)) e conhecimento anterior de certo ítem (com conhecimento (com) e sem conhecimento (sem)). Teríamos então quatro subpopulações compostas pelas referidas covariáveis:

- subpopulação 1: sexo masculino e com conhecimento (M, com)
- subpopulação 2: sexo masculino e sem conhecimento (M, sem)
- subpopulação 3: sexo feminino e com conhecimento (F, com)
- subpopulação 4: sexo feminino e sem conhecimento (F, sem)

Em estudos com respostas categóricas, cada unidade experimental pode assumir um entre K níveis de resposta em cada uma das D condições de avaliação a que fora submetido.

Deste modo, definimos $Y_{h_s sdk}$ como uma função indicadora que assume o valor 1 se o elemento $h_s\,(h_s = 1, \ldots, H_s)$ da subpopulação $s\,(s = 1, \ldots, S)$ assume o nível de resposta $k\,(k = 1, \ldots, K)$ na condição de avaliação $d\,(d = 1, \ldots, D)$ e zero caso contrário, sujeita à restrição:

$$\sum_{k=1}^{K} Y_{h_s sdk} = 1$$

A Tabela 2.1 ilustra a forma estrutural de um planejamento experimental com respostas categóricas de acordo com esta definição dos $Y_{h_s sdk}$.

Tabela 2.1: Estrutura dos dados para respostas categóricas com D condições de avaliação, K níveis e S subpopulações.

		condição						
	nível de	1			⋯	D		
subpopulação	*resposta*	1	⋯	K	⋯	1	⋯	K
	1	Y_{1111}	⋯	Y_{111K}	⋯	Y_{11D1}	⋯	Y_{11DK}
1	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	H_1	Y_{H_1111}	⋯	Y_{H_111K}	⋯	Y_{H_11D1}	⋯	Y_{H_11DK}
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮
	1	Y_{1S11}	⋯	Y_{1S1K}	⋯	Y_{1SD1}	⋯	Y_{1SDK}
S	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	H_S	Y_{H_SS11}	⋯	Y_{H_SS1K}	⋯	Y_{H_SSD1}	⋯	Y_{H_SSDK}

Podemos ainda tomar as sequências de respostas de cada unidade ao longo de todas as condições de avaliação, formando assim, $R = K^D$ perfis de resposta (também chamados de categorias de resposta). Conta-se então a frequência de cada categoria em cada subpopulação. O resultado pode ser visto na Tabela 2.2 e é chamado de tabela de contingência sendo n_{sr} o número de elementos da subpopulação s $(s = 1, ..., S)$ que assume a categoria r $(r = 1, ..., R)$.

A cada cela (s,r) da tabela de contingência, podemos associar uma proporção esperada que é a probabilidade de que uma unidade que foi alocada na subpopulação s tenha como resposta a categoria r . Tais probabilidades são também chamadas de probabilidades

Tabela 2.2: Tabela de Contingência - frequência de cada subpopulação por categoria de resposta.

	categoria				
subpopulação	**1**	**2**	...	**R**	**total**
1	n_{11}	n_{12}	...	n_{1R}	n_{1+}
⋮	⋮	⋮	...	⋮	⋮
S	n_{S1}	n_{S2}	...	n_{SR}	n_{S+}

Tabela 2.3: Probabilidades Conjuntas associadas às frequências de cada subpopulação por categoria de resposta.

	categoria			
subpopulação	**1**	**2**	...	**R**
1	π_{11}	π_{12}	...	π_{1R}
⋮	⋮	⋮	...	⋮
S	π_{S1}	π_{S2}	...	π_{SR}

conjuntas, já que são probabilidades de uma sequência de respostas. Na Tabela 2.3 são apresentadas as probabilidades conjuntas associadas às frequências que compõem a tabela de contingência, em que definimos π_{sr} como a probabilidade conjunta de que uma unidade amostral da subpopulação s assuma a categoria de resposta r.

2.2 Hipóteses de interesse

Em experimentos com Medidas Repetidas envolvendo respostas categóricas, comumente surgem algumas questões básicas:

- Existe diferença significativa entre as subpopulações?
- O efeito de condição de avaliação é significativo?

- O comportamento das subpopulações é influenciado pelas condições de avaliação, isto é, há interação entre subpopulações e condições de avaliação?

Conforme descrito em Koch et al. [15], estas questões podem ser descritas por diversas formulações que envolvem as probabilidades conjuntas referidas acima.

1. Utilizando simplesmente as probabilidades conjuntas:

   (a) Hipótese de igualdade entre subpopulações:

   $$H^{sc} : \pi_{1r} = \pi_{2r} = \ldots = \pi_{Sr} \qquad r = 1, ..., R$$

   ou seja, para cada condição de avaliação r, testamos a igualdade das probabilidades conjuntas de todas as subpopulações. A notação H^{sc} quer dizer que esta é a hipótese de igualdade entre subpopulações (s) baseada nas probabilidades conjuntas (c).

   (b) Hipótese de igualdade entre condições de avaliação:

   $$H^{cc} : \pi_{sr} = \pi_{sp} \qquad r = 1, ..., R \qquad s = 1, ..., S$$

   onde p é uma categoria composta por uma permutação dos níveis de resposta da categoria r.

2. Podemos tomar as probabilidades marginais e utilizá-las na construção das hipóteses.

   Probabilidade marginal é o nome que se dá à probabilidade de que uma unidade amostral da subpopulação s assuma o nível k na condição d. Na prática, temos que a probabilidade marginal, que representaremos por ϕ_{sdk} nada mais é do que a soma das probabilidades conjuntas das categorias r que apresentam o nível k na condição d:

   $$\phi_{sdk} = \sum \pi_{sr_{(dk)}} \quad \text{com} \quad \sum_{k=1}^{K} \phi_{sdk} = 1 \tag{2.1}$$

   Utilizando tais probabilidades temos as hipóteses nulas:

   (a) Hipótese de igualdade entre subpopulações:

   $$H^{sm} : \phi_{1dk} = \phi_{2dk} = \ldots = \phi_{Sdk} \qquad d = 1, ..., D \qquad k = 1, ..., K$$

(b) Hipótese de igualdade entre condições:

$$H^{cm} : \phi_{s1k} = \phi_{s2k} = \ldots = \phi_{sDk} \qquad s = 1, ..., S \qquad k = 1, ..., K$$

Supondo a não existência de interação entre subpopulações e condições, podemos construir um modelo linear, também chamado de modelo marginal, e testar seu ajuste. Se o ajuste for aceito, temos que não há efeito significativo de interação.

$$\phi_{sdk} = \mu_k + \alpha_{s\cdot k} + \beta_{\cdot dk} \qquad s = 1, ..., S \qquad d = 1, ..., D \qquad k = 1, ..., K \qquad (2.2)$$

restrito a:

$$\sum_{k=1}^{K} \mu_k = 1 \; ; \; \sum_{k=1}^{K} \alpha_{s\cdot k} = 0 \; ; \; \sum_{s=1}^{S} \alpha_{s\cdot k} = 0 \; ; \; \sum_{k=1}^{K} \beta_{\cdot dk} = 0 \; ; \; \sum_{d=1}^{D} \beta_{\cdot dk} = 0$$

com
$$\begin{cases} \mu_k : \text{média associada ao nível } k \\ \alpha_{s\cdot k} : \text{efeito devido à subpopulação } s \\ \beta_{\cdot dk} : \text{efeito devido à condição } d \end{cases}$$

Utilizando este modelo, podemos escrever as hipóteses (a) e (b) como:

(a) Hipótese de igualdade entre subpopulações:

$$H^{sm/am} : \alpha_{1\cdot k} = \alpha_{2\cdot k} = \ldots = \alpha_{S\cdot k} = 0 \qquad k = 1, ..., K$$

(b) Hipótese de igualdade entre condições:

$$H^{cm/am} : \beta_{\cdot 1k} = \beta_{\cdot 2k} = \ldots = \beta_{\cdot Dk} = 0 \qquad k = 1, ..., K$$

3. Quando tivermos uma escala ordinal nos níveis de resposta, ou seja, se eles são crescentes ou decrescentes, podemos tomar estruturas de probabilidades marginais acumuladas da forma:

$$\theta_{sdk} = \sum_{i=k+1}^{K} \phi_{sdi} \qquad s = 1, ..., S \qquad d = 1, ..., D \qquad k = 1, ..., K-1$$

que são as probabilidades de que a resposta seja estritamente maior que o nível k para a condição de avaliação d na subpopulação s.

Além disso, se pudermos caracterizar os acréscimos de cada nível em relação ao nível anterior por pesos não negativos $w_1, w_2, ..., w_{K-1}$, com ao menos um diferente de zero, podemos computar os níveis médios ponderados para cada condição de avaliação, dentro de cada subpopulação:

$$\eta_{sd} = \sum_{k=1}^{K-1} w_k \theta_{sdk} \qquad s = 1, ..., S \qquad d = 1, ..., D$$

Com esta estrutura, podemos escrever as hipóteses (a) e (b) como:

(a) Hipótese de igualdade entre subpopulações:

$$H^{snm} : \eta_{1d} = \eta_{2d} = \ldots = \eta_{Sd} \quad d = 1, ..., D$$

(b) Hipótese de igualdade entre condições:

$$H^{cnm} : \eta_{s1} = \eta_{s2} = \ldots = \eta_{sD} \quad s = 1, ..., S$$

A interação entre subpopulações e condições pode ser testada através da hipótese:

$$H^{inm} : \eta_{1d} - \eta_{1D} = \eta_{2d} - \eta_{2D} = \ldots = \eta_{Sd} - \eta_{SD} \qquad d = 1, ..., D-1$$

Se preferirmos, podemos construir um modelo linear com base nos níveis médios:

$$\eta_{sd} = \mu + \alpha_s + \beta_d \tag{2.3}$$

restrito a $\alpha_1 = 0$ e $\beta_1 = 0$,

com
μ : média da subpopulação 1 na condição 1
α_s : incremento devido à subpopulação s
β_d : incremento devido à condição de avaliação d.

Note que os parâmetros α_s e β_d são incrementos, justificando assim as restrições impostas, ou seja, a subpopulação 1 e a condição de avaliação 1 não têm incremento algum.

Se o ajuste do modelo for aceito, podemos concluir que não existe interação entre subpopulações e condições de avaliação. As hipóteses de igualdade entre subpopulações e entre condições podem ser escritas como:

(a) Hipótese de igualdade entre subpopulações:

$$H^{snm/am} : \alpha_2 = \alpha_3 = \ldots = \alpha_S = 0$$

(b) Hipótese de igualdade entre condições de avaliação:

$$H^{cnm/am} : \beta_2 = \beta_3 = \ldots = \beta_D = 0$$

2.2.1 Exemplo

Para melhor entendermos as hipóteses descritas acima, tomemos um exemplo genérico, com três níveis de resposta (A,B,C), duas condições de avaliação (1,2) e duas subpopulações (1,2), como descrito na Tabela 2.4. Neste caso temos $R = K^D = 3^2 = 9$ categorias de respostas formadas pelas combinações dos níveis A, B e C nas condições de avaliação 1 e 2, por exemplo, a categoria AA denota resposta A na primeira condição de avaliação e A na segunda, a categoria AB, A na primeira e B na segunda, etc. Para facilitar a compreensão dos índices, mudaremos a notação das probabilidades marginais para $\pi_{s(kk)}$ ao invés de π_{sr} como havíamos sugerido, assim, onde escreveríamos π_{s1}, escreveremos $\pi_{s(AA)}$ e assim por diante.

Tabela 2.4: Probabilidades conjuntas para 3 níveis de resposta, 2 condições e 2 subpopulações

	categoria								
subpopulação	AA	AB	AC	BA	BB	BC	CA	CB	CC
1	$\pi_{1(AA)}$	$\pi_{1(AB)}$	$\pi_{1(AC)}$	$\pi_{1(BA)}$	$\pi_{1(BB)}$	$\pi_{1(BC)}$	$\pi_{1(CA)}$	$\pi_{1(CB)}$	$\pi_{1(CC)}$
2	$\pi_{2(AA)}$	$\pi_{2(AB)}$	$\pi_{2(AC)}$	$\pi_{2(BA)}$	$\pi_{2(BB)}$	$\pi_{2(BC)}$	$\pi_{2(CA)}$	$\pi_{2(CB)}$	$\pi_{2(CC)}$

Vamos escrever as probabilidades marginais e os níveis médios na forma matricial. Assim:

$$\vec{\phi} = A_1 \, \vec{\pi} \tag{2.4}$$

com $\vec{\phi} = [\vec{\phi_1}, \vec{\phi_2}]'$, $\vec{\phi_s} = [\phi_{s1A}, \phi_{s1B}, \phi_{s1C}, \phi_{s2A}, \phi_{s2B}, \phi_{s2C}]'$, para s = 1,2 e $\vec{\pi} = [\vec{\pi_1}, \vec{\pi_2}]'$, $\vec{\pi_s} = [\pi_{s(AA)}, \pi_{s(AB)}, \pi_{s(AC)}, \pi_{s(BA)}, \pi_{s(BB)}, \pi_{s(BC)}, \pi_{s(CA)}, \pi_{s(CB)}, \pi_{s(CC)}]'$ para s = 1,2

$$A_1 = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \bigotimes I_2$$

onde $\bigotimes$ é o produto de Kronecker e I_2 é uma matriz identidade 2×2.

Assim, a probabilidade marginal do nível A na primeira condição de avaliação da primeira subpopulação, nada mais é do que a soma:

$$\phi_{11A} = \pi_{1(AA)} + \pi_{1(AB)} + \pi_{1(AC)}$$

Podemos obter ainda a seguinte relação:

$$\vec{\theta} = A_2 \, \vec{\phi} = A_2 A_1 \, \vec{\pi} \tag{2.5}$$

com $\vec{\theta} = [\vec{\theta_1} \, , \, \vec{\theta_2}]'$ e $\vec{\theta_s} = [\theta_{s1A} \, , \, \theta_{s1B} \, , \, \theta_{s2A} \, , \, \theta_{s2B}]'$ com s = 1,2

E a matriz de transformação A_2, dada por:

$$A_2 = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \bigotimes I_2$$

Os níveis médios ponderados são descritos pelo vetor $\vec{\eta}$ também em função de $\vec{\theta}$ e consequentemente de $\vec{\pi}$, pois, definindo a matriz de pesos A_3 por:

$$A_3 = \begin{bmatrix} w_A & w_B & 0 & 0 \\ 0 & 0 & w_A & w_B \end{bmatrix} \bigotimes I_2$$

Temos

$$\vec{\eta} = A_3 \, \vec{\theta} = A_3 A_2 A_1 \, \vec{\pi} \tag{2.6}$$

com $\vec{\eta} = [\eta_{11}\ ,\ \eta_{12}\ ,\ \eta_{21}\ ,\ \eta_{22}]'$

Desta forma, o nível médio da condição 1 na subpopulação 1, por exemplo, é dado por:

$$\eta_{11} = w_a(\pi_{1(BA)} + \pi_{1(BB)} + \pi_{1(BC)}) + (w_A + w_B)(\pi_{1(CA)} + \pi_{1(CB)} + \pi_{1(CC)})$$

As hipóteses (a) e (b), como definidas anteriormente, podem ser escritas para o nosso exemplo como:

1. Probabilidades conjuntas: $H^{sc}: \begin{cases} \pi_{1(AA)} = \pi_{2(AA)} \\ \pi_{1(AB)} = \pi_{2(AB)} \\ \vdots \\ \pi_{1(CC)} = \pi_{2(CC)} \end{cases}$ e $H^{cc}: \begin{cases} \pi_{1(AB)} = \pi_{1(BA)} \\ \pi_{1(AC)} = \pi_{1(CA)} \\ \pi_{1(BC)} = \pi_{1(CB)} \\ \pi_{2(AB)} = \pi_{2(BA)} \\ \pi_{2(AC)} = \pi_{2(CA)} \\ \pi_{2(BC)} = \pi_{2(CB)} \end{cases}$

2. Probabilidades marginais: $H^{sm}: \begin{cases} \phi_{11A} = \phi_{21A} \\ \phi_{11B} = \phi_{21B} \\ \phi_{11C} = \phi_{21C} \\ \phi_{12A} = \phi_{22A} \\ \phi_{12B} = \phi_{22B} \\ \phi_{12C} = \phi_{22C} \end{cases}$ e $H^{cm}: \begin{cases} \phi_{11A} = \phi_{12A} \\ \phi_{11B} = \phi_{12B} \\ \phi_{11C} = \phi_{12C} \\ \phi_{21A} = \phi_{22A} \\ \phi_{21B} = \phi_{22B} \\ \phi_{21C} = \phi_{22C} \end{cases}$

3. Níveis médios: $H^{snm}: \begin{cases} \eta_{11} = \eta_{21} \\ \eta_{12} = \eta_{22} \end{cases}$ e $H^{cnm}: \begin{cases} \eta_{11} = \eta_{12} \\ \eta_{21} = \eta_{22} \end{cases}$

E utilizando as matrizes A_1, A_2 e A_3 definidas acima, podemos reescrever os ítens 2 e 3 em função apenas das probabilidades conjuntas para podermos compará-las com o ítem 1.

As hipóteses 2, com base nas probabilidades marginais, podem ser escritas como:

$$H^{sm}: \begin{cases} \pi_{1(AA)} + \pi_{1(AB)} + \pi_{1(AC)} = \pi_{2(AA)} + \pi_{2(AB)} + \pi_{2(AC)} \\ \pi_{1(BA)} + \pi_{1(BB)} + \pi_{1(BC)} = \pi_{2(BA)} + \pi_{2(BB)} + \pi_{2(BC)} \\ \pi_{1(CA)} + \pi_{1(CB)} + \pi_{1(CC)} = \pi_{2(CA)} + \pi_{2(CB)} + \pi_{2(CC)} \\ \pi_{1(AA)} + \pi_{1(BA)} + \pi_{1(CA)} = \pi_{2(AA)} + \pi_{2(BA)} + \pi_{2(CA)} \\ \pi_{1(AB)} + \pi_{1(BB)} + \pi_{1(CB)} = \pi_{2(AB)} + \pi_{2(BB)} + \pi_{2(CB)} \\ \pi_{1(AC)} + \pi_{1(BC)} + \pi_{1(CC)} = \pi_{2(AC)} + \pi_{2(BC)} + \pi_{2(CC)} \end{cases}$$

e

$$H^{cm}: \begin{cases} \pi_{1(AB)} + \pi_{1(AC)} = \pi_{1(BA)} + \pi_{1(CA)} \\ \pi_{1(BA)} + \pi_{1(BC)} = \pi_{1(AB)} + \pi_{1(CB)} \\ \pi_{1(CA)} + \pi_{1(CB)} = \pi_{1(AC)} + \pi_{1(BC)} \\ \pi_{2(AB)} + \pi_{2(AC)} = \pi_{2(BA)} + \pi_{2(CA)} \\ \pi_{2(BA)} + \pi_{2(BC)} = \pi_{2(AB)} + \pi_{2(CB)} \\ \pi_{2(CA)} + \pi_{2(CB)} = \pi_{2(AC)} + \pi_{2(BC)} \end{cases}$$

Podemos dividir cada linha de H^{sm}, dos dois lados por três e temos que a hipótese de igualdade entre subpopulações, na verdade, sugere a igualdade para as duas subpopulações das médias das probabilidades conjuntas de cada nível em cada uma das condições de avaliação. Lembremos que a hipótese H^{sc}, que é a referente a esta mas com base nas probabilidades conjuntas, testa a igualdade das probabilidades conjuntas nas duas subpopulações, de cada categoria separadamente (e não em média), assim H^{sm} é mais ampla que H^{sc}.

A mesma conclusão pode-se tomar a respeito de H^{cm}, dividindo-se os dois lados por dois e temos assim que as hipóteses com base nas probabilidades marginais são mais adequadas que as estruturadas simplesmente com base nas probabilidades conjuntas.

Vamos agora escrever as hipóteses dos níveis médios em função das probabilidades conjuntas, começando por H^{snm}:[1]

$$H^{snm}: \begin{cases} w_A\pi_{1(B+)} + (w_A + w_B)\pi_{1(C+)} = w_A\pi_{2(B+)} + (w_A + w_B)\pi_{2(C+)} \\ w_A\pi_{1(+B)} + (w_A + w_B)\pi_{1(+C)} = w_A\pi_{2(+B)} + (w_A + w_B)\pi_{2(+C)} \end{cases}$$

Sem alterar as igualdades podemos escrever:

$$H^{snm}: \begin{cases} \dfrac{w_A\left(\frac{\pi_{1(B+)}}{3}\right) + (w_A+w_B)\left(\frac{\pi_{1(C+)}}{3}\right)}{2w_A+w_B} = \dfrac{w_A\left(\frac{\pi_{2(B+)}}{3}\right) + (w_A+w_B)\left(\frac{\pi_{2(C+)}}{3}\right)}{2w_A+w_B} \\ \dfrac{w_A\left(\frac{\pi_{1(+B)}}{3}\right) + (w_A+w_B)\left(\frac{\pi_{1(+C)}}{3}\right)}{2w_A+w_B} = \dfrac{w_A\left(\frac{\pi_{2(+B)}}{3}\right) + (w_A+w_B)\left(\frac{\pi_{2(+C)}}{3}\right)}{2w_A+w_B} \end{cases}$$

E temos que H^{snm} testa a igualdade para as duas subpopulações entre médias ponderadas das probabilidades conjuntas médias dos níveis B e C em cada condição de avaliação.

[1] O sinal + nos índices significa a soma das probabilidades conjuntas para todos os níveis na condição assinalada, assim, $\pi_{1(B+)}$ simboliza a soma $\pi_{1(BA)} + \pi_{1(BB)} + \pi_{1(BC)}$

Notemos que como os pesos w_A e w_B se referem a incrementos do nível A para o nível B e do nível B para o C, respectivamente, não temos na comparação as probabilidades conjuntas médias do nível A.

Fazendo o mesmo trabalho com H^{cnm} temos:

$$H^{cnm}: \begin{cases} w_A\pi_{1(AB)} + (w_A + w_B)\pi_{1(AC)} + w_B\pi_{1(BC)} = \\ \quad = w_A\pi_{1(BA)} + (w_A + w_B)\pi_{1(CA)} + w_B\pi_{1(CB)} \\ w_A\pi_{2(AB)} + (w_A + w_B)\pi_{2(AC)} + w_B\pi_{2(BC)} = \\ \quad = w_A\pi_{2(BA)} + (w_A + w_B)\pi_{2(CA)} + w_B\pi_{2(CB)} \end{cases}$$

Sem alterar as igualdades, podemos escrever:

$$H^{cnm}: \begin{cases} \frac{w_A\pi_{1(AB)}+(w_A+w_B)\pi_{1(AC)}+w_B\pi_{1(BC)}}{2(w_A+w_B)} = \frac{w_A\pi_{1(BA)}+(w_A+w_B)\pi_{1(CA)}+w_B\pi_{1(CB)}}{2(w_A+w_B)} \\ \\ \frac{w_A\pi_{2(AB)}+(w_A+w_B)\pi_{2(AC)}+w_B\pi_{2(BC)}}{2(w_A+w_B)} = \frac{w_A\pi_{2(BA)}+(w_A+w_B)\pi_{2(CA)}+w_B\pi_{2(CB)}}{2(w_A+w_B)} \end{cases}$$

que é a igualdade para as duas subpopulações da média ponderada das probabilidades conjuntas com os níveis na ordem natural de acontecimento (AB, AC e BC) e na ordem inversa (BA, CA e CB). Notemos que os "pesos" utilizados para o cálculo das médias são os acréscimos que separam os dois níveis que compõem a categoria, assim, para a categoria AC utilizamos os pesos w_A e w_B.

Também, estas hipóteses são mais amplas que aquelas baseadas em probabilidades marginais e por consequência, mais amplas que as baseadas simplesmente em probabilidades conjuntas, mas por exigirem que os níveis apresentem ordem, tornam-se por vezes inadequadas. Assim, no restante deste trabalho, faremos testes com base nas probabilidades marginais, salvo o Método WC, que utiliza outra função das probabilidades conjuntas. Ainda, daremos preferência à construção de modelos lineares (modelos marginais), pois tais constituem fácil ferramenta para previsão da resposta de novas unidades, que possam se enquadrar em qualquer uma das subpopulações definidas.

2.3 O problema dos dados faltantes

Como estamos trabalhando com experimentos com medidas repetidas, ou seja, estamos analisando alguma variável das unidades amostrais em mais de uma condição de avaliação, é comum encontrarmos situações onde algumas unidades não participem de todas as avaliações, proporcionando assim, respostas incompletas, ou seja, apresentam respostas para algumas condições de avaliação e não para outras. Um bom exemplo para tal situação é o caso de uma pesquisa clínica, onde após comparecer várias vezes a certo laboratório para coletar sangue, um paciente tem algum tipo de problema (que pode estar relacionado ou não à doença diagnosticada pelo exame de sangue) e falta a uma coleta, voltando a coletar sangue nas ocasiões seguintes. O pesquisador tem então em mãos informações incompletas para este paciente, mas às vezes, ele não pode simplesmente descartar tais informações e coletar novas, mas necessita de meios que possibilitem a utilização destas, mesmo que incompletas.

Além das situações que geram dados incompletos e que fogem ao controle do pesquisador como a sugerida acima, existem certos planejamentos que, ou por dificuldade na preservação das unidades amostrais em todas as condições de avaliação ou por outros fatores, prevêem a existência de observações incompletas. Discussões sobre o aparecimento de dados faltantes podem ser vistas em Little e Rubin [18], Little [17] e Woolson e Clarke [28], entre outros.

Vamos utilizar o símbolo "*" para denotar o fato da resposta estar faltando, assim, a categoria "SN*N" denotará por exemplo, resposta sim (S) para a primeira condição de avaliação, não (N) para a segunda, faltando (*) para a terceira e não (N) para a quarta.

Qualquer que seja o motivo do aparecimento dos dados faltantes, nos próximos capítulos nos dedicaremos a apresentar e discutir métodos de análise, através de modelos lineares que incluam as observações incompletas.

Quanto ao mecanismo que gera este tipo de dados, utilizaremos as especificações de Rubin [22, 23] e Little [17], que consideram os dados como:

1. **incompletos por razões completamente aleatórias** (MCAR, *"missing completely at random"*) quando a probabilidade da resposta estar faltando em determinada ocasião

não depende nem dos valores observados nas demais ocasiões, nem do valor a ser observado naquela condição de avaliação.

2. **incompletos por razões aleatórias** (MAR, *"missing at random"*) quando a probabilidade da resposta estar faltando depende dos valores observados, mas não dos valores não observados.

Notemos que o mecanismo MAR é muito mais realista que o mecanismo MCAR. Mais adiante, quando descrevermos os métodos que incorporam as observações incompletas na análise, notaremos a veracidade de tal afirmação e a importância do conhecimento do mecanismo de dados faltantes para a análise.

Necessário se faz ainda comentar que, quando da presença de respostas contínuas, existem testes para verificar qual mecanismo gerou os dados faltantes, como os citados em Simon e Simonoff [25], Little [17] ou Dixon [7].

2.4 Dois passos para a estimação

Considerando a estrutura dos dados como a da Tabela 2.2 e com possibilidade de presença de observações incompletas, vamos dirigir nosso trabalho para a obtenção de um modelo linear que escreva as probabilidades das respostas, ou alguma relação entre elas em função das covariáveis que definiram a subdivisão das unidades amostrais em subpopulações e das condições de avaliação.

Usualmente, utilizamo-nos de dois passos para isto, como podemos ver em Lipsitz et al. [16]:

1. **Passo I:** Estimar os parâmetros π_{sr}, da distribuição esperada caso não existisse falta de dados e sua matriz de variâncias e covariâncias, ou seja, obter $\hat{\pi}_{sr}$ e $V(\hat{\pi}_{sr})$.

2. **Passo II:** Escrever uma função dos parâmetros estimados, $\vec{F}_s$ $(\vec{\hat{\pi}}_s)$ e utilizá-la como resposta no modelo:

$$\vec{F}_s = \vec{F}_s\,(\vec{\hat{\pi}}_s) = X_s\,\vec{\beta} \tag{2.7}$$

As funções $\vec{F}_s$, podem ser tomadas quaisquer, segundo Forthofer e Koch [8], desde que tenham derivadas parciais contínuas até segunda ordem numa região contendo $\vec{\pi}_s$, mas, segundo estes autores, em um grande número de aplicações, apresentam-se como composições de funções:

1. *lineares:* $\vec{F}_s\ (\vec{\hat{\pi}}_s) = A\ \vec{\hat{\pi}}_s$, com A sendo uma matriz composta por 0's e 1's adequadamente posicionados.

2. *logarítmicas:* $\vec{F}_s\ (\vec{\hat{\pi}}_s) = \ln(\vec{\hat{\pi}}_s)$

3. *exponenciais:* $\vec{F}_s\ (\vec{\hat{\pi}}_s) = \exp(\vec{\hat{\pi}}_s)$

A matriz de variâncias e covariâncias de $\vec{F}_s$ é dada por:

$$V(\vec{F}_s) = H_s[V(\vec{\hat{\pi}}_s)]H_s' \tag{2.8}$$

com $H_s = \left[\frac{\partial \vec{F}_s(x)}{\partial x}\Big|_{x=\vec{\hat{\pi}}_s}\right]$ (matriz das derivadas parciais de $\vec{F}_s$ avaliadas nos pontos $\vec{\hat{\pi}}_s$).

Quando $\vec{F}_s$ é uma das funções dadas acima, temos estruturas de H_s já conhecidas, dadas por:

1. *lineares:* $H_s = A$

2. *logarítmicas:* $H_s = Diag(\vec{\hat{\pi}}_s)^{-1}$

3. *exponenciais:* $H_s = Diag(\exp(\vec{\hat{\pi}}_s))$

Para composições destas funções, basta compor adequadamente as H_s. Por exemplo, se temos $\vec{F}\ (\vec{\hat{\pi}}_s) = \exp[B\ln(A\ \vec{\hat{\pi}}_s)]$, com A e B matrizes de especificação, teremos H_s dada por:

$$H_s = Diag\left\{\exp[B\ln(A\ \vec{\hat{\pi}}_s)]\right\} BDiag[A\ \vec{\hat{\pi}}_s]^{-1}A$$

Um estimador BAN (*"best assymptotically normal"*) para $\vec{\beta}$, o vetor de parâmetros do modelo é dado, segundo Neymann [21] por:

$$\vec{b}=\vec{\hat{\beta}}=\left(\sum_{s=1}^{S} X_s'V(\vec{F}_s)^{-1}X_s\right)^{-1}\left(\sum_{s=1}^{S} X_s'V(\vec{F}_s)^{-1}\ \vec{F}_s\right) \tag{2.9}$$

Note que $\vec{b}$ é o valor de $\vec{\beta}$ que minimiza a forma quadrática:

$$\sum_{s=1}^{S} \left(\vec{F}_s - X_s\,\vec{\beta}\right)' V(\vec{F}_s)^{-1} \left(\vec{F}_s - X_s\,\vec{\beta}\right)$$

A matriz de variâncias e covariâncias de $\vec{b}$ é dada por:

$$V(\vec{b}) = V(\hat{\vec{\beta}}) = \left(\sum_{s=1}^{S} X_s' V(\vec{F}_s)^{-1} X_s\right)^{-1} \tag{2.10}$$

Após estimados os parâmetros do modelo, deseja-se testar o ajuste deste, ou seja, testar se os valores estimados para $\vec{F}_s$ estão "próximos" o suficiente dos valores esperados. Para tal, podemos utilizar a estatística de Wald [29]:

$$Q = \sum_{s=1}^{S} \left(\vec{F}_s - X_s\,\vec{b}\right)' V(\vec{F}_s)^{-1} \left(\vec{F}_s - X_s\,\vec{b}\right) \tag{2.11}$$

Q tem distribuição aproximadamente Qui-Quadrado com graus de liberdade dados pela soma do número de linhas das matrizes X_s menos o número de parâmetros a serem estimados. Os níveis de significância usuais são $\alpha = 0.05$ e $\alpha = 0.01$ e se $Q > \chi^2_{1-\alpha}$, rejeitamos a hipótese de ajuste do modelo, caso contrário, a aceitamos.

Também, se queremos testar a nulidade de algum parâmetro ou a igualdade entre alguns deles, podemos escrever tais hipóteses na forma $H_0 : C\,\vec{b} = 0$, com C, matriz de posto completo, composta por 0's, 1's e (-1)'s e utilizar a estatística:

$$Q_C = \sum_{s=1}^{S} \left(C\,\vec{b}\right)' \left[C\left(X_s' V(\vec{F}_s)^{-1} X_s\right)^{-1} C'\right]^{-1} \left(C\,\vec{b}\right) \tag{2.12}$$

cuja distribuição é aproximadamente Qui-Quadrado com graus de liberdade dados pelo número de linhas da matriz C.

Um problema que comumente ocorre com este método de estimação é que a matriz de variâncias e covariâncias de $\vec{F}_s$, $V(\vec{F}_s)$ nem sempre é não singular, o que inviabiliza o método, pois como vemos nas Equações (2.9), (2.10), (2.11) e (2.12), necessitamos de calcular sua inversa. A solução para tal está em substituir $V(\vec{F}_s)$ por uma outra $V^*(\vec{F}_s)$ que seja não singular. Se o número de condições de avaliação e o número de níveis de resposta são pequenos ($D \leq 3$ e $K \leq 3$), podemos obter $V^*(\vec{F}_s)$, segundo Koch et al. [15], aplicando as mesmas operações que geraram $V(\vec{F}_s)$ na matriz de contingência com alguns zeros trocados

por um número adequadamente pequeno. Em Koch et al. [15] está sugerido o valor a ser trocado como "0.5", o qual foi visto por nós, após simulações, como o melhor devido ao fato que, se aumentarmos este valor e utilizarmos valores próximos de 1, teremos um modelo com melhor ajuste, mas erros-padrões relativamente grandes, enquanto que se trocarmos os zeros por valores menores (perto do zero) teremos ajuste ruim, mas erros-padrões menores.

O método de estimação dos parâmetros, desenvolvido no Passo II é conhecido na literatura estatística como Método dos Mínimos Quadrados Ponderados (MMQP) e detalhes no seu desenvolvimento podem ser encontrados em Grizzle et al. [10] e em Koch et al. [15].

De posse do modelo e tendo a metodologia necessária para estimar seus parâmetros e testar seu ajuste, vamos revisar o método de trabalho para dados completos que pode ser encontrado em Koch et al. [14, 15], entre outros. Assim, na próxima seção, apresentaremos o método a ser utilizado no caso em que todas as unidades amostrais participaram da pesquisa em todas as condições de avaliação. Tal método pode ser aplicado também em situações em que existem poucas unidades amostrais com observações incompletas, bastando para isto descartá-las, mas, como veremos no capítulo final, isto implica em perdas de informação.

2.5 Método para Dados Completos

Se não temos presença de observações incompletas, podemos pensar a estrutura dos dados como aparecem na Tabela 2.2 e temos que cada uma das subpopulações terá distribuição multinomial com vetor de parâmetros $\vec{\pi_s} = [\pi_{s1}, \pi_{s2}, ..., \pi_{sR}]'$:

$$\varphi_s = \frac{n_{s+}!}{\prod_{r=1}^{R} n_{sr}!} \prod_{r=1}^{R} \pi_{sr}^{n_{sr}} \tag{2.13}$$

Assim à Tabela 2.3, temos associado um modelo produto multinomial dado por:

$$\varphi = \prod_{s=1}^{S} \left\{ \frac{n_{s+}!}{\prod_{r=1}^{R} n_{sr}!} \prod_{r=1}^{R} \pi_{sr}^{n_{sr}} \right\}$$

Um estimador consistente, de máxima verossimilhança, de cada probabilidade conjunta π_{sr}, é sabido ser a proporção observada da casela (s, r), assim, $\widehat{\pi}_{sr} = p_{sr} = n_{sr}/n_{s+}$.

Temos portanto que o vetor de proporções $\vec{p_s} = [p_{s1}, ..., p_{sR}]'$ estima o vetor $\vec{\pi_s}$ das probabilidades conjuntas. Além disto, sabemos que $\vec{p_s}$ para grandes amostras tem distri-

buição assintoticamente normal com $E[\vec{p_s}] = \vec{\pi_s}$ e matriz de variâncias e covariâncias dada pela inversa da matriz de informação observada que pode ser escrita como:

$$V(\vec{p_s}) = \frac{1}{n_{s+}} \left[Diag(\vec{p_s}) - \vec{p_s}\vec{p_s}' \right] \tag{2.14}$$

onde $Diag(\vec{p}_s)$ denota uma matriz com os elementos de $\vec{p}_s$ na diagonal principal e zero nas demais caselas.

Tomaremos então, as Probabilidades Marginais (2.1) por $\vec{F}$ $(\vec{\hat{\pi}}_s)$ e escreveremos o modelo marginal:

$$\vec{F}\,(\vec{\hat{\pi}}_s) = \vec{\phi}_s = X_s\,\vec{\beta}$$

Para a obtenção das probabilidades marginais pré-multiplicamos o vetor de proporções por uma matriz A, composta por 0's e 1's, posicionados para efetuar as somas necessárias e, desta maneira, temos uma estrutura linear para $\vec{F}$ $(\vec{\hat{\pi}}_s)$, levando-nos a usar $H_s = A$ no cálculo da matriz de variâncias e covariâncias.

Basta utilizarmos agora o Resultado (2.9) e temos o modelo linear estimado para nosso problema, com possibilidade de testarmos seu ajuste, nulidade de parâmetros, etc...

Alguns programas apresentam subrotinas que fazem automaticamente o trabalho de estimação apresentado acima. É o caso do SAS (SAS Institute), com o procedimento Cat-Mod. Ainda no SAS, se quisermos escrever as rotinas de cálculo matricial, podemos fazê-lo através do procedimento IML (Interative Matrix Language) e desta maneira observarmos passo a passo o andamento do algorítimo.

No Capítulo 4, apresentamos alguns exemplos numéricos deste método e o comparamos com os métodos que englobam os dados incompletos.

Apesar de existir a possibilidade de se descartarem as observações incompletas e se trabalhar simplesmente com as observações completas, isto nem sempre é viável na prática, pois a quantidade de observações incompletas pode ser muito grande, ocasionando perdas significativas na precisão das estimativas. Faz se então necessário a busca de métodos que não desprezem a presença de dados incompletos. Dedicaremos então o próximo capítulo à apresentação de três deles.

Capítulo 3

Métodos que utilizam as observações incompletas

Supondo que temos n unidades experimentais divididas em 2 subpopulações e vamos observar suas respostas sobre a aceitação de certa droga em 2 condições de avaliação. São portanto categorias de resposta: SS, SN, NS, NN, mas além disto, temos a existência de algumas unidades que foram observadas em apenas 1 condição, gerando as categorias incompletas: *S, *N, S* e N*. A Tabela 3.1 ilustra esta situação.

Tabela 3.1: Tabela de contingência para situação com observações incompletas. S denota resposta "sim", N "não" e *, resposta não observada.

	categoria								
subpopulação	**SS**	**SN**	**NS**	**NN**	***S**	***N**	**S***	**N***	**Total**
1	$n_{1(SS)}$	$n_{1(SN)}$	$n_{1(NS)}$	$n_{1(NN)}$	$n_{1(*S)}$	$n_{1(*N)}$	$n_{1(S*)}$	$n_{1(N*)}$	n_{1+}
2	$n_{2(SS)}$	$n_{2(SN)}$	$n_{2(NS)}$	$n_{2(NN)}$	$n_{2(*S)}$	$n_{2(*N)}$	$n_{2(S*)}$	$n_{2(N*)}$	n_{2+}

Com n_{sr} denotando a frequência da casela (s, r), $s = 1, 2$ e $r = (SS), (SN), \ldots, (N*)$

A pergunta é: O que devemos, ou podemos, fazer com as frequências das caselas que apresentam observações incompletas? Podemos descartá-las e utilizar o método para dados completos?

Talvez esta seja uma alternativa eficaz em alguns casos mas existem pesquisas nas quais

o número de observações incompletas é alto demais para que simplesmente as desconsideremos. É o caso, por exemplo, de quando se estuda algum aspecto em pacientes com certa doença rara. O número baixo de unidades amostrais não nos permite o descarte de uma delas sequer. Já em outros casos, o fato da unidade amostral "faltar" pode ser tão comum, que o descarte das observações incompletas resulta no "aproveitamento" de pouquíssimas unidades amostrais.

Apresentaremos então, três metodologias que utilizam as observações incompletas na estimação dos parâmetros do modelo e no Capítulo 5, faremos algumas comparações entre elas.

3.1 Método KO

Proposto em Koch et al. [14] consiste em dividir cada subpopulação em estratos de acordo com a estrutura de dados faltando em, estrato "completo", estrato dos que "faltaram na 1ª condição de avaliação", estrato dos que "faltaram na 2ª", e assim por diante. No exemplo genérico listado acima teríamos então três estratos em cada subpopulação: o "completo", o "faltando na 1ª" e o "faltando na 2ª". O número de estratos gerados pela divisão, em cada subpopulação, é dado por $T = 2^D - 1$, já que em cada condição existe a possibilidade de se ter dado faltando ou observado e não estamos interessados no estrato que contém as unidades que faltaram em todas as condições (justificando assim a subtração de 1).

Cada um dos estratos terá uma distribuição multinomial com vetor de parâmetros denotado por $\vec{P}_{s_t}$, cujas estimativas são as proporções observadas [1] e espera-se que exista uma relação destes com os parâmetros da distribuição, também multinomial, que explicaria as frequências caso nenhuma unidade amostral tivesse apresentado observações incompletas:

$$\vec{P}_{s_t} = Z_{s_t}\ \vec{\pi}_s$$

O Passo I para estimação dos parâmetros do modelo linear, consiste exatamente em se obter as estimativas dos parâmetros $\vec{\pi}_s$, da distribuição esperada, e isto pode ser feito como em (2.9), com Z_{s_t} no lugar de X_s e $V(\vec{P}_{s_t})$ ao invés de $V(\vec{F}_s)$ dado pela Estrutura (2.14).

[1] Usaremos a mesma notação $\vec{P}_{s_t}$ para as estimativas.

$$\vec{\hat{\pi}}_s = \left(\sum_{t=1}^{T} Z'_{s_t} V(\vec{P}_{s_t})^{-1} Z_{s_t} \right)^{-1} \left(\sum_{t=1}^{T} Z'_{s_t} V(\vec{P}_{s_t})^{-1} \vec{P}_{s_t} \right) \tag{3.1}$$

Também, a matriz de variâncias e covariâncias de $\vec{\hat{\pi}}_s$, como na Equação (2.10) é dada por:

$$V(\vec{\hat{\pi}}_s) = \left(\sum_{t=1}^{T} Z'_{s_t} V(\vec{P}_{s_t})^{-1} Z_{s_t} \right)^{-1} \tag{3.2}$$

O método segue tomando as probabilidades marginais como função resposta no modelo linear $\vec{F}_s = \vec{F}_s\ (\vec{\hat{\pi}}_s) = X_s\ \vec{\beta}$ e novamente, utilizando MMQP para a obtenção de $\vec{b} = \vec{\hat{\beta}}$, as estimativas do vetor de parâmetros do modelo como na Equação (2.9).

Para o exemplo genérico da Tabela 3.1 teríamos as relações:

$$\vec{P}_{s_1} = \begin{bmatrix} P_{s(SS)} \\ P_{s(SN)} \\ P_{s(NS)} \\ P_{s(NN)} \end{bmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{bmatrix} \pi_{s(SS)} \\ \pi_{s(SN)} \\ \pi_{s(NS)} \\ \pi_{s(NN)} \end{bmatrix} = I_{4\times 4}\ \vec{\pi}_s$$

$$\vec{P}_{s_2} = \begin{bmatrix} P_{s(*S)} \\ P_{s(*N)} \end{bmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \vec{\pi}_s$$

$$\vec{P}_{s_3} = \begin{bmatrix} P_{s(S*)} \\ P_{s(N*)} \end{bmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \vec{\pi}_s \qquad s = 1, 2$$

A Equação (3.1) nos dá $\vec{\hat{\pi}}_s$ e de posse de tais estimativas, podemos obter as probabilidades marginais como:

$$\vec{\phi}_s = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix} \vec{\hat{\pi}}_s = A\ \vec{\hat{\pi}}_s \qquad s = 1, 2$$

Seguiremos escrevendo o modelo

$$\vec{F}_s\ (\vec{\hat{\pi}}_s) = X_s\ \vec{\beta} \tag{3.3}$$

$$\vec{\beta} = [\mu_S\,,\ \alpha_{1\cdot S}\,,\ \gamma_{\cdot 1S}]' \text{ com } \begin{cases} \mu_S : \text{média geral do nível de resposta } S \\ \alpha_{1\cdot S} : \text{efeito da resposta Sim na subpopulação 1} \\ \gamma_{\cdot 1S} : \text{efeito da resposta Sim na condição 1} \end{cases}$$

Note que não precisamos calcular a média geral do nível N, devido à restrição:

$$\sum_{k=S,N} \mu_k = 1$$

Temos ainda outras restrições que diminuem o número de parâmetros a serem estimados. São elas:

$$\sum_{s=1}^{2} \alpha_{s\cdot k} = 0 \quad \sum_{k=S,N} \alpha_{s\cdot k} = 0 \quad \sum_{d=1}^{2} \gamma_{\cdot dk} = 0 \quad \sum_{k=S,N} \gamma_{\cdot dk} = 0$$

Escrevendo os parâmetros $\alpha_{s\cdot k}$ numa tabela 2×2, notamos que, devido às restrições acima, podemos calcular apenas um parâmetro ao invés de quatro:

Tabela 3.2: Efeitos de subpopulações × níveis de resposta.

	nível		
subpopulação	**S**	**N**	**soma**
1	$\alpha_{1\cdot S}$	$-\alpha_{1\cdot S}$	0
2	$-\alpha_{1\cdot S}$	$\alpha_{1\cdot S}$	0
soma	0	0	

Também os parâmetros $\gamma_{\cdot dk}$ podem ser reduzidos seguindo o mesmo pensamento:

Tabela 3.3: Efeitos de condições × níveis de resposta.

	nível		
condição	**S**	**N**	**soma**
1	$\gamma_{\cdot 1S}$	$-\gamma_{\cdot 1S}$	0
2	$-\gamma_{\cdot 1S}$	$\gamma_{\cdot 1S}$	0
soma	0	0	

As matrizes X_s devem ter a estrutura:

$$X_1 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \\ 1 & 1 & -1 \end{pmatrix} \qquad X_2 = \begin{pmatrix} 1 & -1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix}$$

E por fim, podemos aplicar o Resultado (2.9) e temos as estimativas dos parâmetros do Modelo (3.3) pelo Método KO.

Podemos reduzir mais ainda o número de cálculos efetuados, eliminando as probabilidades marginais ϕ_{sdN}, já que a soma sobre os níveis das probabilidades marginais de cada subpopulação em cada condição de avaliação, necessariamente dá um.

Uma particularidade do Método KO está em que, para a divisão das subpopulações supõe-se que a estrutura dos dados faltantes não dependa dos dados observados nem dos dados não observados. Tal estrutura já denotamos anteriormente por MCAR e tal suposição consiste na maior crítica deste método por não ser verificada em muitas situações reais. Mais adiante, quando formos apresentar o Método LLH, dedicaremos maior discussão a esta suposição. Problemas sobre a suposição de estrutura MCAR quando ela não é verdadeira são encontrados em Lipsitz et al. [16].

3.2 Método WC

Muito semelhante ao "Método da Razão", sugerido em Stanish et al. [26], o Método WC, proposto em Woolson e Clarke [28] considera o fato do dado estar faltando como mais um nível de resposta, assim, para nosso exemplo genérico, teremos os níveis S, N e *, e consequentemente as categorias SS, SN, NS, NN, *S, *N, S* e N*. O número de categorias geradas pelas combinações dos $K+1$ níveis de resposta é dado por $R = (K+1)^D - 1$ (o -1 deve-se ao fato de que não nos interessa a categoria que apresenta dado faltando em todas as condições).

Cada subpopulação terá distribuição multinomial com parâmetros $\vec{\pi_s} = [\pi_{s1}, \ldots, \pi_{sR}]'$ e, como no método para dados completos, as estimativas dos parâmetros, $\widehat{\pi}_{sr}$, são as próprias

proporções observadas p_{sr}, com matriz de variâncias e covariâncias como na Equação (2.14).

Toma-se então, por $\vec{F}_s\ (\vec{\hat{\pi}}_s)$, a razão da probabilidade de se observar nível k na condição de avaliação d, sem se importar se naquela categoria existe ou não dados faltando pela probabilidade da resposta ter sido observada na condição d sem se importar com o nível assumido. Tal razão denotaremos por $\lambda_{s_{dk}}$ e está expressa na relação abaixo:

$$F_{s_{dk}}(\hat{\pi}_{s_{dk}}) = \lambda_{s_{dk}} = \frac{\phi^{+}_{s_{dk}}}{\phi^{o}_{s_{d+}}} \quad k = 1, \ldots, K \quad d = 1, \ldots, D \quad s = 1, \ldots, S. \tag{3.4}$$

Com $\phi^{+}_{s_{dk}}$ denotando a soma das probabilidades conjuntas das categorias que apresentam nível k na condição d, sem se importar se existem dados faltando (*) nesta categoria e $\phi^{o}_{s_{d+}}$, a soma das probabilidades conjuntas das categorias que apresentam dado observado na condição d, sem se importar com o nível assumido.

Podemos escrever as razões $\lambda_{s_{dk}}$ em notação matricial aplicando-se logarítimos e exponencias, como $\vec{\lambda}_s = \exp[B \ln(A\ \vec{\hat{\pi}}_s)]$, com A e B, matrizes de especificação, compostas por 1's, 0's e (-1)'s, devidamente posicionados para efetuar as somas e subtrações exigidas.

A matriz de variâncias e covariâncias de $\vec{F}_s\ (\vec{\hat{\pi}}_s)$, pode ser obtida como composição das matrizes H_s para funções lineares, logarítmicas e exponenciais e teremos:

$$V(\vec{F}_s) = V[\vec{F}_s\ (\vec{\hat{\pi}}_s)] = H_s V(\vec{\hat{\pi}}_s) H_s' \tag{3.5}$$

$$\text{com } H_s = Diag\left\{\exp[B \ln(A\ \vec{\hat{\pi}}_s)]\right\} B\, Diag(A\ \vec{\hat{\pi}}_s)^{-1} A$$

Escreve-se então, o já conhecido modelo linear $\vec{F}_s = X_s\ \vec{\beta}$, que tem o vetor $\vec{b}$ como na Equação (2.9) como estimador do seu vetor de parâmetros.

No exemplo genérico listado na Tabela 3.1 temos:

$$\vec{\hat{\pi}}_s = \vec{p}_s = [p_{s(SS)}\ ,\ p_{s(SN)}\ ,\ \cdots\ ,\ p_{s(N*)}]'$$

$$\vec{\phi}_s = \begin{bmatrix} \phi^{+}_{s_{1S}} \\ \phi^{o}_{s_{1+}} \\ \phi^{+}_{s_{2S}} \\ \phi^{o}_{s_{2+}} \end{bmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \end{pmatrix} \vec{p}_s = A\ \vec{p}_s$$

$$\vec{\lambda_s} = \begin{pmatrix} \lambda_{s_1 S} \\ \lambda_{s_2 S} \end{pmatrix} = \exp\left\{ \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \ln\left[\vec{\phi_s}\right] \right\} = \exp\left\{ B \ln\left[A\,\vec{p_s}\right] \right\}$$

O modelo linear utilizado é o mesmo que o do Método KO, ou seja, $\vec{\lambda_s} = X_s\,\vec{\beta}$,

$$\vec{\beta} = [\mu_S\,,\ \alpha_{1\cdot S}\,,\ \gamma_{\cdot 1 S}]' \text{ com } \begin{cases} \mu_S : \text{média geral para o nível } S \\ \alpha_{1\cdot S} : \text{efeito da subpopulação 1 no nível } S \\ \gamma_{\cdot 1 S} : \text{efeito da condição 1 no nível } S \end{cases}$$

$$X_1 : \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \end{pmatrix} \quad \text{e} \quad X_2 = \begin{pmatrix} 1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix}$$

Note-se que para considerar o fato da observação estar faltando como mais um nível de resposta é necessário supor que a estrutura dos dados faltando não depende dos valores observados, ou seja, que os dados são MCAR.

3.3 Função Distribuição de Probabilidades

Como sugerido em Lipsitz et al. [16] podemos verificar o que acontece a nível de função distribuição de probabilidades quando adotamos um destes métodos e, consequentemente, uma estrutura de dados faltantes.

Tomemos então uma variável aleatória R tal que $R_{h_s sd} = 1$ se $Y_{h_s sd}$, a resposta para o indivíduo h_s da subpopulação s na condição d está faltando e $R_{h_s sd} = 0$ se $Y_{h_s sd}$ foi observado. Podemos então dividir o vetor de respostas de cada subpopulação em duas partes e reordenálas: a primeira parte com respostas observadas e consequentemente, com variável R valendo 0 e a segunda parte, com respostas não observadas ($R = 1$), ou seja, $\vec{Y_s} = (\vec{Y_s^o}, \vec{Y_s^f})$.

A função distribuição de probabilidades de R, será denotada por $f_{\vec{\psi_s}}(R_{h_s sd} | Y^o_{h_s sd})$ com $\vec{\psi_s}$ sendo os parâmetros da estrutura dos dados faltantes condicionada aos valores observados. No caso da estrutura ser MCAR, a função distribuição de R não depende dos valores observados, assim podemos escrever $f_{\vec{\psi_s}}(R_{h_s sd} | Y^o_{h_s sd}) = f_{\vec{\psi_s}}(R_{h_s sd})$.

Para a obtenção de $\vec{\hat{\pi}_s}$ necessitamos da distribuição dos valores observados $Y^o_{h_s sd}$ condicionada à estrutura dos dados faltantes, ou seja, de $f_{\vec{\pi_s}\vec{\psi_s}}(Y^o_{h_s sd} | R_{h_s sd})$. No caso de trabalharmos apenas com os dados completos, a distribuição era simples, não condicionada, pois

a variável R não tinha sentido (era constante igual a 0), assim, para estimarmos $\vec{\pi}_s$ bastava utilizar a função distribuição $f_{\vec{\pi}_s}(Y^o_{h_s sd}) = f_{\vec{\pi}_s}(Y_{h_s sd})$.

Utilizando definição de funções condicionadas, temos:

$$f_{\vec{\pi}_s \vec{\psi}_s}(Y^o_{h_s sd}|R_{h_s sd}) = \frac{f_{\vec{\pi}_s \vec{\psi}_s}(Y^o_{h_s sd}, R_{h_s sd})}{f_{\vec{\psi}_s}(R_{h_s sd})} = \frac{f_{\vec{\psi}_s}(R_{h_s sd}|Y^o_{h_s sd}) f_{\vec{\pi}_s}(Y^o_{h_s sd})}{f_{\vec{\psi}_s}(R_{h_s sd})} \tag{3.6}$$

A função distribuição dos valores observados $\vec{Y}^o_s$ pode ser escrita como a função distribuição de todos os valores $\vec{Y}_s$ somada sobre os valores que estão faltando $\vec{Y}^f_s$, ou seja:

$$f_{\vec{\pi}_s}(Y^o_{h_s sd}) = \sum_{Y^f_{h_s sd}} f_{\vec{\pi}_s}(Y_{h_s sd}) \tag{3.7}$$

Substituindo (3.7) em (3.6), temos:

$$f_{\vec{\pi}_s \vec{\psi}_s}(Y^o_{h_s sd}|R_{h_s sd}) = \underbrace{\frac{f_{\vec{\psi}_s}(R_{h_s sd}|Y^o_{h_s sd})}{f_{\vec{\psi}_s}(R_{h_s sd})}}_{I} \underbrace{\sum_{Y^f_{h_s sd}} f_{\vec{\pi}_s}(Y_{h_s sd})}_{II} \tag{3.8}$$

Se a estrutura dos dados faltantes é MCAR, a distribuição de R não depende dos valores $\vec{Y}^o_s$, então:

$$f_{\vec{\psi}_s}(R_{h_s sd}|Y^o_{h_s sd}) = f_{\vec{\psi}_s}(R_{h_s sd})$$

Assim, temos que o numerador e o denominador da Parte I da Equação (3.8) são iguais e a função distribuição dos valores observados dado R_s reduz-se à função distribuição de todos os valores, somada sobre os valores que estão faltando, ou seja:

$$f_{\vec{\pi}_s \vec{\psi}_s}(Y^o_{h_s sd}|R_{h_s sd}) = \sum_{Y^f_{h_s sd}} f_{\vec{\pi}_s}(Y_{n_s sd}) \tag{3.9}$$

Então, a obtenção dos estimadores $\vec{\hat{\pi}}_s$ pode ser feita por qualquer método, desde que este possua uma maneira de "somar" os valores não observados.

Os dois métodos que apresentamos até agora fazem exatamente isto. O Método KO, aplica uma vez MMQP para obter a "soma" dos valores não observados e após, aplica novamente MMQP no modelo linear. Já o Método WC, considera o fato da observação estar faltando como mais um nível de resposta e também aplica MMQP para a obtenção dos estimadores dos parâmetros do modelo linear.

3.4 Método LLH

Nosso objetivo então é obter um método que suponha a estrutura dos dados faltantes ser MAR, ao invés de MCAR. Este que vamos apresentar é sugerido em Lipsitz et al. [16] e consiste em aplicar o Método da Máxima Verossimilhança ao invés do MMQP para obtenção de $\hat{\vec{\pi}}_s$.

Segundo Rubin [22, 23], se os parâmetros $\vec{\pi}_s$ e $\vec{\psi}_s$ são distintos, no sentido que seu espaço paramétrico comum possa ser fatorado em duas partes disjuntas, uma contendo $\vec{\pi}_s$ e outra, contendo $\vec{\psi}_s$, podemos fatorar a função de verossimilhança de $\vec{\pi}_s$ e $\vec{\psi}_s$, também em duas partes, uma dependendo apenas de $\vec{\pi}_s$ e outra, apenas de $\vec{\psi}_s$.

Isto torna-se claro quando observamos que ao aplicarmos "ln" ao produtório sob h_s da Equação (3.8), teremos uma soma de duas partes, cada uma dependendo de apenas um tipo de parâmetros.

$$\begin{aligned} L(\vec{\pi}_s, \vec{\psi}_s) &= \prod_{h_s=1}^{H_s} \left\{ \frac{f_{\vec{\psi}_s}(R_{h_s sd} | Y^o_{h_s sd})}{f_{\vec{\psi}_s}(R_{h_s sd})} \sum_{Y^f_{h_s sd}} f_{\vec{\pi}_s}(Y_{h_s sd}) \right\} = \\ &= \prod_{h_s=1}^{H_s} \left[\frac{f_{\vec{\psi}_s}(R_{h_s sd} | Y^o_{h_s sd})}{f_{\vec{\psi}_s}(R_{h_s sd})} \right] \prod_{h_s=1}^{H_s} \left[\sum_{Y^f_{h_s sd}} f_{\vec{\pi}_s}(Y_{h_s sd}) \right] = \\ &= L(\vec{\psi}_s) L(\vec{\pi}_s) \\ \Rightarrow \quad \ln L(\vec{\pi}_s, \vec{\psi}_s) &= \ln L(\vec{\psi}_s) + \ln L(\vec{\pi}_s) \end{aligned}$$

Assim, para obtermos $\hat{\vec{\pi}}_s$, basta maximizarmos:

$$L(\vec{\pi}_s) = \prod_{h_s=1}^{H_s} \left[\sum_{Y^f_{h_s sd}} f_{\vec{\pi}_s}(Y_{h_s sd}) \right] \tag{3.10}$$

O mecanismo gerador de dados faltantes, tal qual descrito acima, com estrutura MAR e parâmetros $\vec{\psi}_s$ e $\vec{\pi}_s$ distintos é convenientemente chamado em Rubin [23] e Little [17] de "mecanismo ignorável para estimação por máxima verossimilhança".

Os estimadores de máxima verossimilhança para amostras grandes têm, a exemplo dos estimadores de mínimos quadrados, distribuição assintóticamente normal singular com média $\vec{\pi}_s$ (são não-viesados) e matriz de variâncias e covariâncias dada pela inversa da matriz de informação observada.

A maximização da Equação (3.10) exige a aplicação de algum algorítimo iterativo. Apresentaremos então dois algorítmos que podem ser utilizados neste sentido: O Algorítmo EM e o de Newton Raphson.

Após obtermos $\vec{\widehat{\pi}}_s$, podemos prosseguir como no Método KO, ou no Método para Dados Completos, tomando as probabilidades marginais $\vec{\phi}_s$ como $\vec{F}_s$ $(\vec{\widehat{\pi}}_s)$ e aplicando o MMQP a fim de obtermos os valores para $\vec{\widehat{\beta}}$ como em (2.9).

Notemos pois que o que diferencia o Método LLH dos demais é o uso do Método da Máxima Verossimilhança ao invés do Método dos Mínimos Quadrados Ponderados para a estimação de $\vec{\pi}_s$, que foi rotulado no Capítulo 2 como Passo I. O Passo II permanece o mesmo para todos os métodos, com exceção ao Método WC, que não utiliza as probabilidades marginais como função resposta no modelo, mas sim, uma razão de probabilidades.

3.4.1 O Algorítimo EM

A referência mais antiga que encontramos do algorítmo EM é de 1926, num artigo de Mckendrick que o utiliza na área médica. Hartley em 1958 o introduz como um procedimento para calcular estimadores de máxima verossimilhança dada uma amostra aleatória de tamanho n de uma variável discreta. Em 1973, Carters e Myers propuseram o EM para estimação de máxima verossimilhança de combinações lineares de funções de distribuição de probabilidades discretas. A utilização do EM em dados faltantes surge em 1974, com Brown estimando frequências esperadas sob modelo de independência em tabelas de contingência 2×2. Já Fienberg e Chen apresentam em 1976 sua utilização em casos especias de dados *"cross-classified"* com estrutura monótona[2] de dados faltantes. O nome "EM" se deve a Dempster et al. [6], quando em 1977 apresentam de maneira bastante didática diversos casos com observações incompletas onde podemos utilizar o algorítmo EM.

Se temos então a Função (3.10) para maximizar, o algorítmo EM sugere-nos dois passos:

[2] A estrutura de dados faltantes monótona se caracteriza, segundo Little e Rubin [18], pelo fato da primeira condição de avaliação ser mais observada que a segunda, que por sua vez, é mais observada que a terceira, e assim por diante. Desta forma as categorias do tipo "SSN*S" simplesmente não existem, pois após a ocorrência de um *, as outras condições, forçosamente apresentarão * como resposta.

- **Passo E:** *Esperança* - Dado $\widehat{\pi}_{sr}^{i}$, calcular a frequência esperada de cada casela da subpopulação s, ou seja, calcular:

$$m_{sr}^{i} = E[n_{sr}|\widehat{\pi}_{sr}^{i}] \tag{3.11}$$

- **Passo M:** *Maximização* - Tomando os valores de m_{sr}^{i} calculados no Passo E, maximizar a função de verossimilhança da distribuição esperada, ou seja, obter $\widehat{\pi}_{sr}^{i+1}$, a estimativa de máxima verossimilhança dos parâmetros da distribuição esperada, caso não existisse observações incompletas.

O algorítmo continua aumentando-se 1 ao contador i e repetindo-se os dois passos até que tenhamos algum critério de parada satisfeito.

O valor de $\vec{\widehat{\pi}}_{s}^{i}$ inicial, ou seja $\vec{\widehat{\pi}}_{s}^{1}$, deve ser escolhido o quanto mais próximo do valor esperado possível. A convergência do algorítmo EM depende da escolha deste valor inicial.

Em Fuchs [9] temos a sugestão da utilização das proporções observadas como valores iniciais. Um cuidado deve ser tomado ao aplicar tal sugestão: se houver alguma casela com frequência zero, teremos que a estimativa do parâmetro relativo a tal casela será forçosamente zero. Isto não é bom, no sentido de que podem existir observações incompletas que viriam a se classificar em tal casela caso completas e o Algorítmo EM deveria ser sensível a tais casos. Veja por exemplo: se temos a categoria SS com frequência zero e as categorias *S e S* com frequências não nulas. Algumas das observações de *S e S* deveriam ser alocadas a SS pelo Passo E, mas se utilizarmos as proporções observadas como valores iniciais, certamente isto não irá acontecer, pois todas as parcelas da esperança serão nulas. Uma sugestão de como resolver este impasse é substituir a proporção zero por $\frac{1}{n_{s+}}$. Poderemos entender melhor tal substituição mais adiante quando escrevermos os passos E e M para o exemplo genérico, ou no capítulo de exemplos numéricos, no exemplo que estuda obesidade em crianças.

Pode servir como critério de parada o teste $\max_r |m_{sr}^{i} - m_{sr}^{i+1}| < \epsilon$, conhecido como "erro absoluto", onde ϵ é um número adequadamente pequeno, ou ainda, $\max_r \frac{|m_{sr}^{i} - m_{sr}^{i+1}|}{|m_{sr}^{i}|} < \epsilon$, que é o "erro relativo".

Para facilitar a compreensão de como trabalha o algorítmo EM, vamos escrever os dois passos para o exemplo genérico descrito na Tabela 3.1. Tomando $p_{sr} = \frac{n_{sr}}{n_{s+}}$ como $\widehat{\pi}_{sr}^{1}$, temos

os dois passos definidos por:

- **Passo E:**

$$\begin{cases} m^i_{1(SS)} = n_{1(SS)} + \frac{\widehat{\pi}^i_{1(SS)}}{\widehat{\pi}^i_{1(SS)}+\widehat{\pi}^i_{1(SN)}} n_{1(S*)} + \frac{\widehat{\pi}^i_{1(SS)}}{\widehat{\pi}^i_{1(SS)}+\widehat{\pi}^i_{1(NS)}} n_{1(*S)} \\ \vdots \\ m^i_{1(NN)} = n_{1(NN)} + \frac{\widehat{\pi}^i_{1(NN)}}{\widehat{\pi}^i_{1(NN)}+\widehat{\pi}^i_{1(NS)}} n_{1(N*)} + \frac{\widehat{\pi}^i_{1(NN)}}{\widehat{\pi}^i_{1(NN)}+\widehat{\pi}^i_{1(SN)}} n_{1(*N)} \\ m^i_{2(SS)} = n_{2(SS)} + \frac{\widehat{\pi}^i_{2(SS)}}{\widehat{\pi}^i_{2(SS)}+\widehat{\pi}^i_{2(SN)}} n_{2(S*)} + \frac{\widehat{\pi}^i_{2(SS)}}{\widehat{\pi}^i_{2(SS)}+\widehat{\pi}^i_{2(NS)}} n_{2(*S)} \\ \vdots \\ m^i_{2(NN)} = n_{2(NN)} + \frac{\widehat{\pi}^i_{2(NN)}}{\widehat{\pi}^i_{2(NN)}+\widehat{\pi}^i_{2(NS)}} n_{2(N*)} + \frac{\widehat{\pi}^i_{2(NN)}}{\widehat{\pi}^i_{2(NN)}+\widehat{\pi}^i_{2(SN)}} n_{2(*N)} \end{cases}$$

- **Passo M:**

$$\begin{cases} \widehat{\pi}^{i+1}_{1(SS)} = \frac{m^i_{1(SS)}}{n_{1+}} \\ \vdots \\ \widehat{\pi}^{i+1}_{1(NN)} = \frac{m^i_{1(NN)}}{n_{1+}} \\ \qquad\qquad \text{com } i = 1, 2, \ldots \\ \widehat{\pi}^{i+1}_{2(SS)} = \frac{m^i_{2(SS)}}{n_{2+}} \\ \vdots \\ \widehat{\pi}^{i+1}_{2(NN)} = \frac{m^i_{2(NN)}}{n_{2+}} \end{cases}$$

O comentário da utilização das proporções observadas como valores iniciais e da substituição dos zeros por $\frac{1}{n_{s+}}$ pode ser melhor compreendida aqui. Veja que tal substituição faz com que tenhamos no Passo E:

$$m^i_{sr} = 0 + (a) + (b) \quad i = 1, 2, \ldots$$

com a e b valores possívelmente não nulos, ou seja, m^i_{sr} terá a contribuição das observações incompletas, mas a observação completa, que possui frequência nula, assim permanecerá até a última iteração.

Para implementação de tal algorítmo, pode-se utilizar qualquer software de computação algébrica, ou ainda, pode-se fazer um programa, relativamente simples em qualquer linguagem de programação. Nós utilizamos o "software" Mathematica da Wolfram Research.

O desenvolvimento do Algorítmo EM para distribuições multinomiais, como é o nosso caso, pode ser encontrado de maneira bem clara em Fuchs [9].

3.4.2 O Algorítmo de Newton-Raphson (NR)

O Algorítmo de Newton-Raphson (NR), como relatado em Bard [3], é um processo iterativo para se encontrar pontos de máximo de funções e consiste em aproximar a função, em nosso caso $L(\vec{\pi_s})$, nas proximidades de pontos $\hat{\pi}_{sd}$, por sua expansão em Séries de Taylor de ordem 2 e maximizar a expansão, ao invés da função propriamente dita.

Temos então:

$$L(\vec{\pi_s}) = L(\vec{\hat{\pi}_s}) + D(\vec{\hat{\pi}_s})'(\vec{\pi_s} - \vec{\hat{\pi}_s}) + \frac{1}{2}(\vec{\pi_s} - \vec{\hat{\pi}_s})' H(\vec{\hat{\pi}_s})(\vec{\pi_s} - \vec{\hat{\pi}_s}) \tag{3.12}$$

Com $D(\vec{\hat{\pi}_s})$, sendo o vetor de derivadas parcias de primeira ordem e $H(\vec{\hat{\pi}_s})$, a matriz Hessiana de $L(\vec{\pi_s})$, avaliados no vetor $\vec{\hat{\pi}_s}$, ou seja:

$$D(\vec{\hat{\pi}_s}) = \begin{bmatrix} \left.\frac{\partial L(\vec{\pi_s})}{\partial \pi_{s1}}\right|_{\pi_{s1}=\hat{\pi}_{s1}} \\ \vdots \\ \left.\frac{\partial L(\vec{\pi_s})}{\partial \pi_{sD}}\right|_{\pi_{sD}=\hat{\pi}_{sD}} \end{bmatrix}$$

e

$$H(\vec{\hat{\pi}_s}) = \begin{bmatrix} \left.\frac{\partial^2 L(\vec{\pi_s})}{\partial \pi_{s1}^2}\right|_{\pi_{s1}=\hat{\pi}_{s1}} & \left.\frac{\partial^2 L(\vec{\pi_s})}{\partial \pi_{s1}\partial \pi_{s2}}\right|_{\pi_{s1}=\hat{\pi}_{s1},\pi_{s2}=\hat{\pi}_{s2}} & \cdots & \left.\frac{\partial^2 L(\vec{\pi_s})}{\partial \pi_{s1}\partial \pi_{sD}}\right|_{\pi_{s1}=\hat{\pi}_{s1},\pi_{sD}=\hat{\pi}_{sD}} \\ \left.\frac{\partial^2 L(\vec{\pi_s})}{\partial \pi_{s2}\partial \pi_{s1}}\right|_{\pi_{s1}=\hat{\pi}_{s1},\pi_{s2}=\hat{\pi}_{s2}} & \left.\frac{\partial^2 L(\vec{\pi_s})}{\partial \pi_{s2}^2}\right|_{\pi_{s2}=\hat{\pi}_{s2}} & \cdots & \left.\frac{\partial^2 L(\vec{\pi_s})}{\partial \pi_{s2}\partial \pi_{sD}}\right|_{\pi_{s2}=\hat{\pi}_{s2},\pi_{sD}=\hat{\pi}_{sD}} \\ \vdots & \vdots & \cdots & \vdots \\ \left.\frac{\partial^2 L(\vec{\pi_s})}{\partial \pi_{sD}\partial \pi_{s1}}\right|_{\pi_{s1}=\hat{\pi}_{s1},\pi_{sD}=\hat{\pi}_{sD}} & \left.\frac{\partial^2 L(\vec{\pi_s})}{\partial \pi_{sD}\partial \pi_{s2}}\right|_{\pi_{s2}=\hat{\pi}_{s2},\pi_{sD}=\hat{\pi}_{sD}} & \cdots & \left.\frac{\partial^2 L(\vec{\pi_s})}{\partial \pi_{sD}^2}\right|_{\pi_{sD}=\hat{\pi}_{sD}} \end{bmatrix}$$

Derivando a Equação (3.12) em função de $\vec{\pi_s}$ temos:

$$D(\vec{\pi_s}) = D(\vec{\hat{\pi}_s}) + H(\vec{\hat{\pi}_s})(\vec{\pi_s} - \vec{\hat{\pi}_s}) \tag{3.13}$$

Por fim, igualando $D(\vec{\pi_s})$ a zero, temos:

$$\vec{\pi_s} = \vec{\hat{\pi}_s} - H(\vec{\hat{\pi}_s})^{-1} D(\vec{\hat{\pi}_s}) \tag{3.14}$$

Basta agora considerarmos $\vec{\hat{\pi}_s} = \vec{\hat{\pi}_s}^{\,i}$, como sendo o passo i e $\vec{\pi_s} = \vec{\hat{\pi}_s}^{\,i+1}$, o passo $i+1$ do algorítmo iterativo de Newton-Raphson, que fica então definido por:

$$\vec{\hat{\pi}}_s^{\,i+1} = \vec{\hat{\pi}}_s^{\,i} - H(\vec{\hat{\pi}}_s^{\,i})^{-1} D(\vec{\hat{\pi}}_s^{\,i}) \qquad i = 1, 2, \ldots \tag{3.15}$$

Assim como no algorítmo EM, o critério de parada pode ser $\max_r |m_{sr}^i - m_{sr}^{i+1}| < \epsilon$, o erro absoluto, onde ϵ é um número adequadamente pequeno, ou ainda $\max_r \frac{|m_{sr}^i - m_{sr}^{i+1}|}{|m_{sr}^i|} < \epsilon$, o erro relativo.

A rapidez na convergência depende dos valores iniciais $\vec{\hat{\pi}}_s^{\,1}$, justificando a procura por valores presumivelmente próximos dos valores esperados.

Uma modificação por vezes vantajosa neste algorítmo consiste em substituir $H(\vec{\hat{\pi}}_s^{\,i})$, a matriz Hessiana da função de verossimilhança $L(\vec{\pi}_s)$ calculada nos pontos $\vec{\hat{\pi}}_s^{\,i}$, que em algumas situações se torna computacionalmente difícil de se obter pela matriz de informação de Fisher com o sinal trocado, $-I(\vec{\hat{\pi}}_s^{\,i}) = E[H(\vec{\hat{\pi}}_s^{\,i})]$. Esta modificação é conhecida como "Scoring" de Fisher (veja Andreoni [2] ou Jennrich e Schluchter [12]). Temos assim, o algorítmo iterativo determinado pela equação:

$$\vec{\hat{\pi}}_s^{\,i+1} = \vec{\hat{\pi}}_s^{\,i} + I(\vec{\hat{\pi}}_s^{\,i})^{-1} D(\vec{\hat{\pi}}_s^{\,i}) \qquad i = 1, 2, \ldots \tag{3.16}$$

Para distribuições multinomiais em presença de dados faltantes, temos algumas sugestões de aplicação do NR que reduzem as dificuldades computacionais, descritos em Hocking e Oxspring [11] .

Este artigo sugere que dividamos as subpopulações como no Método KO em T estratos: estrato "completo", estrato "faltando na 1ª condição", estrato "faltando na 2ª", e assim por diante, sendo que cada extrato tem distribuição multinomial com parâmetros que são combinações lineares dos parâmetros da distribuição esperada caso não existisse falta de dados. Tal relação pode ser escrita matricialmente como:

$$\vec{\pi}_{s_t} = C_t \, \vec{\pi}_s \tag{3.17}$$

As matrizes C_t são matrizes compostas por 0's e 1's, com a restrição de que exista apenas um 1 em cada coluna e que não exista linhas contendo apenas 0's. A matriz C_1, dos estratos completos é convenientemente uma matriz Identidade.

Ainda, devido ao fato de que $\sum_{r=1}^{R} \pi_{sr} = 1$, podemos escrever:

$$\pi_{s_tR} = 1 - \sum_{r=1}^{R-1} \pi_{s_tr} = 1 - \vec{J}_t' \vec{P}_{s_t} \tag{3.18}$$

Com $\vec{J}_t$ sendo um vetor com todos os elementos iguais a 1 e $\vec{P}_{s_t} = (\pi_{s_t1}, \ldots, \pi_{s_tR-1})'$

A Relação (3.18) nos diz que podemos estimar apenas $R-1$ parâmetros em cada extrato, já que o último sai como consequência. Assim, trabalharemos para obter $\hat{\vec{P}}_{s_t}$.

Também podemos escrever a frequência da casela (s, R) em relação às $R-1$ anteriores como:

$$n_{s_tR} = n_{s_t+} - \vec{J}_t' \vec{N}_{s_t} \tag{3.19}$$

com $\vec{J}_t$ como acima e $\vec{N}_{s_t} = (n_{s_t1}, \ldots, n_{s_tR-1})'$.

As derivadas parciais de primeira ordem da função de verossimilhança em relação a $\vec{\pi}_{s_t}$ são dadas na forma matricial por:

$$D_{\vec{\pi}_{s_t}}(L(\vec{\pi}_{s_t})) = \frac{\partial \ln L(\vec{\pi}_{s_t})}{\partial \vec{P}_{s_t}} = U_{s_t}^{-1}(\vec{N}_{s_t} - n_{s_t+} \vec{P}_{s_t}) \tag{3.20}$$

com $U_{s_t} = Diag(\vec{P}_{s_t}) - \vec{P}_{s_t}\vec{P}_{s_t}'$ e $U_{s_t}^{-1} = \left[Diag(\vec{P}_{s_t})\right]^{-1} + \frac{1}{\pi_{s_tR}} \vec{J}_t\vec{J}_t'$

As matrizes de segundas derivadas de $L(\vec{\pi}_{s_t})$, ou seja, as matrizes Hessianas, neste contexto podem ser escritas como:

$$H_{\vec{\pi}_{s_t}}(L(\vec{\pi}_{s_t})) = -Diag\left(\frac{\vec{N}_{s_t}}{\vec{P}_{s_tr}^{\,2}}\right) - \frac{n_{s_tR}}{\pi_{s_tR}^2} \vec{J}_t\vec{J}_t' \tag{3.21}$$

A esperança negativa da Equação (3.21), ou seja, a matriz informação de Fisher é:

$$I_{s_t} = n_{s_t+} U_{s_t}^{-1} \tag{3.22}$$

E invertendo I_{s_t}, temos a matriz de variâncias e covariâncias para grandes amostras dada por:

$$V_{s_t} = \frac{1}{n_{s_t+}} U_{s_t} \tag{3.23}$$

Podemos escrever os Vetores de Primeiras Derivadas (3.20) e a Matrizes Hessianas (3.21) em função dos parâmetros $\vec{\pi}_s$, bastando para isto utilizarmos a Igualdade (3.17):

$$D_{\vec{\pi}_s}(L(\vec{\pi}_{s_t})) = \frac{\partial \ln L(\vec{\pi}_{s_t})}{\partial \vec{\pi}_s} = C_t' \frac{\partial \ln L(\vec{\pi}_{s_t})}{\partial \vec{\pi}_{s_t}} \tag{3.24}$$

$$H_{\vec{\pi}_s}(L(\vec{\pi}_{s_t})) = \frac{\partial^2 \ln L(\vec{\pi}_{s_t})}{\partial \vec{\pi}_s^{\,2}} = C_t' \frac{\partial^2 \ln L(\vec{\pi}_{s_t})}{\partial \vec{\pi}_{s_t}^{\,2}} C_t = C_t' H_{\vec{\pi}_{s_t}}(L(\vec{\pi}_{s_t}))C_t \tag{3.25}$$

Também é de nosso interesse, devido á estrutura da matriz de variâncias e covariâncias geral que apresentaremos mais adiante, escrever a matriz de variâncias e covariâncias de cada estrato em relação aos parâmetros π_s:

$$V_{s_t} = \left(\frac{1}{n_{s_t}}\right)\left(C_t U_s C_t'\right)$$

A matriz U_s tem estrutura como definida em (3.20), com os parâmetros da distribuição esperada.

A função de verossimilhança conjunta de todos os estratos de cada subpopulação, e que deve ser a função de verossimilhança da distribuição esperada, pode ser escrita como o produto das verossimilhanças dos estratos, ou seja:

$$L(\vec{\pi}_s) = \prod_{t=1}^{T} L(\vec{\pi}_{s_t}) \tag{3.26}$$

A Equação (3.15) do algorítmo de Newton-Raphson utiliza os vetores de primeiras derivadas e a matrizes hessianas das Funções de Verossimilhança (3.26) em relação aos parâmetros $\vec{\pi}_s$ das distribuições esperadas das subpopulações,

Utilizando as equações (3.24), (3.25) e (3.26), chegamos às relações:

$$D_{\vec{\pi}_s}(L(\vec{\pi}_s)) = \sum_{t=1}^{T} D_{\vec{\pi}_s}(L(\vec{\pi}_{s_t})) = \sum_{t=1}^{T} C_t' n_{s_t} + U_{s_t}^{-1} \left(\frac{\vec{N}_{s_t}}{n_{s_t R}} - \vec{P}_{s_t}\right) \tag{3.27}$$

$$H_{\vec{\pi}_s}(L(\vec{\pi}_s)) = \sum_{t=1}^{T} C_t' H_{\vec{\pi}_{s_t}}(L(\vec{\pi}_{s_t}))C_t \tag{3.28}$$

Podemos então substituir (3.27) e (3.28) em (3.15) e teremos o algorítmo NR dado por:

$$\hat{\vec{\pi}}_s^{\,i+1} = \hat{\vec{\pi}}_s^{\,i} - \left[\sum_{t=1}^{T} C_t' H_{\vec{\pi}_{s_t}{}^i}(L(\vec{\pi}_{s_t}^{\,i}))C_t\right]^{-1} \sum_{t=1}^{T} C_t' D_{\vec{\pi}_{s_t}{}^i}(L(\vec{\pi}_{s_t}^{\,i})) \quad \text{com} \quad i = 1, 2, \ldots \tag{3.29}$$

Se quisermos utilizar o "Scoring" de Fisher (veja a Estrutura (3.16)), podemos tomar a esperança negativa de (3.28), que, utilizando a Relação (3.22) fica:

$$I_s = \sum_{t=1}^{T} C_t' I_{s_t} C_t \tag{3.30}$$

Neste caso, o Algorítmo NR com "Scoring" de Fisher fica determinado por:

$$\vec{\hat{\pi}}_s^{\,i+1} = \vec{\hat{\pi}}_s^{\,i} + \left[\sum_{t=1}^{T} C_t' I(\vec{\pi}_{s_t}^{\,i}) C_t\right]^{-1} \sum_{t=1}^{T} C_t' D_{\vec{\pi}_{s_t}^{\,i}}(L(\vec{\pi}_{s_t}^{\,i})) \quad \text{com} \quad i = 1, 2, \ldots \tag{3.31}$$

Ainda podemos aproveitar a Estrutura (3.30) para obtenção das matrizes de variâncias e covariâncias de $\vec{\hat{\pi}}_s$, que como mencionado anteriormente, é a inversa de I_s. Hocking e Oxspring [11] mostram que podemos obter $V(\vec{\hat{\pi}}_s)$ através da estrutura acumulada:

$$V_s^t = (I + B_{s_t} C_t)\, V_s^{t-1} \quad \text{com} \quad t = 1, 2, \ldots, T \tag{3.32}$$

$$B_{s_t} = -V_s^{t-1} C_t' \left(C_t V_s^{t-1} C_t' + V_{s_t}\right)^{-1} \tag{3.33}$$

Por convenção, quando t=1, tomaremos $V_s^1 = V_{s_1} = \frac{1}{n_{s_1}} U_{s_1}$

Alternativamente, podemos escrever a expressão (3.32) como:

$$V_s^t = \prod_{l=j+1}^{t} (I + B_{s_l} C_{s_l}) V_s^j \quad \text{com} \quad j = 1, \ldots, t-1 \tag{3.34}$$

Nos exemplos numéricos apresentados no capítulo a seguir, utilizaremos as Estruturas (3.32) ou (3.34) para a obtenção de $V(\vec{\hat{\pi}}_s)$, mesmo no caso de utilização do Algorítmo EM.

Quanto à rapidez de convergência, notamos que o Algorítmo NR e o SF são melhores que o EM, mas o EM é nítidamente mais simples computacionalmente para distribuições multinomiais que os demais.

Obviamente, nos exemplos numéricos, chegamos aos mesmos resultados quando da utilização de um ou de outro algorítmo, assim, optamos por utilizar o Algorítmo NR, já que usaremos a estrutura de matriz de variâncias e covariâncias como definida nas Equações (3.32) e (3.33), com exceção do último exemplo, no qual utilizamos o Algorítmo EM.

Capítulo 4

Exemplos Numéricos

4.1 Exemplo 1:

A Tabela 4.13 do Apêndice, refere-se a um estudo no qual foi medido o grau de acuidade visual para 8577 mulheres com idade entre 30 e 39 anos e sem auxílio de lentes corretivas. Para 7477 delas temos as classificações de ambos os olhos e a análise de igualdade entre os graus de ambos os olhos, ou seja, de homogeneidade marginal foi feita, entre outros em Stuart [27] e Grizzle et al. [10]. As demais 1100 observações são dados artificiais introduzidos em Koch et al. [14] sendo 600 mulheres com graus medidos para o olho esquerdo apenas e 500 para o olho direito.

Faremos então uma análise de homogeneidade marginal segundo os métodos descritos no capítulo anterior, ou seja, escreveremos um modelo linear, colocando alguma relação das probabilidades conjuntas em função do nível de resposta assumido e da condição de avaliação (olho esquerdo ou direito), utilizando para estimação dos parâmetros os diversos métodos já discutidos e por fim, testaremos a igualdade das condições de avaliação.

De início, vamos utilizar somente os "dados completos" para a análise, ou seja, tomaremos apenas a parte completa da Tabela 4.13, ou seja, a Tabela 4.14 do Apêndice.

O Passo I é facilmente obtido já que $\widehat{\pi}_{sr}$ é a própria proporção observada $p_{sr} = \frac{n_{sr}}{n_{s+}}$. Já que temos apenas uma subpopulação, vamos deixar de colocar o índice referente a ela nos parâmetros e em seus estimadores, assim, por exemplo onde escreveríamos π_{1r} para denotar

o parâmetro relativo à categoria r e à subpopulação 1, escreveremos apenas π_r. Temos assim que as estimativas dos parâmetros da distribuição Multinomial associada à Tabela 4.14 são:

$$\vec{\hat{\pi}}=\vec{p}=[p_{(11)}=\frac{1520}{7477}\,,\,p_{(12)}=\frac{266}{7477}\,,\,\cdots\,,\,p_{(43)}=\frac{179}{7477}]'$$

A estimativa de $\pi_{(44)}$ não precisa ser calculada pois temos por restrição que a soma de todos os parâmetros vale um, assim:

$$p_{(44)}=1-\sum_{r=1}^{15}p_r$$

A matriz de variâncias e covariâncias de $\vec{p}$ é a matriz informação observada e pode ser escrita como:

$$V(\vec{p}) = \frac{1}{n}\left[Diag(\vec{p})-\vec{p}\,\vec{p}\,'\right] =$$

$$= \frac{1}{7477}\left[\begin{pmatrix} \frac{1520}{7477} & 0 & \cdots & 0 \\ 0 & \frac{266}{7477} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{179}{7477} \end{pmatrix} - \begin{pmatrix} \frac{1520}{7477} \\ \frac{266}{7477} \\ \vdots \\ \frac{179}{7477} \end{pmatrix} \begin{pmatrix} \frac{1520}{7477} & \frac{266}{7477} & \cdots & \frac{179}{7477} \end{pmatrix}\right] =$$

$$= \frac{1}{7477}\begin{pmatrix} \frac{1520}{7477}\left(1-\frac{1520}{7477}\right) & -\frac{1520\times 266}{7477^2} & \cdots & -\frac{1520\times 179}{7477^2} \\ -\frac{266\times 1520}{7477^2} & \frac{266}{7477}\left(1-\frac{266}{7477}\right) & \cdots & -\frac{266\times 179}{7477^2} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{179\times 1520}{7477^2} & -\frac{179\times 266}{7477^2} & \cdots & \frac{179}{7477}\left(1-\frac{179}{7477}\right) \end{pmatrix}$$

As probabilidades marginais podem ser obtidas através da operação:

$$\vec{\phi}=\begin{pmatrix} \phi_{11} \\ \phi_{12} \\ \phi_{13} \\ \phi_{21} \\ \phi_{22} \\ \phi_{23} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1520}{7477} \\ \frac{266}{7477} \\ \vdots \\ \frac{179}{7477} \end{pmatrix} = A\,\vec{p}$$

Também aqui notamos que não é necessário calcular todas as probabilidades marginais, visto que a soma das probabilidades de todos os níveis em cada uma das condições de avaliação vale 1, assim, não precisamos calcular ϕ_{14} nem ϕ_{24}.

Tomemos o modelo $\phi_{dk} = \mu_k + \gamma_d$ com μ_k sendo a média para o nível k e γ_d, o efeito devido ao olho pesquisado (esquerdo ou direito)[1]. Podemos escrever matricialmente tal modelo, como:

$$\vec{\phi} = X\,\vec{\beta} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \gamma_1 \end{pmatrix} \tag{4.1}$$

Sujeito às restrições:

$$\sum_{k=1}^{4} \mu_k = 1 \quad \text{e} \quad \sum_{d=1}^{2} \gamma_d = 0$$

E é por causa destas restrições que não estão no modelo a média para o nível 4 nem o efeito da condição 2.

As estimativas para os parâmetros do modelo são dadas pela Equação (2.9) e podem ser vistas na Tabela (4.1), juntamente com seus erros-padrões que nada mais são do que as raizes quadradas de suas variâncias.

Tabela 4.1: Estimativas dos parâmetros do modelo do Exemplo 1 e seus erros-padrões segundo o Método para Dados Completos.

Parâmetro	**Estimativa**	**Erro-Padrão**
μ_1	0.2596222	0.0046831
μ_2	0.2993638	0.0046424
μ_3	0.3319486	0.004827
γ_1	0.0014349	0.0005534

Testando o ajuste do modelo através da Estatística de Wald (2.11) chegamos ao valor $Q = 5.2515659$ que tem distribuição Qui-Quadrado com 6 graus de liberdade, assim, o valor

[1] A condição de avaliação para este exemplo nada mais é do que a posição do olho pesquisado, ou seja, se o olho referido é o olho direito ou o esquerdo

da probabilidade acumulada acima de Q é $p = 0.5120$, levando à aceitação do ajuste do modelo.

Para testarmos a homogeneidade marginal, podemos formular a hipótese de que os efeitos das duas condições são iguais e, além disto, iguais a zero (devido à restrição), ou seja, basta testarmos a hipótese nula $H_0 : \gamma_1 = 0$, que pode ser escrita como $H_0 : C\beta = 0$ com $C = (\ 0 \quad 0 \quad 0 \quad 1\ )$. Para tal teste utilizamos a Estatística Q_C (2.12), que resultou em $Q_C = 6.7242543$ cuja distribuição é aproximadamente Qui-Quadrado com 1 grau de liberdade, e assim, o valor $p = 0.0095$, levando-nos a rejeitar a hipótese de igualdade entre as duas condições, logo, podemos concluir que existe diferença significativa entre os dois olhos quanto à acuidade visual das 7477 mulheres pesquisadas.

O Método KO pode ser aplicado utilizando-se uma divisão da subpopulação em três estratos: o estrato completo, o estrato das mulheres que foram observadas apenas quanto ao olho direito e o estrato das que foram observadas apenas quanto ao olho esquerdo. Temos as proporções de cada um destes estratos dadas por:

$$\vec{p}_{comp} = \begin{pmatrix} \frac{1520}{7477} \\ \frac{266}{7477} \\ \vdots \\ \frac{179}{7477} \end{pmatrix} \quad \vec{p}_{od} = \begin{pmatrix} \frac{140}{500} \\ \frac{150}{500} \\ \frac{160}{500} \end{pmatrix} \quad \text{e} \quad \vec{p}_{oe} = \begin{pmatrix} \frac{160}{600} \\ \frac{180}{600} \\ \frac{200}{600} \end{pmatrix}$$

As matrizes de variâncias e covariâncias dos $\vec{p_i}$ são da forma:

$$V(\vec{p_i}) = \frac{1}{n_i}\left(Diag(\vec{p_i}) - \vec{p_i}\vec{p_i}'\right) \quad i = comp, od, oe$$

As relações destas proporções com os parâmetros da distribuição esperada podem ser escritos como :

$$\vec{p_i} = Z_i\,\vec{\pi} \quad \text{com} \quad i = comp, od, oe$$

$$Z_{comp} = I_{15\times 15} \quad , \quad Z_{od} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

$$Z_{oe} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

As estimativas dos parâmetros da distribuição esperada são obtidas através do MMQP, ou seja:

$$\vec{\hat{\pi}} = \left(\sum_{i=comp}^{oe} Z_i' V(\vec{p_i})^{-1} Z_i \right)^{-1} \left(\sum_{i=comp}^{oe} Z_i' V(\vec{p_i})\, \vec{p_i} \right)$$

Substituindo as matrizes Z_i, $V(\vec{p_i})$ e os vetores $\vec{p_i}$, obtivemos:

$$\vec{\hat{\pi}} = \begin{pmatrix} 0.2046481 \\ 0.0357297 \\ 0.0166384 \\ 0.0087793 \\ 0.0313863 \\ 0.2023284 \\ 0.0577469 \\ 0.010336 \\ 0.015674 \\ 0.0483819 \\ 0.2365794 \\ 0.0271315 \\ 0.0048157 \\ 0.0109432 \\ 0.0238628 \end{pmatrix}$$

Prosseguiremos tomando as probabilidades marginais, bastando para isso utilizar a mesma matriz A do Método para Dados Completos, e escrevendo o Modelo (4.1). As estimativas para os parâmetros do modelo são obtidas de maneira idêntica ao Método para Dados Completos e podem ser vistas na Tabela 4.2.

A Estatística Q, utilizada para testar o ajuste do modelo vale 5.5723783, que com 6 graus de liberdade tem probabilidade acumulada acima valendo $p = 0.4727$ levando-nos a

Tabela 4.2: Estimativas dos parâmetros do modelo do Exemplo 1 e seus erros-padrões segundo o Método KO.

Parâmetro	**Estimativa**	**Erro-Padrão**
μ_1	0.2610862	0.0044214
μ_2	0.2995101	0.0044002
μ_3	0.3314	0.0045678
γ_1	0.0013763	0.0005442

aceitar o modelo.

Testando a hipótese de homogeneidade marginal $C\vec{\beta}= 0$, com a mesma matriz C do Método para Dados Completos, obtemos $Q_C = 6.3966186$ que com 1 grau de liberdade fornece o valor $p = 0.0114$, sugerindo ao nível de significância de 5% a não igualdade dos efeitos dos dois olhos.

Para aplicarmos o Método WC, basta considerarmos o vetor

$$\vec{p}= [\frac{1520}{8577}, \cdots, \frac{492}{8577}, \frac{140}{8577}, \cdots, \frac{50}{8577}, \frac{160}{8577}, \cdots, \frac{60}{8577}]'$$

como vetor de estimativas dos parâmetros da distribuição esperada já que estamos considerando o fato da observação estar faltando como mais um nível de resposta.

Em seguida, tomamos as razões das probabilidades marginais de cada nível k pelas probabilidades marginais da resposta estar presente em cada condição de avaliação:

$$\vec{\lambda}= [\frac{\phi^+_{11}}{\phi^o_{1+}}, \cdots, \frac{\phi^+_{13}}{\phi^o_{1+}}, \frac{\phi^+_{21}}{\phi^o_{2+}}, \cdots, \frac{\phi^+_{23}}{\phi^o_{2+}}]' = \exp[B \ln(A\ \vec{p})]$$

$$\text{com} \quad A = \begin{pmatrix}
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1
\end{pmatrix}$$

$$\text{e} \quad B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{pmatrix}$$

O modelo fica definido então como $\vec{\lambda} = X\ \vec{\beta}$ com X e $\vec{\beta}$ idênticos aos do Método para Dados Completos. Prossegue-se com a estimação dos parâmetros pelo MMQP. As estimativas e seus erros-padrões podem ser vistas na Tabela 4.3.

Tabela 4.3: Estimativas dos parâmetros do modelo do Exemplo 1 e seus erros-padrões segundo o Método WC.

Parâmetro	**Estimativa**	**Erro-Padrão**
μ_1	0.2604814	0.004458
μ_2	0.2994318	0.0044266
μ_3	0.3316666	0.0045996
γ_1	0.0012968	0.0005569

A estatística teste do ajuste resultou em $Q = 5.1964489$ que com 6 graus de liberdade tem probabilidade acumulada acima $p = 0.5189$, sugerindo a adequacidade do modelo.

O teste de homogeneidade marginal resultou em $Q_C = 5.4211156$ que apresenta valor $p = 0.0199$ e, assim como nos outros métodos já utilizados, levam à conclusão de que os efeitos relativos ao olho esquerdo e ao direito não são iguais.

Para a utilização do Método LLH vamos utilizar o Algorítmo NR para estimar os parâmetros da distribuição esperada e, seguindo os passos indicados em Hocking e Oxspring [11], tomamos os mesmos estratos do Método KO e utilizamos as matrizes de ligação entre os parâmetros das distribuições de cada estrato, também as mesmas do Método KO, ou seja, $C_i = Z_i$ para $i = comp, od, oe$. Com o auxílio do "software" *Mathematica* e utilizando as Estruturas de (3.17) a (3.15), chegamos aos valores $\vec{\pi} =$ [0.204668, 0.0357289,

0.0166378, 0.00878651, 0.0313832, 0.202287, 0.0577349, 0.0103433, 0.0156729, 0.0483736, 0.236539, 0.0271523, 0.00481558, 0.0109421, 0.0238603, 0.0650741]'.

As probabilidades marginais bem como o modelo linear utilizado no Passo II são os mesmos dos métodos anteriores e temos as estimativas dos parâmetros listadas na Tabela 4.4.

Tabela 4.4: Estimativas dos parâmetros do modelo do Exemplo 1 e seus erros-padrões segundo o Método LLH.

Parâmetro	**Estimativa**	**Erro-Padrão**
μ_1	0.2611112	0.0041591
μ_2	0.2994578	0.0041356
μ_3	0.3313531	0.0042931
γ_1	0.0013821	0.0005084

A estatística teste de ajuste do modelo resultou em $Q = 6.358485$, que com 6 graus de liberdade tem probabilidade acumulada acima $p = 0.3842$, levando-nos a aceitar o modelo.

Já a estatística teste de nulidade do efeito das condições, ou seja, de igualdade entre efeito dos dois olhos, resultou em $Q_C = 7.3915528$, com probabilidade acumulada acima (1 g.l.) $p = 0.0065$, e como nos demais métodos, concluímos que os dois olhos não possuem efeitos iguais quanto à acuidade.

Por fim, para efeito de comparação, observemos a Tabela 4.5 que contém as estimativas dos parâmetros, seus erros-padrões (entre parênteses), os valores das estatísticas Q e Q_C e suas probabilidades acumuladas acima p (entre parênteses).

Podemos notar, olhando para a tabela acima que o Método que apresenta menores erros-padrões é o LLH, mas por outro lado, o ajuste deste método é o pior de todos eles. Deixaremos para analisar estes fatos no Capítulo final.

Tabela 4.5: Estimativas dos parâmetros do modelo do Exemplo 1 e seus erros-padrões segundo todos os métodos descritos.

Método	Completos	KO	WC	LLH
μ_1	0.2596222	0.2610862	0.2604814	0.2611112
	(0.0046831)	(0.0044214)	(0.004458)	(0.0041591)
μ_2	0.2993638	0.2995101	0.2994318	0.2994578
	(0.0046424)	(0.0044002)	(0.0044266)	(0.0041356)
μ_3	0.3319486	0.3314	0.3316666	0.3313531
	(0.004827)	(0.0045678)	(0.0045996)	(0.0042931)
γ_1	0.0014349	0.0013763	0.0012968	0.0013821
	(0.0005534)	(0.0005442)	(0.0005569)	(0.0005084)
Q	5.2515659	5.5723783	5.1964489	6.358485
	(0.5120)	(0.4727)	(0.5189)	(0.3842)
$\mathbf{Q_C}$	6.7242543	6.3966186	5.4211156	7.3915528
	(0.0095)	(0.0114)	(0.0199)	(0.0065)

4.2 Exemplo 2:

Foi feita uma análise clínica de dois períodos com pacientes subdivididos em dois grupos de acordo com idade (idosos e jovens) e cada grupo subdividido em três subgrupos que receberam sequências diferentes de tratamentos. Verifica-se então se a resposta aos tratamentos foi classificada como favorável ou não-favorável. Temos então $s = 6$ subpopulações (resultantes das combinações de $g = 2$ grupos e $t = 3$ sequências de tratamentos), $d = 2$ condições de avaliação e $k = 2$ níveis de resposta, gerando assim, $r = 4$ categorias de resposta. A apresentação e análise das observações completas está feita em Koch et al. [15]. A inclusão das observações incompletas foi feita por nós para podermos mostrar a utilização dos métodos descritos no Capítulo 3. Nosso interesse com tal inserção está apenas em mostrar que os três métodos funcionam de maneira semelhante e compará-los, mesmo que com dados artificiais. Técnicas específicas para inserção de dados podem ser utilizadas, a fim de minimizar

a tendenciosidade, mas não foram consideradas neste estudo. Resultados mais expressivos estatisticamente podem ser vistos no exemplo 3, já que todos os dados são reais. Os dados completos (reais) e incompletos (artificiais) podem ser vistos na Tabela 4.15 do Apêndice.

Como fizemos no exemplo anterior, vamos aplicar todos os métodos para podermos comparar os resultados, iniciando pelo Método para Dados Completos.

As estimativas das probabilidades conjuntas π_{sr} são as proporções observadas, com matrizes de variâncias e covariâncias dadas por:

$$V(\vec{p_s}) = \frac{1}{n_s}\left(Diag(\vec{p_s}) - \vec{p_s}\vec{p_s}'\right)$$

Construimos então as probabilidades marginais de resposta favorável (F) através da operação:

$$\vec{\phi_s} = A\,\vec{p_s} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}\vec{p_s} \quad \text{com} \quad \vec{\phi_s} = \begin{pmatrix} \phi_{s1F} \\ \phi_{s2F} \end{pmatrix} \quad s = 1, \ldots, 6.$$

Podemos tomar o modelo:

$$\vec{\phi_g} = \vec{\mu_g} + \vec{\psi_{gt}} + \vec{\tau_{gd}} \qquad g = I, J \qquad t = A, B, P \qquad d = 1, 2$$

com
- μ_g : média do grupo g,
- ψ_{gt} : efeito devido ao tratamento t no grupo g
- τ_{gd} : efeito devido à condição d no grupo g.

Temos, na forma matricial, o modelo definido por:

$$\vec{\phi_g} = X_g\,\vec{\beta} \qquad g = I, J \qquad \text{com}$$

$$X_I = \begin{pmatrix} 1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} \quad X_J = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix} \text{ e}$$

$$\vec{\beta} = [\mu_I\,,\, \psi_{IA}\,,\, \psi_{IB}\,,\, \tau_{I2}\,,\, \mu_J\,,\, \psi_{JA}\,,\, \psi_{JB}\,,\, \tau_{J2}]'$$

As estimativas destes parâmetros podem ser vistas na Tabela 4.6, bem como seus erros-padrões.

Tabela 4.6: Estimativas dos parâmetros do modelo do Exemplo 2 e seus respectivos erros-padrões.

Par.	μ_I	ψ_{IA}	ψ_{IB}	τ_{I2}	μ_J	ψ_{JA}	ψ_{JB}	τ_{J2}
Est.	0.34444	0.16570	-0.03988	0.04883	0.60052	0.15003	-0.04082	0.12765
E.P.	0.02931	0.03687	0.03198	0.02210	0.02768	0.03121	0.03861	0.02309

A estatística teste de ajuste resultou em $Q = 0.2439817 (valorp = 0.9931)$ com 4 graus de liberdade, levando-nos a aceitar o modelo. A Tabela 4.7 apresenta o resultado de alguns testes sobre as estimativas dos parâmetros que sugere-nos a simplificação do modelo já que aceitamos a hipótese de que $\psi_{IA} = \psi_{JA}$ e $\psi_{IB} = \psi_{JB}$, ou seja, de que para cada tratamento, não existe diferença de grupo.

Tabela 4.7: Testes de igualdades de alguns parâmetros do modelo do Exemplo 2.

Hipóteses	Q_C	p	$g.l.$
$\psi_{IA} = \psi_{JA}; \psi_{IB} = \psi_{JB}$	0.1609848	0.9227	2
$\tau_{I2} = \tau_{J2}$	6.0765112	0.0137	1
$\mu_I = \mu_J$	40.344797	2.12872×10^{-10}	1

Então podemos condensar o vetor de parâmetros para $\vec{\beta} = [\mu_I, \mu_J, \psi_A, \psi_B, \tau_{I2}, \tau_{J2}]'$
E as matrizes X_g ficam:

$$X_I = \begin{pmatrix} 1 & 0 & 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & -1 & 0 \\ 1 & 0 & -1 & -1 & 1 & 0 \\ 1 & 0 & -1 & -1 & -1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} \quad \text{e} \quad X_J = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & -1 & 0 & 1 \\ 0 & 1 & -1 & -1 & 0 & -1 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Com este modelo, aplicamos novamente o MMQP e obtivemos as estimativas listadas na Tabela 4.8 que, por simplicidade de apresentação, contém as estimativas dos parâmetros do modelo melhorado para todos os métodos aplicados. Também o valor da estatística teste de ajuste com sua respectiva probabilidade acumulada acima, está numa tabela geral (Tabela 4.9), que contém também os valores de tal estatística para os demais métodos.

Como neste trabalho não estamos interessados na escolha do melhor modelo que explica os dados, vamos tomar apenas o modelo melhorado para utilização nos outros métodos.

Para a aplicação do Método KO, cada subpopulação foi subdividida em 3 estratos: o estrato completo, o estrato dos que faltaram na 1ª condição de avaliação e o dos que faltaram na 2ª. A relação entre os parâmetros das distribuições destes estratos pode ser escrita como:

$$\vec{p_{s_t}} = Z_t \, \vec{\pi_s} \quad s = 1, \ldots, 6 \; t = 1, 2, 3$$

$$\text{com } Z_1 = I_{3\times 3} \; Z_2 = \begin{pmatrix} 1 & 0 & 1 \end{pmatrix} \text{ e } Z_3 = \begin{pmatrix} 1 & 1 & 0 \end{pmatrix}$$

Concluímos então o Passo I aplicando o MMQP para obtenção de $\vec{\widehat{\pi}_s}$. Após, construímos as probabilidades marginais de resposta favorável e seguimos com o Passo II idêntico ao Método para Dados Completos. Os resultados podem ser vistos na Tabela 4.8 e a estatística teste de ajuste na Tabela 4.9.

O Método WC considera o fato da observação estar faltando como um nível de resposta, assim, teremos 3 níveis e consequentemente 8 categorias de resposta.

Escrevemos o modelo $\exp(B \log(A \, \vec{p_s})) = X_s \, \vec{\beta}$ com:

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix} \; ; \; B = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

$$X_1 = \begin{pmatrix} 1 & 0 & 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \end{pmatrix} \; ; \; X_2 = \begin{pmatrix} 1 & 0 & 0 & 1 & -1 & 0 \\ 1 & 0 & -1 & -1 & 1 & 0 \end{pmatrix}$$

$$X_3 = \begin{pmatrix} 1 & 0 & -1 & -1 & -1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} \; ; \; X_4 = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

$$X_5 = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & -1 & 0 & 1 \end{pmatrix} ; X_6 = \begin{pmatrix} 0 & 1 & -1 & -1 & 0 & -1 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

As estimativas dos elementos do vetor $\vec{\beta}$ estão na Tabela 4.8 e o valor da estatística Q, na Tabela 4.9.

Por fim, utilizamo-nos do Algotítmo NR para obtermos as estimativas da distribuição esperada e após escrevemos o modelo marginal do Método para Dados Completos e obtivemos as estimativas segundo o Método LLH.

Também para este, as estimativas encontram-se na Tabela 4.8 e a estatística de ajuste na Tabela 4.9.

Tabela 4.8: Estimativas e erros-padrões dos parâmetros do modelo melhorado do Exemplo 2, segundo todos os métodos descritos.

Parâmetro	**Comp.**	**KO**	**WC**	**LLH**
μ_I	0.3430271	0.3308148	0.336865	0.3375042
	(0.0290878)	(0.0252253)	(0.0307619)	(0.0129872)
μ_J	0.5990329	0.5924538	0.5985752	0.6061112
	(0.0274203)	(0.0237526)	(0.0291725)	(0.0163152)
ψ_A	0.1560485	0.1711929	0.1644925	0.154454
	(0.0237196)	(0.0214857)	(0.0260028)	(0.0086975)
ψ_B	-0.038874	-0.032059	-0.032916	-0.042657
	(0.024384)	(0.0221964)	(0.0266975)	(0.0094915)
τ_{I2}	0.0483156	0.0501	0.0487034	0.0476185
	(0.022025)	(0.0200943)	(0.0243465)	(0.0139581)
τ_{J2}	0.1284236	0.1400268	0.1326604	0.1294085
	(0.0230053)	(0.0206573)	(0.0253886)	(0.0168927)

Um fato interessante que notamos observando os resultados nas Tabelas 4.8 e 4.9 é que o Método WC apresenta erros-padrões muito grandes em relação aos outros métodos, mas apresenta ajuste melhor que os outros que utilizam as observações incompletas. Já o Método

Tabela 4.9: Valores das Estatísticas Q do Exemplo 2 e suas respectivas probabilidades acumuladas acima segundo todos os métodos.

Parâmetro	Comp.	KO	WC	LLH
Q	0.4049664	2.5967056	0.7150943	6.5924146
p	0.9989	0.8575	0.9942	0.3602

LLH, assim como havíamos notado no exemplo anterior, apresenta os menores erros-padrões e o pior ajuste de todos os métodos.

4.3 Exemplo 3:

Este exemplo refere-se a uma parte de um Estudo de Fatores de Riscos Coronários, relatado em Woolson e Clarke [28], que teve início em 1971 e seguiu de dois em dois anos até 1981. Trata-se da classificação de crianças em idade escolar quanto à obesidade. As crianças foram divididas em 10 subpopulações segundo sexo (M e F) e idade no início da pesquisa (de 5 a 7 anos, de 7 a 9,..., de 13 a 15) e o critério adotado para classificação de obeso (O) ou não obeso (N) foi o peso relativo: tomou-se a razão ($\times$ 100) do peso da criança pela mediana de seu grupo e se o valor resultante superasse 110, a criança era classificada como obesa. Para facilitar nosso trabalho, tomamos apenas as três últimas etapas, ou seja, os dados relativos aos anos de 1977, 1979 e 1981. As frequências de cada categoria de resposta estão na Tabela 4.16 do Apêndice.

Para aplicação do Método para Dados Completos basta tomarmos as proporções de cada casela como estimativas dos parâmetros das distribuições esperadas para cada subpopulação e escrevermos o modelo marginal:

$$\vec{\phi}_s = A\,\vec{p}_s = X_s\,\vec{\beta} \qquad \text{com} \tag{4.2}$$

$$A = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix};$$

$$X_1 = \begin{pmatrix} 1 & 6 & 36 & 0 & 0 & 0 \\ 1 & 8 & 64 & 0 & 0 & 0 \\ 1 & 10 & 100 & 0 & 0 & 0 \end{pmatrix} ; X_2 = \begin{pmatrix} 1 & 8 & 64 & 0 & 0 & 0 \\ 1 & 10 & 100 & 0 & 0 & 0 \\ 1 & 12 & 144 & 0 & 0 & 0 \end{pmatrix}$$

$$X_3 = \begin{pmatrix} 1 & 10 & 100 & 0 & 0 & 0 \\ 1 & 12 & 144 & 0 & 0 & 0 \\ 1 & 14 & 196 & 0 & 0 & 0 \end{pmatrix} ; X_4 = \begin{pmatrix} 1 & 12 & 144 & 0 & 0 & 0 \\ 1 & 14 & 196 & 0 & 0 & 0 \\ 1 & 16 & 256 & 0 & 0 & 0 \end{pmatrix}$$

$$X_5 = \begin{pmatrix} 1 & 14 & 196 & 0 & 0 & 0 \\ 1 & 16 & 256 & 0 & 0 & 0 \\ 1 & 18 & 324 & 0 & 0 & 0 \end{pmatrix} ; X_6 = \begin{pmatrix} 0 & 0 & 0 & 1 & 6 & 36 \\ 0 & 0 & 0 & 1 & 8 & 64 \\ 0 & 0 & 0 & 1 & 10 & 100 \end{pmatrix}$$

$$X_7 = \begin{pmatrix} 0 & 0 & 0 & 1 & 8 & 64 \\ 0 & 0 & 0 & 1 & 10 & 100 \\ 0 & 0 & 0 & 1 & 12 & 144 \end{pmatrix} ; X_8 = \begin{pmatrix} 0 & 0 & 0 & 1 & 10 & 100 \\ 0 & 0 & 0 & 1 & 12 & 144 \\ 0 & 0 & 0 & 1 & 14 & 196 \end{pmatrix}$$

$$X_9 = \begin{pmatrix} 0 & 0 & 0 & 1 & 12 & 144 \\ 0 & 0 & 0 & 1 & 14 & 196 \\ 0 & 0 & 0 & 1 & 16 & 256 \end{pmatrix} \text{ e } X_{10} = \begin{pmatrix} 0 & 0 & 0 & 1 & 14 & 196 \\ 0 & 0 & 0 & 1 & 16 & 256 \\ 0 & 0 & 0 & 1 & 18 & 324 \end{pmatrix}$$

$$\vec{\beta} = [\mu_M , \alpha_M , \gamma_M , \mu_F , \alpha_F , \gamma_F]'$$

com
- μ_M : média para os grupos de sexo masculino,
- α_M : efeito linear da idade para o sexo masculino,
- γ_M : efeito quadrático da idade para o sexo masculino,
- μ_F : média para os grupos de sexo feminino,
- α_F : efeito linear da idade para o sexo feminino
- γ_F : efeito quadrático da idade para o sexo feminino.

As estimativas segundo o MMQP podem ser vistas na Tabela 4.11. A estatística teste de ajuste está na Tabela 4.10 e tem distribuição aproximadamente Qui-quadrado com 24 graus de liberdade; assim, aceitamos a adequacidade do modelo.

Podemos ainda testar a hipótese de não influência da covariável sexo na resposta (obesidade), bastando para isso testarmos a hipótese:

$$H_0 : C\,\vec{\beta} = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{pmatrix} \vec{\beta}$$

Temos então os valores da estatística Q_C, que tem distribuição aproximadamente Qui-quadrado com 3 graus de liberdade, sugerindo ao nível $\alpha = 1\%$ a simplificação do modelo para:

$$\vec{\phi_s} = X_s\,\vec{\beta} \quad \text{com} \quad \vec{\beta} = [\mu\,,\,\alpha\,,\,\gamma]'$$

$$\text{onde} \begin{cases} \mu : \text{média geral} \\ \alpha : \text{efeito linear para a idade} \\ \gamma : \text{efeito quadrático para a idade} \end{cases}$$

$$X_1 = X_6 = \begin{pmatrix} 1 & 6 & 36 \\ 1 & 8 & 64 \\ 1 & 10 & 100 \end{pmatrix} \quad ; \quad X_2 = X_7 = \begin{pmatrix} 1 & 8 & 64 \\ 1 & 10 & 100 \\ 1 & 12 & 144 \end{pmatrix}$$

$$X_3 = X_8 = \begin{pmatrix} 1 & 10 & 100 \\ 1 & 12 & 144 \\ 1 & 14 & 196 \end{pmatrix} \quad ; \quad X_4 = X_9 = \begin{pmatrix} 1 & 12 & 144 \\ 1 & 14 & 196 \\ 1 & 16 & 256 \end{pmatrix} \quad e$$

$$X_5 = X_{10} = \begin{pmatrix} 1 & 14 & 196 \\ 1 & 16 & 256 \\ 1 & 18 & 324 \end{pmatrix}$$

As estimativas dos parâmetros deste modelo estão na Tabela 4.12, onde também se encontram as estimativas para os demais métodos. Os valores da estatística teste de ajuste para este e para os outros métodos estão listadas na Tabela 4.10.

Para aplicação do Método KO, cada subpopulação foi dividida em 7 estratos: completo (1), faltando na 1ª condição (2), faltando na 2ª (3), na 3ª (4), na 1ª e 2ª (5), na 1ª e 3ª (6) e faltando na 2ª e 3ª (7).

As matrizes que relacionam os parâmetros dos estratos com os parâmetros da distribuição esperada são:

$$Z_1 = I_{7\times 7} \quad ; \quad Z_2 = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \quad ;$$

$$Z_3 = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix} \quad ; \quad Z_4 = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \quad ;$$

$$Z_5 = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \quad ; \quad Z_6 = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix} \quad \text{e}$$

$$Z_7 = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Aplicamos então o MMQP e obtivemos $\vec{\widehat{\pi}}_s$. A partir de então, seguimos com a construção das probabilidades marginais de resposta "obeso" e com o modelo diferenciando os grupos de sexo masculino dos grupos de sexo feminino (Modelo (4.2)) e obtivemos os resultados que aparecem na Tabela 4.11. A estatística teste de ajuste pode ser vista na Tabela 4.10 e indica que segundo o Método KO, este modelo não é adequado.

Também com o modelo sem efeito de sexo não obtivemos ajuste com este método. As estimativas dos parâmetros estão na Tabela 4.12 e uma coisa interessante de se notar é que os erros-padrões de tais estimativas são menores que os dos outros métodos.

A aplicação do Método WC aos dados também foi feita utilizando-se para isto as

matrizes:

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

Os resultados estão nas Tabelas 4.11, 4.12 e 4.10 e como nos outros exemplos podemos notar que este é o método que apresenta melhores ajustes (valores de p maiores).

Por fim, utilizamos o Algorítmo EM para obtenção do Passo I do Método LLH e, com o Passo II idêntico ao Método para Dados Completos chegamos aos resultados que estão também nas Tabelas 4.11, 4.12 e 4.10.

Tabela 4.10: Valores das Estatísticas Q do Exemplo 3 e suas respectivas probabilidades acumuladas acima p.

	Comp.	**KO**	**WC**	**LLH**
Q(com efeito sexo)	32.236638	43.748186	26.549767	35.996469
	(0.1212)	(0.0082)	(0.3259)	(0.0549)
Q_C(nulidade	0.980469	6.2147751	3.9337466	6.6043266
do efeito sexo)	(0.8060)	(0.1016)	(0.2687)	(0.0856)
Q(sem efeito sexo)	33.217107	49.962961	30.483514	42.600796
	(0.1900)	(0.0046)	(0.2929)	(0.0287)

Para encerrarmos este capítulo, gostaríamos de lembrar que os algorítmos iterativos foram implementados no "software" Mathematica (Wolfrang Research) e as estimações pelo MMQP e os cálculos matriciais foram feitos no SAS (SAS Institute).

Tabela 4.11: Estimativas e erros-padrões para os parâmetros do modelo com efeito de sexo do Exemplo 3, segundo todos os métodos descritos.

	Comp.	**KO**	**WC**	**LLH**
μ_M	-0.20769	-0.230182	-0.222128	-0.251909
	(0.07155)	(0.0570458)	(0.0643767)	(0.0574844)
α_M	0.06752	0.0680778	0.0692826	0.0738298
	(0.01263)	(0.0101957)	(0.0114472)	(0.0102823)
γ_M	-0.00264	-0.002586	-0.002669	-0.002824
	(0.00053)	(0.000436)	(0.0004897)	(0.0004403)
μ_F	-0.13639	-0.091773	-0.178661	-0.15346
	(0.08345)	(0.0659704)	(0.0765295)	(0.067898)
α_F	0.05202	0.046853	0.0656772	0.0589985
	(0.01443)	(0.0114338)	(0.013251)	(0.0117967)
γ_F	-0.00193	-0.001697	-0.002534	-0.002148
	(0.00060)	(0.0004796)	(0.0005543)	(0.0004952)

Tabela 4.12: Estimativas e erros-padrões para os parâmetros do modelo sem efeito de sexo do Exemplo 3, segundo todos os métodos descritos.

	Comp.	**KO**	**WC**	**LLH**
μ	-0.176469	-0.174438	-0.20687	-0.21372
	(0.0542924)	(0.0430519)	(0.0491913)	(0.0437945)
α	0.0605182	0.0594689	0.0683587	0.0680952
	(0.0094896)	(0.0075885)	(0.0086416)	(0.0077309)
γ	-0.002317	-0.002217	-0.002632	-0.002552
	(0.0003985)	(0.0003218)	(0.000366)	(0.0003281)

4.4 Apêndice: Tabelas de Dados

Tabela 4.13: Frequências de mulheres entre 30 e 39 anos de idade classificadas de acordo com o grau de acuidade visual em cada olho sem auxílio de lentes corretivas - Exemplo 1.

	olho esquerdo						
olho direito	Maior grau (1)	2º grau (2)	3º grau (3)	Menor grau (4)	subtot.	só olho dir.	total
Maior grau (1)	1520	266	124	66	1976	140	2116
2º grau (2)	234	1512	432	78	2256	150	2406
3º grau (3)	117	362	1772	205	2456	160	2616
Menor grau (4)	36	82	179	492	789	50	839
subtotal	1907	2222	2507	841	7477	500	7977
só olho esq.	160	180	200	60	600	-	-
total	2067	2402	2707	901	8077	-	8577

Tabela 4.14: Parte completa da tabela de frequências de mulheres entre 30 e 39 anos de idade classificadas de acordo com o grau de acuidade visual em cada olho sem auxílio de lentes corretivas - Exemplo 1.

	olho esquerdo				
	Maior grau	2º grau	3º grau	Menor grau	
olho direito	(1)	(2)	(3)	(4)	total
Maior grau (1)	1520	266	124	66	1976
2º grau (2)	234	1512	432	78	2256
3º grau (3)	117	362	1772	205	2456
Menor grau (4)	36	82	179	492	789
total	1907	2222	2507	841	7477

Tabela 4.15: Frequências de respostas favorável (F) ou não- favorável (N) para unidades agrupadas segundo idade e sequência de tratamentos - Exemplo 2.

		categorias de resposta								
Idade	*Seq.*	FF	FN	NF	NN	*F	*N	F*	N*	Tot.
I	A:B	12	12	6	20	2	5	7	8	72
I	B:P	8	5	6	31	1	7	3	9	70
I	P:A	5	3	22	20	5	3	1	12	71
J	B:A	19	3	25	3	8	1	5	5	69
J	A:P	25	6	6	13	10	4	8	2	74
J	P:B	13	5	21	11	8	3	1	9	71

Tabela 4.16: Frequências de respostas O (obeso), N (não obeso) e * (resposta não observada) - Exemplo 3.

	M,5-7 (1)	M,7-9 (2)	M,9-11 (3)	M,11-13 (4)	M,13-15 (5)	F,5-7 (6)	F,7-9 (7)	F,9-11 (8)	F,11-13 (9)	F,13-15 (10)
NNN	90	150	152	119	101	75	154	148	129	91
NNO	9	15	11	7	4	8	14	6	8	9
NON	3	8	8	8	2	2	13	10	7	5
NOO	7	8	10	3	7	4	19	8	9	3
ONN	0	8	7	13	8	2	2	12	6	6
ONO	1	9	7	4	0	2	6	0	2	0
OON	1	7	9	11	6	1	6	8	7	6
OOO	8	20	25	16	15	8	21	27	14	15
NN*	16	38	48	42	82	20	25	36	36	83
NO*	5	3	6	4	9	0	3	0	9	15
ON*	0	1	2	4	8	0	1	7	4	6
OO*	0	11	14	13	12	4	11	17	13	23
N*N	9	16	13	14	6	7	16	8	31	5
N*O	3	6	5	2	1	2	3	1	4	0
O*N	0	1	0	1	0	0	0	1	2	0
O*O	0	3	3	4	1	1	4	4	6	1
*NN	129	42	36	18	13	109	47	39	19	11
*NO	18	2	5	3	1	22	4	6	1	1
*ON	6	3	4	3	2	7	1	7	2	2
*OO	13	13	3	1	2	24	8	13	2	3
N**	32	45	59	82	95	23	47	53	58	89
O**	5	7	17	24	23	5	7	16	37	32
N	33	33	31	23	34	27	23	25	21	43
O	11	4	9	6	12	5	5	9	1	15
**N	70	55	40	37	15	65	39	23	23	14
**O	24	14	9	14	3	19	13	8	10	5

Capítulo 5

Considerações finais

Foram apresentados três métodos que visam utilizar as informações das unidades amostrais que não participaram do estudo em todas as condições de avaliação, chamadas de observações incompletas: o Método KO, o Método WC e o Método LLH. Entre eles, o LLH destaca-se por supor que a estrutura dos dados faltando é MAR (fato do dado estar faltando pode depender das respostas observadas, mas não da resposta que seria observada naquela condição de avaliação), que é muito mais realista do que a estrutura MCAR (fato da observação nao depende nem dos valores já observados nem do valor que seria observado naquela condição de avaliação) suposta pelos outros dois métodos. Lembremos que a estrutura MCAR exige que, o fato do valor não ter sido observado, não dependa nem dos valores observados em outras condições de avaliação, nem do valor que seria observado naquela condição, enquanto que a estrutura MAR permite que ocorra o primeiro tipo de dependência, mas não o segundo.

Quanto à facilidade computacional o Método WC supera os demais. O Método KO peca neste aspecto pois a divisão em estratos torna os algorítmos de cálculos matriciais muito extensos. Além disso, a matriz de variâncias e covariâncias das estimativas de $\vec{\pi}_s$, os parâmetros da distribuição esperada, é quase sempre singular quando da presença de zeros, tornando necessário substituí-los por outros valores. Através de exemplos numéricos verificamos que as substituições desses zeros por valores próximos de 1 levam a ajustes bons, mas erros-padrões muito acima dos esperados, enquanto que substituições por valores

próximos de 0, os ajustes perdem em qualidade. Possivelmente seja devido a estes fatos que em Koch et al. [15] está sugerido a utilização do valor 0.5 como substituto dos zeros.

A necessidade de se utilizar um algorítmo iterativo faz com que, em aspectos computacionais, o Método LLH não seja vantajoso já que rotinas iterativas "carregam" o sistema provocando aumento no tempo de execução e maior necessidade de memória. Quanto aos algorítmos que apresentamos, Fuchs [9] descreve o EM de maneira bastante simples para distribuições multinomiais, mas sua matriz de variâncias e covariâncias, que é a inversa da matriz de informação observada de Fisher exige o cálculo de derivadas de segunda ordem, esperanças e inversas que são rotinas "grandes" em relação a espaço de memória.

Nós optamos neste trabalho por utilizarmos mais o Algorítmo NR, já que existem grandes facilidades proporcionadas por Hocking e Oxspring [11] que, inclusive, detalham uma boa maneira de se obter a matriz de variâncias e covariâncias. O que fizemos foi adaptar esta estrutura para determinar a matriz de variâncias e covariâncias para o Algorítmo EM.

Ainda quanto às diferenças entre os algorítmos iterativos, o NR tem convergência mais rápida que o EM, mas existem métodos para acelerar a convergência do EM, não abordados neste trabalho, já que fogem ao escopo, que talvez venham a compensar tal desvantagem.

Quanto aos resultados dos exemplos numéricos apresentados, observamos que o Método LLH apresenta menores erros-padrões que os demais, contudo com o Método WC, obtêm-se melhores ajustes.

No Exemplo 1, podemos notar que os resultados dos três métodos descritos não sugere melhorias significativas em relação ao Método para Dados Completos. Isto indica que as unidades que apresentam observações incompletas fornecem poucas informações novas em relação às unidades que apresentam observações completas. Talvez este fato se deva à cautela que Koch et al. [14] tiveram ao introduzir as observações incompletas artificialmente, gerando observações tendenciosas, e ainda ao fato do número de unidades amostrais ser muito grande.

O mesmo fato acontece no Exemplo 2. Cabe ainda notar que nestes dois exemplos, os métodos que supõem estrutura de dados faltando MCAR (Métodos KO e WC) forneceram ajustes melhores que o Método LLH, que supõe estrutura MAR.

O bom comportamento do Método LLH pode ser visto no Exemplo 3, já que este

fornece erros-padrões pequenos e bom ajuste, enquanto que os outros dois, apresentam-se eficazes em um aspecto, mas não em outro: o KO fornece erros-padrões pequenos mas o ajuste do modelo é rejeitado, enquanto que o WC, bom ajuste mas erros-padrões maiores que os dos outros métodos.

Outros pontos, relacionados com este trabalho, e que poderão ser explorados posteriormente são o estudo dos vieses gerados pela utilização dos Métodos KO ou WC quando da presença da estrutura MAR, bem como, a adaptação dos testes MCAR $\times$ MAR para dados categóricos.

Bibliografia

[1] Agresti, A. (1990). *Categorical data analysis.* New York, John Wiley & Sons, Inc, 558p.

[2] Andreoni, S. (1989). *Modelos de efeitos aleatórios para análise de dados longitudinais não balanceados em relação ao tempo.* Dissertação apresentada ao IME - USP, para obtenção do título de Mestre em Estatística. São Paulo. S.P.

[3] Bard, Y. (1974). *Nonlinear parameter estimation.* New York, Academic Press, 341p.

[4] Bryant, E., Gillings, D. (1985). Statistical analysis of longitudinal repeated measures design. *Biostatistics*, 251-282.

[5] Dall'Agnol, I. (1990). *Alguns aspectos da análise em estudos longitudinais com respostas categorizadas.* Dissertação apresentada ao IMECC - Unicamp, para obtenção do título de Mestre em Estatística. Campinas, S.P.

[6] Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of Royal Statistical Society, Series B 39*, 1-38.

[7] Dixon, W.J. (editor) (1983). *BMDP statistical software.* Berkeley. University of California Press.

[8] Forthofer, R.N., Koch, G.G. (1973). An analysis for compounded functions of categorical data. *Biometrics 28*, 143-157.

[9] Fuchs, C. (1982). Maximum likelihood estimation and model selection in contingency tables with missing data. *Journal of the American Statistical Association 77*, 270-278.

[10] Grizzle, J.E., Starmer, C.F., Koch, G.G. (1969). Analysis of categorical data by linear models. *Biometrics 25*, 489-504.

[11] Hocking, R.R., Oxspring, H.H. (1971). Maximum likelihood estimation with incomplete multinomial data. *Journal of the American Statistical Association 66*, 65-70.

[12] Jennrich, R.I., Schluchter, M.D. (1986). Unbalanced repeated measures models with structured covariance matrices. *Biometrics 42*, 805-820.

[13] Koch, G.G., Elashoff, J.D., Amara, I.A. (1985). Repeated measurements studies: design and analysis. *Encyclopedia of Statistical Sciences*. N.L. Johnson and S. Kotz (eds.). New York. John Wiley & Sons, Inc.

[14] Koch, G.G., Imrey, P.B., Reinfurt, D.W. (1972). Linear model analysis of categorical data with incomplete response vectors. *Biometrics 28*, 663-692.

[15] 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 measurements of categorical data. *Biometrics 33*, 133-158.

[16] Lipsitz, S.R., Laird, N.M., Harrington, D.P. (1994). Weighted least squares analysis of repeated categorical measurements with outcomes subject to nonresponse. *Biometrics 50*, 11-24.

[17] Little, R.J.A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association 83*, 1198-1202.

[18] Little, R.J.A., Rubin, D.B. (1987). *Statistical analysis with missing data*. New York. John Wiley & Sons, Inc, 278p.

[19] Louis, I.A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B 44*, 226-233.

[20] Mood, A.M., Graybill, F.A., Boes, D.C. (1974). *Introduction to the theory of statistics*, 3ª edição. McGraw-Hill, 564p.

[21] Neyman, J. (1949). Contribution to the theory of the χ^2 test. *Proceedings of the Berkeley symposium on mathematical statistics and probability.* Berkeley and Los Angeles: University of California Press, 239-272.

[22] Rubin, D.B. (1974). Characterizing the estimation of parameters in incomplete data problems. *Journal of the American Statistical Association 69*, 467-474.

[23] Rubin, D.B. (1976). Inference and missing data. *Biometrika 63*, 581-592.

[24] Searle, S.R. (1971). *Linear Models.* New York. John Wiley & Sons, Inc.532p.

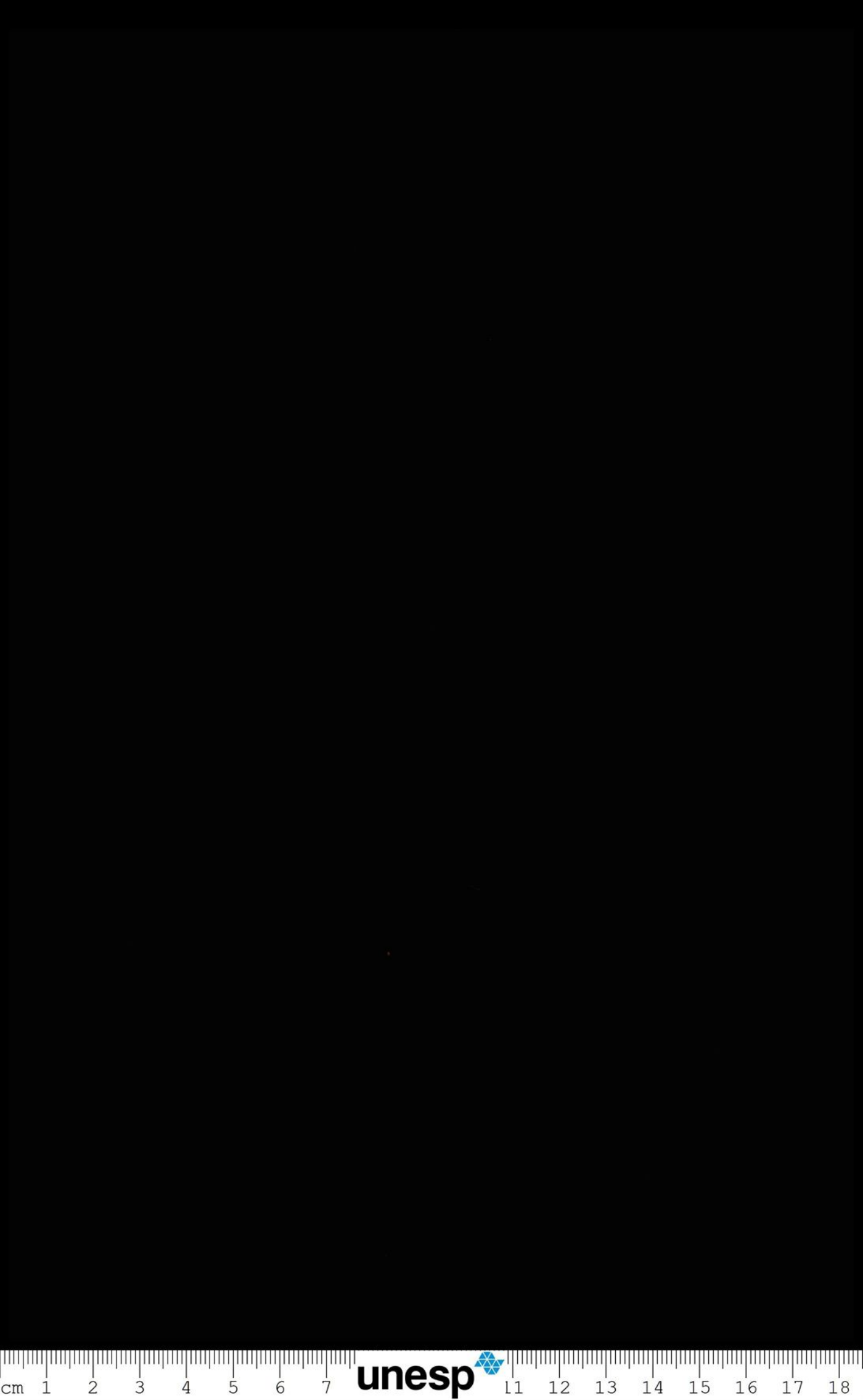
[25] Simon, G.A., Simonoff, J.S. (1986). Diagnostic plots for missing data in least squares regression. *Journal of the American Statistical Association 81*, 501-509.

[26] Stanish, W.M., Gillings, D.B., Koch, G.G. (1978). An aplication of multivariate ratio methods for the analysis of a longitudinal clinical trial with missing data. *Biometris 34*, 305-317.

[27] Stuart, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika 42*, 412-416.

[28] Woolson, R.F., Clarke, W.R. (1984). Analysis of categorical incomplete longitudinal data. *Journal of Royal Statistical Society, Series A 147*, 87-99.

[29] Wald, A. (1943). Tests of statistical hypotheses concerning general parameters when the number of observations in large. *Transactions of the American Mathematical Society 54*, 426-482.


cm 1 2 3 4 5 6 7 unesp 11 12 13 14 15 16 17 18