

ANÁLISE DE DADOS CATEGORIZADOS COM OMISSÃO

Frederico Zanqueta Poletto

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA



ANÁLISE DE DADOS CATEGORIZADOS COM OMISSÃO

Frederico Zanqueta Poletto

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

Área de Concentração: **Estatística**

Orientador: **Prof. Dr. Julio da Motta Singer**

*Durante a elaboração deste trabalho o autor
recebeu apoio financeiro do CNPq*

São Paulo, agosto de 2006

ANÁLISE DE DADOS CATEGORIZADOS COM OMISSÃO

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Frederico Zanqueta Poletto, e aprovada pela Comissão Julgadora.

São Paulo, 30 de agosto de 2006.

Comissão Julgadora:

- Prof. Dr. Julio da Motta Singer (orientador) – IME, USP
- Prof. Dr. Carlos Alberto de Bragança Pereira – IME, USP
- Prof. Dr. Carlos Daniel Mimoso Paulino – IST, Universidade Técnica de Lisboa

“Resista um pouco mais, mesmo que as feridas latejem e que sua coragem esteja cochilando.

Resista mais um minuto e será fácil resistir aos demais.

Resista mais um instante, mesmo que a derrota seja um ímã, mesmo que a desilusão caminhe em sua direção.

Resista mais um pouco, mesmo que os invejosos digam para você parar, mesmo que sua esperança esteja no fim.

Resista mais um momento, mesmo que você não possa avistar ainda a linha de chegada, mesmo que as inseguranças brinquem de roda à sua volta.

Resista um pouco mais, mesmo que a sua vida esteja sendo pesada como a consciência dos insensatos e você se sinta indefeso como um pássaro de asas quebradas.

Resista, porque o último instante da madrugada é sempre aquele que puxa a manhã pelo braço e essa manhã bonita, ensolarada, sem algemas, nascerá para você em breve, desde que você resista.

Resista, porque estamos sentados na arquibancada do tempo, torcendo ansiosos para que você vença e ganhe de Deus o troféu que você merece: a felicidade!”

(autor desconhecido)

*Aos meus pais Dagmar e Sérgio,
à minha avó Aurora,
aos meus irmãos Natália e Guilherme
e à minha namorada Lillian*

Resumo

Neste trabalho aborda-se aspectos teóricos, computacionais e aplicados de análises clássicas de dados categorizados com omissão. Uma revisão da literatura é apresentada enquanto se introduz os mecanismos de omissão, mostrando suas características e implicações nas inferências de interesse por meio de um exemplo considerando duas variáveis respostas dicotômicas e estudos de simulação. Amplia-se a modelagem descrita em Paulino (1991, *Brazilian Journal of Probability and Statistics* **5**, 1-42) da distribuição multinomial para o produto de multinomiais para possibilitar a inclusão de variáveis explicativas na análise. Os resultados são desenvolvidos em formulação matricial adequada para a implementação computacional, que é realizada com a construção de uma biblioteca para o ambiente estatístico R, a qual é disponibilizada para facilitar o traçado das inferências descritas nesta dissertação. A aplicação da teoria é ilustrada por meio de cinco exemplos de características diversas, uma vez que se ajusta modelos estruturais lineares (homogeneidade marginal), log-lineares (independência, razão de chances adjacentes comum) e funcionais lineares (*kappa*, *kappa* ponderado, sensibilidade/especificidade, valor preditivo positivo/negativo) para as probabilidades de categorização. Os padrões de omissão também são variados, com omissões em uma ou duas variáveis, confundimento de células vizinhas, sem ou com subpopulações.

Palavras-chave: Dados categorizados e, faltantes, incompletos ou omissos; MAR, MCAR e MNAR; Mecanismo ignorável e não-ignorável; Modelos de seleção.

Abstract

We consider theoretical, computational and applied aspects of classical categorical data analyses with missingness. We present a literature review while introducing the missingness mechanisms, highlighting their characteristics and implications in the inferences of interest by means of an example involving two binary responses and simulation studies. We extend the multinomial modeling scenario described in Paulino (1991, *Brazilian Journal of Probability and Statistics* **5**, 1-42) to the product-multinomial setup to allow for the inclusion of explanatory variables. We develop the results in matrix formulation and implement the computational procedures via subroutines written under R statistical environment. We illustrate the application of the theory by means of five examples with different characteristics, fitting structural linear (marginal homogeneity), log-linear (independence, constant adjacent odds ratio) and functional linear models (*kappa*, weighted *kappa*, sensibility/specificity, positive/negative predictive value) for the marginal probabilities. The missingness patterns includes missingness in one or two variables, neighbor cells confounded, with or without explanatory variables.

Key words: Categorical and, missing or incomplete data; MAR, MCAR and MNAR; Ignorable and non-ignorable mechanism; Selection models.

Agradecimentos

Agradeço imensamente a orientação constantemente presente durante todo o desenvolvimento desta dissertação e os incansáveis ensinamentos dos professores Julio da Motta Singer e Carlos Daniel Mimoso Paulino. Tenho a felicidade de poder contar com conselhos do Julio desde 2001, quando fui seu monitor no curso de dados categorizados. Mesmo que naquele tempo não pudesse imaginar o tema que trabalharia futuramente no mestrado, tenho certeza que seus encaminhamentos e a sua disposição em sempre me ajudar contribuíram significativamente para a realização deste trabalho. Em março de 2004 tive muita sorte de (i) o Daniel estar visitando o departamento justamente no momento em que eu iniciava a busca pelo assunto da dissertação (até então, pretendia apenas que fosse algo relacionado a dados categorizados), (ii) o Julio ter me sugerido o tema em que o Daniel é especialista e (iii) o Daniel ter aceitado me orientar mesmo após descobrirmos que a orientação não poderia ser formalizada pelas regras do IME-USP, que não aceita co-orientação no mestrado. Lamento o fato de o IME adotar normas que não propiciam a interação entre pesquisadores, o que obviamente não favorece a produção científica. Se não fosse pela louvável atitude de o Daniel prosseguir com a orientação informalmente, sem dúvida alguma este trabalho seria extremamente prejudicado. O longo convívio resultou em amizades afabilíssimas, ainda mais porque tanto o Julio quanto o Daniel separaram perfeitamente as figuras de orientador e amigo. Além disso, não posso deixar de ressaltar os esforços que os dois fizeram em dar suporte à distância, quando necessário, e a hospitalidade atenciosa que o Daniel me ofereceu durante o mês que passei em Lisboa. Este tempo que pesquisei próximo dele, embora curto, foi fundamental para que eu compreendesse melhor o assunto no início do projeto. Sobretudo, sou grato às laboriosas e freqüentes leituras desta dissertação, sempre muito cuidadosas e repletas de sugestões, que ambos fizeram nesses quase dois anos e meio. Muitíssimo obrigado por tudo!!!

Agradeço ao professor Carlos Alberto de Bragança Pereira por estar sempre disponível

para discutir as mais variadas questões relacionadas à estatística e contribuir em mais uma etapa da minha formação.

Sou grato à professora Lúcia Pereira Barroso pelos auxílios prestados sempre que precisei e pela orientação em minha iniciação científica sobre imputação de dados num problema específico. Apesar de os assuntos da iniciação científica e do mestrado estarem relacionados por (puro?) acaso, o fato de ter pensado previamente nos problemas associados aos dados omissos certamente contribuiu para um maior amadurecimento.

Agradeço a todos os professores e funcionários do Departamento de Estatística do IME–USP que de alguma forma contribuíram para este trabalho. Em especial, destaco a Elaine Fischer Bosko e os professores que tive o prazer de manter um contato mais próximo: Antonio Carlos Pedroso de Lima, Carmen Diva Saldiva de André, Clélia Maria de Castro Tolói, Denise Aparecida Botter, Elisabeti Kira, Gilberto Alvarenga Paula, Heleno Bolfarine, Luis Gustavo Esteves, Márcia D’Elia Branco, Marcos Nascimento Magalhães, Nelson Ithiro Tanaka, Pedro Alberto Morettin, Rinaldo Artes, Sérgio Wechsler, Silvia Lopes de Paula Ferrari, Vladimir Belitsky e Wagner de Souza Borges.

Agradeço aos professores Paul S. Clarke e Stuart G. Baker pelas solicitações nos esclarecimentos acerca de seus artigos e ao professor Geert Molenberghs por, além disso, estar sempre disponível para me ajudar e ter feito uma cuidadosa revisão de um artigo originado do Exemplo 5 e que acabou refletida na dissertação.

Sou grato aos professores Carlos Daniel Mimoso Paulino, Giovani Loiola Silva, Julio da Motta Singer, Paulo José de Jesus Soares e Viviana Giampaoli por terem divulgado as rotinas computacionais, que desenvolvi para o ambiente estatístico R, em cursos de análise de dados categorizados que ministraram na graduação e pós-graduação do IME–USP e do Instituto Superior Técnico da Universidade Técnica de Lisboa (IST–UTL) nos anos de 2005 e 2006, possibilitando que as rotinas fossem mais testadas e, por conseguinte, aprimoradas. Agradeço aos alunos destes cursos pelas sugestões, críticas e comentários.

Agradeço ao Nuno Sepúlveda e ao professor Paulo José de Jesus Soares por seus amáveis amparos quando estive em Lisboa. Sou grato também ao Nuno por ter lido parte de uma versão preliminar do Capítulo 1 e ter tecido comentários construtivos que incorporei dentro do possível. Agradeço também ao professor Paulo por sua cautelosa revisão de um texto que preparei como documentação preliminar das rotinas desenvolvidas para o R e pelas sugestões de aperfeiçoamento.

Sou grato ao Elias Teixeira Krainski, Fernando Henrique Ferraz Pereira da Rosa e aos professores Francisco José de Azevêdo Cysneiros, Laura A. Thompson e Paulo Justiniano Ribeiro Júnior por gentilmente terem me auxiliado com esclarecimentos de dúvidas sobre o R.

Agradeço à Simone da Silva Aranha do Centro de Estudos de Opinião Pública (CESOP) da UNICAMP e à Paula Miamoto Yamakawa e Márcia Cavallari Nunes do IBOPE pelos auxílios fundamentais no processo de obtenção dos dados referentes às pesquisas de intenção de voto.

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro, sem o qual não poderia ter me dedicado com afinco durante os primeiros dois anos. Sou grato ao IME–USP e agências de fomento pelos auxílios fornecidos para idas a simpósios, reuniões, escolas, conferências, encontros e *workshops*, tanto nos que compareci apenas como ouvinte, quanto, principalmente, nos que apresentei comunicações relacionadas a este trabalho: (i) pôster “Análise de dados categorizados omissos segundo uma abordagem freqüentista em dois estágios” no 16º Simpósio Nacional de Probabilidade e Estatística (SINAPE) em julho de 2004, (ii) pôster “Análise de dados categorizados com omissão: aspectos computacionais” na 9ª Escola de Modelos de Regressão em fevereiro de 2005 e (iii) comunicação oral “*Comparing diagnostic tests with missing data*” no 17º SINAPE em julho de 2006. Estas apresentações trouxeram contribuições muito importantes para aprimorar as partes correspondentes da dissertação. Agradeço muito ao Centro de Matemática (CEMAT) do IST–UTL por ter custeado gastos durante a viagem a Lisboa, sem o qual eu não poderia ter usufruído dessa oportunidade fantástica. Sou grato ainda ao IME por disponibilizar equipamentos (computacionais) e meios de reprodução gráfica (*e.g.*, impressões e cotas para fotocópias) sem custos, dentre outras facilidades que ajudaram a traçar o meu caminho.

Agradeço muito a meus pais, irmãos e à minha avó por terem assegurado as melhores condições possíveis para compor meu crescimento intelectual e profissional, pelos ensinamentos, amor e formação de meu caráter.

Sou grato à minha namorada Lílian Natis pelos encorajamentos, apoios, ajudas (estatísticas e não-estatísticas), companheirismo, amor e por iluminar meus dias.

Agradeço a meus sogros pelo acolhimento, carinho e suporte.

Sou grato a Deus por sempre me guiar, proteger e dar força para superar os obstáculos diários.

Agradeço aos colegas de trabalho e amigos que formei durante os curtos períodos que trabalhei no Citibank e na Credicard, pois, sem dúvida alguma, aprendi muito e tive excelentes

oportunidades de fazer conexões da teoria com a prática, isso ainda sem mencionar as diversas discussões sobre estatística que travamos nos dias (e noites). Em especial, sou grato ao Ricardo Tadashi Takeyama, Hélio Arizono, Marcelo Leal Santos, Darren Stewart, Gustavo Cruz, Heloisa Dias de Carvalho, Érica Elisa Gumiero, Carlos Alberto Auricchio Júnior, Tatiana Salomão Miamoto, Karina Pretto, Guilherme Perez, Jacqueline Sant'Eufemia David, Augusto César Giovanetti de Andrade, Igor Luiz Quidim, Mariana Mizutani Ribeiro, Salete Doniani e Roberta Duarte. Agradeço ao Edson Roberto da Silva e à Roseli Garcia pela oportunidade de continuar a freqüentar as reuniões do Grupo Foco da Associação Comercial de São Paulo, que é uma iniciativa interessante de interação de conhecimento entre o mercado e a academia, mesmo após parar de trabalhar com análises de risco de crédito.

Sou grato a todos os meus parentes, amigos e colegas que de alguma forma contribuíram para a realização desse projeto, como, por exemplo, com discussões sobre estatística, conversas várias, momentos de lazer, viagens para eventos relacionados à estatística ou para passear, trocas de mensagens eletrônicas, palavras de encorajamento, torcidas, telefonemas longos, ligações curtas, enfim, tudo o que me ajudou a tornar essa trajetória mais fácil e agradável. As amizades foram originadas em diversos âmbitos, como ensino secundário, graduação, pós-graduação, esportes, trabalhos (*e.g.*, Núcleo de Estudos da Violência), *etc.* Alguns já foram mencionados anteriormente e não serão repetidos. Adicionalmente, mas não apenas, destaco: Alessandra Montini, Alice Sobral Singer, André Azevedo, André Luiz Silva Samartini, André Rodrigo Sanchez, Ângela Tavares Paes, Caio Hiroshi Oshiro, Caio Lucidius Naberezny Azevedo, Camila Poplawski, Carine Savalli Redígolo, Carlos Eduardo Crespo, Cátia Petri, Cátia Yumi Nagae, César Torres, Cléber da Costa Figueiredo, Cristian Bayes, Cristina Brech, Daniela Carine Ramirez, Daniela Soares Costa, Daniel Kamo, Daniel Kashiwamura Scheffer, Danilo Clemente Coelho, Dario Gabai Júnior, Delhi Teresa Paiva Salinas, Démerson André Polli, Dóris Satie Maruyama Fontes, Edélcio Gonçalves de Souza, Edijane Paredes Garcia, Eduardo de Arruda Issei, Eduardo Manoel Brito, Élen Natis, Elias Chaibub, Émerson Almeida, Emilene Parlato, Estevão Freitas de Souza, Eurico de Mello Braga Filho, Fábio Esteves Nogueira, Felipe Osório, Felipe Villarino Prieto, Fernando Boza Costa Barra, Fernando Correa Lima, Fernando Frota Redígolo, Fernando Lemes, Flávia Carpinetti Pinto, Flávio A. Moraes, Flávio Vargas, Francisco Marcelo Monteiro da Rocha, Gianni Yanaguibashi, Gilberto da Silva Matos, Gisela Tunes da Silva, Iesus Carvalho Diniz, Iracema Hiroko Iramina Arashiro, João Fernando Serrajordia Rocha Mello, João Paulo dos Santos Pacífico, João Ricardo Sato, Joelmir Feliciano Barcellos, José Roberto Villar Perez, Jose Santos Romeo Nunez, Juscelino Jun Kang, Juvêncio Santos

Nobre, Karina Bezerra de Figueiredo, Karine Lage Martins, Leandro Fiorini Aurichi, Levindo David dos Santos, Liliam Pereira de Lima, Lincoln Cardoso Yamasaki, Livia Borges, Lourdes Contreras Montenegro, Luís Gustavo do Amaral Vinha, Luiz Fernando Freitas Pereira, Marcel Frederico de Lima Taga, Marcelo Batista Nery, Marcelo Hiroshi Ogava, Márcio Spera Poletto, Marco Aurélio Hirata, Marco César dos Santos Barbosa, Marcos Antônio Vincenzi, Marcos Santos de Oliveira, Maria Kelly Venezuela, Maria Paula Zanardi Chicarino Rosa, Mariana Curi, Marina Andretta, Mauro Sznelwar, Mayra Ivanoff Lora, Michelli Karinne Barros, Milena Reis, Nancy Cardia, Olímpio Ribeiro da Fonseca Neto, Patrícia L. Espinheira, Patrícia Pelegrini, Paulo de Tarso Marques Rosa, Raquel da Cunha Valle, Raydonal Ospina Martínez, Regina Albanese Pose, Regina Poletto, Rejane Augusta de Oliveira Figueiredo, Renata Aguemí, Ricardo Olivares de Magalhães, Rodrigo Andrade Tavares, Rogério Ruscitto do Prado, Rosana Francisco Alves, Sérgio Adorno, Sérgio Ricardo Martins, Silvina San Martino, Silvio Watanabe, Simone Curti, Stela Valeska Kretchetoff, Sumaia Abdel Latif, Susana Miyuki Okaze Tomazela, Tatiana Melhado Terabayashi, Ulisses Umbelino dos Anjos, Valéria Troncoso Baltar e Willy Patané Nielsen.

Apesar de todas as assistências recebidas, quaisquer erros e imprecisões que eventualmente tenham permanecido nesta dissertação, obviamente, são de minha inteira responsabilidade.

FREDERICO ZANQUETA POLETO

São Paulo, setembro de 2006

Conteúdo

Resumo	ix
Abstract	xi
Agradecimentos	xiii
1 Introdução	1
1.1 Um modelo probabilístico para duas variáveis dicotômicas	4
1.2 Mecanismos de omissão e modelos estruturais associados	7
1.2.1 Omissão aleatória	7
1.2.2 Omissão completamente aleatória	15
1.2.3 Omissão informativa	19
1.3 Seleção da estrutura para o mecanismo de omissão	40
1.4 Exemplo de análise	44
1.5 Outras abordagens	49
2 Formulação geral	53
2.1 Descrição do problema e notação	54
2.2 Modelo probabilístico e mecanismos de omissão	60
2.3 Estimação das probabilidades de categorização para modelos estruturais saturados	63
2.3.1 Estimação por máxima verossimilhança para modelos estruturais saturados sob os mecanismos MAR e MCAR	64
2.3.2 Estimação por mínimos quadrados generalizados para modelos estruturais saturados sob o mecanismo MCAR	69
	xix

2.4	Estimação das probabilidades de categorização para modelos estruturais não saturados	71
2.4.1	Exemplos de modelos estruturais	71
2.4.2	Estimação por máxima verossimilhança para modelos estruturais não saturados sob os mecanismos MAR e MCAR	78
2.4.3	Estimação por mínimos quadrados generalizados para modelos estruturais não saturados sob os mecanismos MAR, MCAR e MNAR	87
3	Aspectos práticos	91
3.1	Implementação da biblioteca para o ambiente estatístico R	91
3.2	Análises de dados	94
4	Discussão	137
A	Detalhes da obtenção de algumas expressões	141
A.1	Expressão (2.10)	141
A.2	Expressões (2.12) e (2.14)	142
A.3	Expressão (2.17)	144
A.4	Expressão (2.21)	144
A.5	Expressões (2.73) e (2.74)	145
B	Alocações de casos extremos	147
B.1	Exemplo 2	147
B.2	Exemplo 3	149
	Referências bibliográficas	151
	Índice de autores	169
	Índice de assuntos	175

“(Colonel Ross) *Is there any other point to which you would wish to draw my attention?*
(Holmes) *To the curious incident of the dog in the night-time.*
(Ross) *The dog did nothing in the night-time.*
That was the curious incident!, remarked Sherlock Holmes.”
Dawid e Dickey (1977)

CAPÍTULO 1

Introdução

Dados categorizados são freqüentes na prática e, no caso de estarem completos, sua análise sob os mais diferentes modelos tem sido estudada por inúmeros autores, dentre os quais destaca-se Bishop, Fienberg e Holland (1975), Forthofer e Lehnen (1981), Agresti (2002), Fleiss, Levin e Paik (2003) e Paulino e Singer (2006). A metodologia inferencial empregada para esse propósito encontra-se implementada em vários ambientes computacionais estatísticos (SAS, R, S-Plus, SPSS, Minitab, Statistica, *etc.*). No entanto, é comum encontrar problemas em que algumas pessoas não cumpriram uma ou mais diretrizes do protocolo do estudo (*non-compliance*, também interpretado como recusa em participação do estudo), abandonaram-no (*dropout*) durante sua realização ou não responderam a certas questões (*non-response*); ou uma parte do banco de dados está faltando (*missing*) por outro motivo qualquer. Nestes casos, as respostas em algumas variáveis para uma parte das unidades experimentais não são observadas e então, diz-se que o conjunto de dados obtido tem **omissão**. Um exemplo é exposto a seguir.

Exemplo 1 (Baker, Rosenberger e DerSimonian, 1992)

Em um estudo prospectivo do Departamento de Saúde dos Estados Unidos, mães grávidas, fumantes ou não, foram acompanhadas até ao parto e os recém-nascidos classificados conforme seu peso fosse < 2.5 kg ou ≥ 2.5 kg. O objetivo do estudo foi avaliar a associação entre o hábito de fumo da mãe e o peso do recém-nascido. Os dados observados estão expostos na Tabela 1.1.

Por razões desconhecidas, de 57 061 pares de mães/recém-nascidos, apenas 93% foram completamente categorizados. Não há informação sobre o peso do recém-nascido, hábito de fumo da mãe ou ambos para, respectivamente, 4%, 1% e 2% da amostra. Algumas especulações possíveis para as omissões são:

1. Problemas no armazenamento dos dados.
2. Mães fumantes se negariam a informar o hábito de fumo mais do que as não-fumantes.
3. Recém-nascidos com pesos menores poderiam ter complicações que impediriam sua pesagem no instante desejado.

Tabela 1.1: Frequências observadas

Mãe	Peso do recém-nascido (kg)		
	< 2.5	≥ 2.5	omisso
fumante			
sim	4 512	21 009	1 049
não	3 394	24 132	1 135
omisso	142	464	1 224

■

A literatura sobre a análise de dados com omissão é menos abundante apesar de estar crescendo consideravelmente. Além disso, métodos desenvolvidos para esse fim não estão computacionalmente implementados com a mesma abrangência que aqueles talhados para dados completos.

A abordagem mais simples e fácil para esses casos consiste em ignorar os dados com omissão e analisar apenas os dados completamente categorizados da maneira usual, supondo que constituem uma amostra aleatória da população de interesse. Essa prática é conhecida como análise de casos completos (*complete case analysis*) e será referida por **ACC**.

Blumenthal (1968) aparentemente foi o primeiro a considerar um processo de categorização parcial, embora o fizesse num contexto específico. Hocking e Oxspring (1971, 1974) consideraram o caso em que algumas unidades amostrais são completamente e outras parcialmente categorizadas, supondo que isto teria sido fixado pelo planejamento amostral. As equações de verossimilhança apresentadas sugerem um processo iterativo para a obtenção das estimativas de máxima verossimilhança e as condições que garantem sua convergência foram descritas em Hartley e Hocking (1971). Chen e Fienberg (1974) generalizaram os trabalhos de Blumenthal (1968) e Hocking e Oxspring (1971) obtendo estimadores de máxima verossimilhança sob a hipótese de independência e introduzindo modelos para o processo de omissão, mas permitindo que estes dependam apenas do que foi observado. Nessas mesmas condições, Chen e

Fienberg (1976) apresentaram um processo iterativo para o ajuste de modelos log-lineares. Haber, Chen e Williamson (1991) e Williamson e Haber (1994) realizaram outras análises de dados categorizados com omissão na mesma linha de Chen e Fienberg (1974).

Muitas tentativas foram feitas para se definir um procedimento geral para a obtenção de estimadores de máxima verossimilhança aplicável a qualquer padrão de dados omissos, tanto para respostas contínuas como categorizadas, dentre as quais destacam-se Woodbury (1971), Orchard e Woodbury (1972) e Sundberg (1974, 1976). Dempster, Laird e Rubin (1977) esclareceram, unificaram e generalizaram esses métodos chamando-os de EM (Esperança-Maximização). Por meio desse trabalho, pode-se perceber que os métodos de Hartley e Hocking (1971) e Chen e Fienberg (1976) são aplicações do algoritmo EM.

No contexto mais específico de dados categorizados com omissão, Choi e Stablein (1982) propuseram testes de igualdade de duas proporções binomiais; Shih (1987) apresentou um teste de razão de verossimilhanças para homogeneidade marginal em tabelas $r \times r$; Lipsitz e Fitzmaurice (1996) propuseram testes para a hipótese de independência em tabelas de contingência $r \times c$ com variáveis nominais ou ordinais e Tang e Tang (2004) desenvolveram testes exatos para comparar proporções pareadas.

Todos os autores citados até aqui assumiram implícita ou explicitamente que o mecanismo de omissão não depende das categorias não-observadas. Rubin (1976) formalizou condições gerais que permitem a ignorabilidade do processo que gera os dados omissos e considerou casos em que o processo é não-ignorável, cujos detalhes serão discutidos na Seção 1.2. Outras situações serão tratadas posteriormente.

O objetivo desta dissertação é discutir teórica e computacionalmente técnicas de análise de dados categorizados com omissão. Neste capítulo introduz-se conceitos de dados categorizados com omissão, apresentando um modelo probabilístico para o caso em que há apenas duas variáveis respostas dicotômicas na Seção 1.1, estruturas para os mecanismos de omissão na Seção 1.2, uma discussão sobre a seleção da estrutura para o mecanismo de omissão na Seção 1.3, um exemplo de análise na Seção 1.4 e outras abordagens na Seção 1.5. No decorrer da Seção 1.2 mostra-se que a ACC implica, em geral, perda de eficiência dos estimadores (erros padrões maiores) e/ou vieses nas inferências de interesse, dependendo do mecanismo de omissão. No Capítulo 2 estende-se o modelo probabilístico para permitir a utilização de variáveis explicativas e mais categorias de respostas. Apresenta-se fórmulas matriciais gerais, que permitem a análise sob modelos lineares, log-lineares e funcionais lineares, para uma parte

dos mecanismos de omissão. No Capítulo 3 descreve-se a implementação computacional dessa teoria em uma biblioteca para o ambiente estatístico R (R Development Core Team, 2006) e ilustra-se a sua flexibilidade ao realizar análises de alguns conjuntos de dados.

1.1 Um modelo probabilístico para duas variáveis dicotômicas

Para facilitar a exposição dos conceitos, utiliza-se um caso particular em que se deseja classificar as respostas das unidades experimentais segundo duas variáveis binárias, Y_1 e Y_2 , que podem assumir os valores 1 e 2. Essas variáveis podem representar, por exemplo, a categorização de cada unidade amostral com relação a duas questões de interesse ou a mesma questão medida em duas ocasiões. Teoricamente, os seguintes cenários são possíveis: (1) classificação completa (em Y_1 e Y_2); (2) classificação em Y_1 e omissão em Y_2 ; (3) omissão em Y_1 e classificação em Y_2 e (4) omissão completa (em Y_1 e Y_2). Supõe-se aqui que foram observados todos esses cenários, conforme representação na Tabela 1.2. A Tabela 1.1 é um exemplo de um conjunto de dados com estas características.

Tabela 1.2: Cenários de omissão em tabelas 2×2

Y_1	Y_2		omisso
	1	2	
1	(1)		(2)
2	(1)		(2)
omisso	(3)		(4)

Apesar de o interesse inicial se concentrar apenas nas variáveis Y_1 e Y_2 , a ocorrência de unidades amostrais nos cenários de omissão (2), (3) e (4) sugere a utilização de uma terceira variável, W , com possíveis valores 1, 2, 3 e 4, representativa dos diferentes padrões de omissão.

Supõe-se que a distribuição do vetor aleatório (W, Y_1, Y_2) possui 16 parâmetros denotados por $\gamma_{tij} = P(W = t, Y_1 = i, Y_2 = j)$, $i, j = 1, 2$, $t = 1, 2, 3, 4$, sendo que apenas 15 são linearmente independentes, pois $\sum_{t=1}^4 \sum_{i=1}^2 \sum_{j=1}^2 \gamma_{tij} = 1$. Esses parâmetros estão expostos

na Tabela 1.3 e as freqüências observáveis, na Tabela 1.4¹. Comparando as Tabelas 1.3 e 1.4 nota-se que o vetor (W, Y_1, Y_2) não é completamente observável, uma vez que há apenas 9 freqüências observáveis, sendo que somente 8 são linearmente independentes, pois supõe-se que $\sum_{t=1}^4 \sum_{i=1}^2 \sum_{j=1}^2 n_{tij} = n_{+++}$ é fixado de antemão.

Tabela 1.3: Parâmetros $\gamma = \{\gamma_{tij}\}$ definidores da distribuição do vetor aleatório (W, Y_1, Y_2)

$Y_1 \setminus Y_2$	$W = 1 (Y_1^o, Y_2^o)$		$W = 2 (Y_1^o, Y_2^m)$		$W = 3 (Y_1^m, Y_2^o)$		$W = 4 (Y_1^m, Y_2^m)$	
	1	2	1	2	1	2	1	2
1	γ_{111}	γ_{112}	γ_{211}	γ_{212}	γ_{311}	γ_{312}	γ_{411}	γ_{412}
2	γ_{121}	γ_{122}	γ_{221}	γ_{222}	γ_{321}	γ_{322}	γ_{421}	γ_{422}

Tabela 1.4: Freqüências observáveis associadas ao vetor aleatório (W, Y_1, Y_2)

$Y_1 \setminus Y_2$	$W = 1 (Y_1^o, Y_2^o)$		$W = 2 (Y_1^o, Y_2^m)$			$W = 3 (Y_1^m, Y_2^o)$		$W = 4 (Y_1^m, Y_2^m)$		
	1	2	1	2	total	1	2	1	2	total
1	n_{111}	n_{112}	?	?	n_{21+}	?	?	?	?	
2	n_{121}	n_{122}	?	?	n_{22+}	?	?	?	?	
total						n_{3+1}	n_{3+2}			n_{4++}

? representa uma freqüência não-observável

Da Tabela 1.4 percebe-se que as freqüências observáveis são somas das freqüências associadas aos valores de (W, Y_1, Y_2) , com suposta distribuição multinomial² octodimensional cujos parâmetros são somas dos elementos de $\{\gamma_{tij}\}$. Ou seja, apesar de as freqüências n_{211} e n_{212} estarem, respectivamente, associadas aos parâmetros γ_{211} e γ_{212} , elas não são observáveis, sendo este o motivo da representação via interrogações na Tabela 1.4; no entanto, a soma das duas, n_{21+} , é observável e está associada ao parâmetro γ_{21+} . Isto evidencia a sobre-parametrização do modelo probabilístico e sugere a necessidade de modelos identificáveis, que, em geral, são formulados sob certas fatorações dos $\{\gamma_{tij}\}$ que se descreve a seguir.

As probabilidades conjuntas γ_{tij} podem ser escritas como o produto das probabilidades

¹Para esclarecer a notação utilizada, a substituição de algum índice pelo sinal “+” indica a soma dos valores sobre esse índice. Adicionalmente, Y_k^o representa que o resultado de Y_k é observado e Y_k^m indica que a classificação de Y_k está omissa (*missing*), $k = 1, 2$.

²Obviamente, a utilização do modelo multinomial está associada à suposição de que as n_{+++} unidades amostrais foram observadas independentemente e com as mesmas probabilidades $\{\gamma_{tij}\}$.

marginais de (Y_1, Y_2) , denotadas por θ_{ij} , pelas probabilidades condicionais de $W|Y_1, Y_2$, representadas por $\lambda_{t(ij)}$, ou seja,

$$\begin{aligned}\gamma_{tij} &= P(W = t, Y_1 = i, Y_2 = j) \\ &= P(Y_1 = i, Y_2 = j) P(W = t|Y_1 = i, Y_2 = j) = \theta_{ij} \lambda_{t(ij)}.\end{aligned}\quad (1.1)$$

Essa fatoração facilita a explicitação de modelos estruturais para as **probabilidades marginais de categorização**, de interesse primordial, e para as **probabilidades condicionais de omissão**³, de interesse secundário. Como as probabilidades condicionais de omissão $\{\lambda_{t(ij)}\}$ podem estar associadas a uma auto-seleção das unidades amostrais em algum dos padrões de omissão W , dado $Y_1 = i, Y_2 = j$, na literatura de dados omissos, os modelos estruturais propostos por meio da fatoração (1.1) são chamados de **modelos de seleção** (*selection models*).

Pode-se também escrever as probabilidades conjuntas γ_{tij} como o produto das probabilidades marginais dos padrões de omissão W , representadas por ϕ_t , pelas probabilidades condicionais de $Y_1, Y_2|W$, denotadas por $\eta_{ij(t)}$, isto é,

$$\gamma_{tij} = P(W = t) P(Y_1 = i, Y_2 = j|W = t) = \phi_t \eta_{ij(t)}.\quad (1.2)$$

A designação de **modelos de mistura de padrões** (*pattern-mixture models*) para essa fatoração é fundamentada no fato de as probabilidades marginais do processo de categorização, θ_{ij} , serem expressas como uma mistura das correspondentes probabilidades para os diversos padrões observados de omissão, ou seja,

$$\theta_{ij} = \sum_{t=1}^T \phi_t \eta_{ij(t)}.\quad (1.3)$$

Molenberghs, Michiels, Kenward e Diggle (1998) analisaram as duas fatorações, sugerindo que se a questão de interesse é formulada ignorando o mecanismo de omissão, ou seja, marginalmente, então os modelos de seleção parecem ser uma escolha natural; mas se o interesse for

³No cenário $W = 1$ as variáveis Y_1 e Y_2 são observadas, portanto as probabilidades $\{\lambda_{1(ij)}\}$ devem ser interpretadas como probabilidades de ausência de omissão para que a nomenclatura designada seja coerente. Note também que, devido às restrições $\sum_{t=1}^4 \lambda_{t(ij)} = 1, i, j = 1, 2$, pode-se obter unicamente as probabilidades condicionais de omissão, estabelecendo uma estrutura apenas para $\lambda_{t(ij)}, t = 2, 3, 4, i, j = 1, 2$ e tomando $\lambda_{1(ij)} = 1 - \lambda_{2(ij)} - \lambda_{3(ij)} - \lambda_{4(ij)}, i, j = 1, 2$. Isso demonstra que as probabilidades condicionais de ausência de omissão são funções das outras probabilidades condicionais de omissão, justificando a convenção utilizada.

no estudo do fenômeno sob cada padrão de omissão, os modelos de mistura de padrões podem ser preferidos. Estas recomendações justificam a adoção pelos modelos de seleção no restante deste trabalho e esclarecem a preferência por essa fatoração na literatura. Algumas exceções, com discussões sobre as duas modelagens, estão descritas em Glynn, Laird e Rubin (1986), Little (1993), Ekholm e Skinner (1998), Molenberghs, Michiels e Kenward (1998), Michiels, Molenberghs e Lipsitz (1999) e Birmingham e Fitzmaurice (2002).

No contexto da Tabela 1.4, pode-se assumir que o vetor $\mathbf{N} = (n_{111}, n_{112}, n_{121}, n_{122}, n_{21+}, n_{22+}, n_{3+1}, n_{3+2}, n_{4++})'$ com as frequências observáveis segue uma distribuição multinomial especificada pela função de verossimilhança

$$L(\{\theta_{ij}\}, \{\lambda_{t(ij)}\} | \mathbf{N}) \propto \prod_{i=1}^2 \prod_{j=1}^2 (\theta_{ij} \lambda_{1(ij)})^{n_{1ij}} \times \prod_{i=1}^2 (\theta_{i1} \lambda_{2(i1)} + \theta_{i2} \lambda_{2(i2)})^{n_{2i+}} \times \prod_{j=1}^2 (\theta_{1j} \lambda_{3(1j)} + \theta_{2j} \lambda_{3(2j)})^{n_{3+j}} \times \left(\sum_{i=1}^2 \sum_{j=1}^2 \theta_{ij} \lambda_{4(ij)} \right)^{n_{4++}}, \quad (1.4)$$

com $\sum_{i=1}^2 \sum_{j=1}^2 \theta_{ij} = 1$ e $\sum_{t=1}^4 \lambda_{t(ij)} = 1$, $i, j = 1, 2$. Note que o primeiro termo corresponde ao cenário $W = 1$, o segundo, ao cenário $W = 2$, e assim sucessivamente.

1.2 Mecanismos de omissão e modelos estruturais associados

Uma vez que o interesse se concentra nos $\{\theta_{ij}\}$, estruturas mais reduzidas para as probabilidades condicionais de omissão $\{\lambda_{t(ij)}\}$ são consideradas para tornar o modelo identificável. Os mecanismos de omissão originalmente descritos por Rubin (1976) são apresentados a seguir.

1.2.1 Omissão aleatória

O mecanismo conhecido por **omissão aleatória**, **omissão não-informativa** ou **MAR** (*missing at random*) supõe que as probabilidades condicionais de omissão dependem apenas do que é observado, o que é formalmente expresso por

$$\lambda_{1(ij)} = \alpha_{1(ij)}, \quad \lambda_{2(ij)} = \alpha_{2(i)}, \quad \lambda_{3(ij)} = \alpha_{3(j)}, \quad \lambda_{4(ij)} = \alpha_4. \quad (1.5)$$

A Tabela 1.5 ilustra a disposição dessas probabilidades entre as celas da tabela de contingência que se tem analisado. Utilizando as restrições $\sum_{t=1}^4 \lambda_{t(ij)} = 1$, obtém-se $\alpha_{1(ij)} = 1 - \alpha_{2(i)} - \alpha_{3(j)} - \alpha_4$, e assim, pode-se definir o mecanismo de omissão MAR apenas em função de 5 parâmetros englobados no vetor $\boldsymbol{\alpha} = (\alpha_{2(1)}, \alpha_{2(2)}, \alpha_{3(1)}, \alpha_{3(2)}, \alpha_4)'$. Tendo em conta as 3 probabilidades marginais de interesse (θ_{11} , θ_{12} e θ_{21} , uma vez que $\theta_{22} = 1 - \theta_{11} - \theta_{12} - \theta_{21}$) resumidas no vetor $\boldsymbol{\theta}$, o modelo probabilístico correspondente baseia-se em 8 parâmetros, sendo portanto, saturado.

Tabela 1.5: Probabilidades condicionais de omissão sob o mecanismo MAR

$Y_1 \setminus Y_2$	$W = 1 (Y_1^o, Y_2^o)$		$W = 2 (Y_1^o, Y_2^m)$		$W = 3 (Y_1^m, Y_2^o)$		$W = 4 (Y_1^m, Y_2^m)$	
	1	2	1	2	1	2	1	2
1	$\alpha_{1(11)}$	$\alpha_{1(12)}$	$\alpha_{2(1)}$	$\alpha_{2(1)}$	$\alpha_{3(1)}$	$\alpha_{3(2)}$	α_4	α_4
2	$\alpha_{1(21)}$	$\alpha_{1(22)}$	$\alpha_{2(2)}$	$\alpha_{2(2)}$	$\alpha_{3(1)}$	$\alpha_{3(2)}$	α_4	α_4

$\alpha_{1(ij)} = 1 - \alpha_{2(i)} - \alpha_{3(j)} - \alpha_4$

Apesar de não se atribuir explicitamente resultados para as respostas não observadas, isto é implicitamente realizado ao adotar qualquer estrutura para o mecanismo de omissão. Note, por exemplo, que sob o mecanismo MAR, a alocação das unidades amostrais n_{21+} nas celas ($W = 2, Y_1 = 1, Y_2 = 1$) e ($W = 2, Y_1 = 1, Y_2 = 2$) (veja a Tabela 1.4) pode ser especificada a partir de

$$\begin{aligned}
P(Y_2 = 1|Y_1 = 1, W = 2) &= \frac{P(W = 2, Y_1 = 1, Y_2 = 1)}{P(W = 2, Y_1 = 1)} = \frac{\gamma_{211}}{\gamma_{21+}} = \frac{\gamma_{211}}{\gamma_{211} + \gamma_{212}} \\
&= \frac{\theta_{11}\alpha_{2(1)}}{\theta_{11}\alpha_{2(1)} + \theta_{12}\alpha_{2(1)}} = \frac{\theta_{11}}{\theta_{11} + \theta_{12}} = \frac{\theta_{11}}{\theta_{1+}} = \frac{P(Y_1 = 1, Y_2 = 1)}{P(Y_1 = 1)} \\
&= P(Y_2 = 1|Y_1 = 1), \tag{1.6}
\end{aligned}$$

com lógica similar para $P(Y_2 = 2|Y_1 = 1, W = 2) = P(Y_2 = 2|Y_1 = 1)$. Esse resultado reflete o significado do termo omissão aleatória ou não-informativa, que embute a suposição de que as unidades amostrais com omissão estariam distribuídas da mesma forma como se não houvesse omissão. Isto não quer dizer que se possa ignorar as unidades amostrais com omissão em

alguma(s) variável(eis) fazendo uma ACC. Para esclarecer isto, note-se inicialmente que

$$\begin{aligned}
P(Y_1 = i, Y_2 = j | W = 1) &= \frac{P(W = 1, Y_1 = i, Y_2 = j)}{P(W = 1)} = \frac{\gamma_{1ij}}{\gamma_{1++}} = \frac{\gamma_{1ij}}{\sum_{i=1}^2 \sum_{j=1}^2 \gamma_{1ij}} \\
&= \frac{\theta_{ij} \alpha_{1(ij)}}{\sum_{i=1}^2 \sum_{j=1}^2 \theta_{ij} \alpha_{1(ij)}} = \frac{\theta_{ij} (1 - \alpha_{2(i)} - \alpha_{3(j)} - \alpha_4)}{\sum_{i=1}^2 \sum_{j=1}^2 \theta_{ij} (1 - \alpha_{2(i)} - \alpha_{3(j)} - \alpha_4)} \\
&= \frac{\theta_{ij} (1 - \alpha_{2(i)} - \alpha_{3(j)} - \alpha_4)}{1 - \alpha_{2(1)} \theta_{1+} - \alpha_{2(2)} \theta_{2+} - \alpha_{3(1)} \theta_{+1} - \alpha_{3(2)} \theta_{+2} - \alpha_4}. \quad (1.7)
\end{aligned}$$

De (1.7), nota-se que $P(Y_1 = i, Y_2 = j | W = 1) = \theta_{ij} = P(Y_1 = i, Y_2 = j)$ apenas se $\alpha_{2(1)} = \alpha_{2(2)}$ e $\alpha_{3(1)} = \alpha_{3(2)}$ — caso particular do mecanismo MAR, que será exposto na subseção seguinte⁴. Portanto, uma ACC sob a suposição do mecanismo MAR leva, em geral, a inferências enviesadas sobre θ . Isso é uma conseqüência de a tabela completamente categorizada ($W = 1$) não ser uma amostra aleatória da população de interesse, mas da população de que é possível obter (Y_1, Y_2) . Conseqüentemente, a distribuição marginal de Y_1 [Y_2] do cenário $W = 1$ difere da obtida em $W = 2$ [$W = 3$], por exemplo⁵.

A função de verossimilhança (1.4) sob o mecanismo MAR é dada por

$$\begin{aligned}
L(\theta, \alpha | \mathbf{N}; \text{MAR}) &\propto \prod_{i=1}^2 \prod_{j=1}^2 (\theta_{ij} \alpha_{1(ij)})^{n_{1ij}} \prod_{i=1}^2 (\theta_{i1} \alpha_{2(i)} + \theta_{i2} \alpha_{2(i)})^{n_{2i+}} \times \\
&\quad \prod_{j=1}^2 (\theta_{1j} \alpha_{3(j)} + \theta_{2j} \alpha_{3(j)})^{n_{3+j}} \left(\sum_{i=1}^2 \sum_{j=1}^2 \theta_{ij} \alpha_4 \right)^{n_{4++}}
\end{aligned}$$

⁴Contudo, dependendo da inferência de interesse e dos padrões de omissão, a ACC pode em alguns casos ser consistente mesmo sob um mecanismo MAR; veja um exemplo na nota de rodapé 7.

⁵Compare

$$P(Y_1 = i | W = 1) = \frac{\theta_{i+} (1 - \alpha_{2(i)} - \alpha_4) - \theta_{i1} \alpha_{3(1)} - \theta_{i2} \alpha_{3(2)}}{1 - \alpha_{2(1)} \theta_{1+} - \alpha_{2(2)} \theta_{2+} - \alpha_{3(1)} \theta_{+1} - \alpha_{3(2)} \theta_{+2} - \alpha_4}$$

e

$$P(Y_2 = j | W = 1) = \frac{\theta_{+j} (1 - \alpha_{3(j)} - \alpha_4) - \theta_{1j} \alpha_{2(1)} - \theta_{2j} \alpha_{2(2)}}{1 - \alpha_{2(1)} \theta_{1+} - \alpha_{2(2)} \theta_{2+} - \alpha_{3(1)} \theta_{+1} - \alpha_{3(2)} \theta_{+2} - \alpha_4}$$

com, respectivamente,

$$P(Y_1 = i | W = 2) = \frac{\theta_{i+} \alpha_{2(i)}}{\theta_{1+} \alpha_{2(1)} + \theta_{2+} \alpha_{2(2)}} \quad \text{e} \quad P(Y_2 = j | W = 3) = \frac{\theta_{+j} \alpha_{3(j)}}{\theta_{+1} \alpha_{3(1)} + \theta_{+2} \alpha_{3(2)}},$$

expressões obtidas com desenvolvimento análogo a (1.7).

$$\begin{aligned}
&= \prod_{i=1}^2 \prod_{j=1}^2 \theta_{ij}^{n_{1ij}} \prod_{i=1}^2 (\theta_{i1} + \theta_{i2})^{n_{2i+}} \prod_{j=1}^2 (\theta_{1j} + \theta_{2j})^{n_{3+j}} \times \\
&\quad \prod_{i=1}^2 \prod_{j=1}^2 \alpha_{1(ij)}^{n_{1ij}} \prod_{i=1}^2 \alpha_{2(i)}^{n_{2i+}} \prod_{j=1}^2 \alpha_{3(j)}^{n_{3+j}} \alpha_4^{n_{4++}} \\
&\equiv L_1(\boldsymbol{\theta} | \mathbf{N}) L_2(\boldsymbol{\alpha} | \mathbf{N}; \text{MAR}), \tag{1.8}
\end{aligned}$$

em que

$$L_1(\boldsymbol{\theta} | \mathbf{N}) = \prod_{i=1}^2 \prod_{j=1}^2 \theta_{ij}^{n_{1ij}} \prod_{i=1}^2 (\theta_{i1} + \theta_{i2})^{n_{2i+}} \prod_{j=1}^2 (\theta_{1j} + \theta_{2j})^{n_{3+j}}$$

e

$$L_2(\boldsymbol{\alpha} | \mathbf{N}; \text{MAR}) = \prod_{i=1}^2 \prod_{j=1}^2 \alpha_{1(ij)}^{n_{1ij}} \prod_{i=1}^2 \alpha_{2(i)}^{n_{2i+}} \prod_{j=1}^2 \alpha_{3(j)}^{n_{3+j}} \alpha_4^{n_{4++}}.$$

Note que as frequências observadas que não sofreram qualquer tipo de categorização, n_{4++} , não trazem qualquer informação para a estimação de $\boldsymbol{\theta}$. Tendo em vista a fatoração da verossimilhança numa parte concernente a $\boldsymbol{\theta}$, denotada por $L_1(\boldsymbol{\theta} | \mathbf{N})$, e noutra referente a $\boldsymbol{\alpha}$, representada por $L_2(\boldsymbol{\alpha} | \mathbf{N}; \text{MAR})$, se $\boldsymbol{\theta}$ e $\boldsymbol{\alpha}$ forem funcionalmente independentes (distintos, segundo Rubin, 1976) pode-se chegar às seguintes conclusões:

- Os processos de estimação por máxima verossimilhança de $\boldsymbol{\theta}$ e $\boldsymbol{\alpha}$ podem ser conduzidos separadamente. Portanto, existindo interesse apenas na estimação de $\boldsymbol{\theta}$, não há necessidade em se preocupar com $\boldsymbol{\alpha}$.
- As partes da matriz de informação observada referentes a $\boldsymbol{\theta}$ e a $\boldsymbol{\alpha}$ são diagonais em blocos e o bloco correspondente a $\boldsymbol{\theta}$ não depende de $\boldsymbol{\alpha}$. Assim, sob o ponto de vista de inferência sobre $\boldsymbol{\theta}$ baseada apenas na verossimilhança, o processo de omissão é ignorável. No entanto, como $E(n_{tij}) = n_{++++} \theta_{ij} \lambda_{t(ij)}(\boldsymbol{\alpha})$ depende claramente de $\boldsymbol{\alpha}$, a parte da matriz de informação de Fisher referente a $\boldsymbol{\theta}$ também depende de $\boldsymbol{\alpha}$ e o processo de omissão não é ignorável sob o ponto de vista de inferências frequentistas sobre $\boldsymbol{\theta}$.

Dizer que “o processo de omissão (não) é **ignorável**” significa que (não) se pode ignorar as probabilidades condicionais de omissão $\boldsymbol{\alpha}$ para fazer inferências sobre $\boldsymbol{\theta}$. Kenward e Molenberghs (1998) ressaltaram essas distinções sob os contextos frequentista e verossimilhancista, ilustrando analiticamente e por meio de simulações que o uso da matriz de informação de

Fisher ignorando-se o mecanismo de omissão MAR pode levar a estimadores viesados para os elementos da matriz de covariâncias dos estimadores de máxima verossimilhança (EMV) de $\boldsymbol{\theta}$.

Little (1982, 1983) estudou a definição de ignorabilidade apresentada por Rubin (1976) em inquéritos por amostragem. Heitjan e Rubin (1991) e Heitjan (1994) estenderam a definição de ignorabilidade para dados grosseiros (*coarse data*), que incluem dados arredondados, agrupados, censurados e omissos como casos especiais. Exemplos na área biomédica foram apresentados em Heitjan (1993). Heitjan (1997) explorou a relação entre ignorabilidade, suficiência e ancilaridade em dados grosseiros.

Em geral, a equação $\partial \ln L_1(\boldsymbol{\theta} | \mathbf{N}) / \partial \boldsymbol{\theta} = \mathbf{0}$, com $L_1(\boldsymbol{\theta} | \mathbf{N})$ definida em (1.8), não tem solução explícita, o que implica que os EMV de $\boldsymbol{\theta}$ devem ser obtidos por meio de métodos iterativos como Newton-Raphson, *scoring* de Fisher ou EM [McLachlan e Krishnan (1996) e Tanner (1996)]. Com o intuito de apresentar uma idéia intuitiva sobre os EMV na presença de dados categorizados com algum tipo de omissão, suponha uma situação similar, em que nenhuma unidade amostral foi omissa em Y_1 e observada em Y_2 (cenário $W = 3$). Isso caracteriza um **padrão de omissão monótono**⁶ (Rubin, 1974), que permite a obtenção de EMV em forma fechada. Substituindo

$$\theta_{ij} = P(Y_1 = i, Y_2 = j) = P(Y_1 = i) P(Y_2 = j | Y_1 = i) = \theta_{i+} \theta_{j(i)} \quad (1.9)$$

em $L_1(\boldsymbol{\theta} | \mathbf{N})$, eliminando os termos correspondentes ao padrão de omissão $W = 3$ e adicionando as restrições pertinentes, obtém-se a expressão

$$L_1(\boldsymbol{\theta} | \mathbf{N}, n_{3+1} = n_{3+2} = 0) = \theta_{1+}^{n_{11+} + n_{21+}} \theta_{2+}^{n_{12+} + n_{22+}} \theta_{1(1)}^{n_{111}} \theta_{2(1)}^{n_{112}} \theta_{1(2)}^{n_{121}} \theta_{2(2)}^{n_{122}}$$

que conduz aos EMV

$$\hat{\theta}_{i+}(\mathbf{n} | n_{3+1} = n_{3+2} = 0) = \frac{n_{1i+} + n_{2i+}}{n_{1++} + n_{2++}}, i = 1, 2, \quad \hat{\theta}_{j(i)}(\mathbf{n} | n_{3+1} = n_{3+2} = 0) = \frac{n_{1ij}}{n_{1i+}}, j, i = 1, 2$$

e, pela propriedade da invariância à

$$\hat{\theta}_{ij}(\mathbf{n} | n_{3+1} = n_{3+2} = 0) = \frac{n_{1i+} + n_{2i+}}{n_{1++} + n_{2++}} \times \frac{n_{1ij}}{n_{1i+}}, \quad i, j = 1, 2. \quad (1.10)$$

⁶No caso de três variáveis, Y_1 , Y_2 e Y_3 , o padrão monótono é obtido sob os seguintes cenários: (i) as três variáveis são observadas; (ii) há omissão apenas em Y_3 ; (iii) há omissão em Y_2 e Y_3 e (iv) há omissão nas três variáveis. Subconjuntos desses cenários também caracterizam padrões monótonos, mas a observação de cenários adicionais, não. O padrão de omissão monótono é útil principalmente em estudos longitudinais, quando pacientes que abandonam o estudo nunca mais são observados.

Note que o estimador da probabilidade marginal de $Y_1 = i$ utiliza a informação dos dois cenários de omissão e, na ausência de unidades amostrais pertencentes ao cenário $W = 2$, os estimadores definidos em (1.10) tornam-se iguais aos já conhecidos EMV de $\boldsymbol{\theta}$ considerando apenas as unidades amostrais da tabela completamente categorizada, ou seja,

$$\hat{\theta}_{ij}(\mathbf{n} | n_{21+} = n_{22+} = n_{3+1} = n_{3+2} = 0) = \frac{n_{1i+} + 0}{n_{1++} + 0} \times \frac{n_{1ij}}{n_{1i+}} = \frac{n_{1ij}}{n_{1++}}, \quad i, j = 1, 2.$$

Suponha que $(\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}) = (0.1, 0.3, 0.4, 0.2)$ e que foram observados apenas os dois primeiros cenários de omissão, $W = 1, 2$, sob o mecanismo MAR com $\alpha_{2(1)} = \alpha - \delta$ e $\alpha_{2(2)} = \alpha + \delta$. Portanto, α é a probabilidade média de se ter omissão em Y_2 e 2δ é a diferença entre as probabilidades de omissão em Y_2 quando $Y_1 = 2$ e quando $Y_1 = 1$. A Tabela 1.6 ilustra os valores de $P(Y_1 = i, Y_2 = j | W = 1)$ considerando as combinações de $\alpha = 0.3, 0.5, 0.7$ e $\delta = \pm 0.1, \pm 0.2$.

Tabela 1.6: Valores de $P(Y_1 = i, Y_2 = j | W = 1)$

δ	$\{\theta_{ij}\}$	α		
		0.3	0.5	0.7
-0.2	0.1	0.0676	0.0556	0.0294
	0.3	0.2027	0.1667	0.0882
	0.4	0.4865	0.5185	0.5882
	0.2	0.2432	0.2593	0.2941
-0.1	0.1	0.0833	0.0769	0.0625
	0.3	0.2500	0.2308	0.1875
	0.4	0.4444	0.4615	0.5000
	0.2	0.2222	0.2308	0.2500
0.1	0.1	0.1176	0.1250	0.1429
	0.3	0.3529	0.3750	0.4286
	0.4	0.3529	0.3333	0.2857
	0.2	0.1765	0.1667	0.1429
0.2	0.1	0.1364	0.1522	0.1923
	0.3	0.4091	0.4565	0.5769
	0.4	0.3030	0.2609	0.1538
	0.2	0.1515	0.1304	0.0769

Dos casos particulares da Tabela 1.6, conclui-se que o viés absoluto,

$$|P(Y_1 = i, Y_2 = j|W = 1) - \theta_{ij}|,$$

na estimação de $\{\theta_{ij}\}$ ao se ignorar as unidades amostrais com omissão em Y_2 é maior quando a probabilidade de omissão é mais elevada (α maior) e quando a diferença entre $\alpha_{2(1)}$ e $\alpha_{2(2)}$ é mais acentuada ($|\delta|$ maior). Note também que se o interesse fosse avaliar a simetria, enquanto para os $\{\theta_{ij}\}$ verdadeiros a diferença entre θ_{21} e θ_{12} é de 0.1, para o caso em que $\alpha = 0.3$ e $\delta = 0.1$ essa diferença seria nula e para $\alpha = 0.7$ e $\delta = 0.2$ ($\delta = -0.2$), a diferença seria de -0.4231 (0.5000). Isso demonstra que, dependendo das inferências de interesse⁷, realizar uma ACC pode fazer com que as conclusões sejam enviesadas.

Com o objetivo de avaliar a magnitude dos vieses ilustrados na Tabela 1.6 em amostras finitas, realizou-se um estudo de simulação gerando um milhão de réplicas de Monte Carlo de distribuições multinomiais de cada uma das combinações de $(\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}) = (0.1, 0.3, 0.4, 0.2)$ com $\alpha = 0.3, 0.5, \delta = 0.1, 0.2$ e $n_{+++} = 10, 20, 50, 100, 200, 500$. Foram calculadas as médias de todas as estimativas de $\{\theta_{ij}\}$ sob a ACC e sob o mecanismo MAR. Imaginava-se que para tamanhos de amostras “pequenos” (*e.g.*, $n_{+++} = 10, 20$) não seriam observados vieses tão grandes como os ilustrados na Tabela 1.6 ao ignorar as unidades com omissão em Y_2 ou que os estimadores utilizando essas unidades não fossem tão menos enviesados. No entanto, as médias de todas as estimativas obtidas sob o mecanismo MAR [sob a ACC] apresentaram uma diferença absoluta máxima em relação a $(0.1, 0.3, 0.4, 0.2)$ [valores enviesados correspondentes à combinação de α e δ da Tabela 1.6] menor que 0.0119, 0.0015 e 0.0002, respectivamente, para $n_{+++} = 10, 20$ e $n_{+++} \geq 50$, conforme se pode notar pelos valores apresentados na Tabela 1.7. Isso ilustra que mesmo para tamanhos de amostra pequenos, a utilização das unidades com omissão em alguma(s) variável(eis) sob a suposição do mecanismo MAR deva ser encarada como absolutamente necessária para não se fazerem inferências enviesadas sobre θ .

⁷Contudo, para os dois cenários de omissão considerados, as inferências sobre a associação entre Y_1 e Y_2 são consistentes conforme pode-se verificar pela razão de chances

$$\frac{P(Y_1 = 1, Y_2 = 1|W = 1)P(Y_1 = 2, Y_2 = 2|W = 1)}{P(Y_1 = 1, Y_2 = 2|W = 1)P(Y_1 = 2, Y_2 = 1|W = 1)} = \frac{\frac{\theta_{11}(1-\alpha_{2(1)})}{1-\alpha_{2(1)}\theta_{1+}-\alpha_{2(2)}\theta_{2+}} \times \frac{\theta_{22}(1-\alpha_{2(2)})}{1-\alpha_{2(1)}\theta_{1+}-\alpha_{2(2)}\theta_{2+}}}{\frac{\theta_{12}(1-\alpha_{2(1)})}{1-\alpha_{2(1)}\theta_{1+}-\alpha_{2(2)}\theta_{2+}} \times \frac{\theta_{21}(1-\alpha_{2(2)})}{1-\alpha_{2(1)}\theta_{1+}-\alpha_{2(2)}\theta_{2+}}} = \frac{\theta_{11}\theta_{22}}{\theta_{12}\theta_{21}}.$$

Em compensação, se o cenário $W = 3$ também for observado, em geral, quaisquer inferências serão inconsistentes.

Tabela 1.7: Estimativas de Monte Carlo para os valores esperados de estimadores dos $\{\theta_{ij}\}$

n_{+++}	$\{\theta_{ij}\}$	$\delta = 0.1$				$\delta = 0.2$			
		$\alpha = 0.3$		$\alpha = 0.5$		$\alpha = 0.3$		$\alpha = 0.5$	
		ACC	MAR	ACC	MAR	ACC	MAR	ACC	MAR
10	0.1	0.1177	0.1007	0.1254	0.1035	0.1364	0.1002	0.1525	0.1017
	0.3	0.3530	0.2995	0.3746	0.2964	0.4091	0.2999	0.4559	0.2985
	0.4	0.3532	0.3997	0.3330	0.3948	0.3029	0.3979	0.2609	0.3881
	0.2	0.1762	0.2002	0.1670	0.2053	0.1516	0.2020	0.1307	0.2116
20	0.1	0.1176	0.0999	0.1250	0.1002	0.1365	0.1001	0.1521	0.1000
	0.3	0.3531	0.3000	0.3752	0.3000	0.4090	0.2999	0.4566	0.3000
	0.4	0.3529	0.4000	0.3331	0.3995	0.3031	0.4001	0.2609	0.3985
	0.2	0.1765	0.2001	0.1666	0.2003	0.1514	0.1999	0.1304	0.2015
50	0.1	0.1178	0.1001	0.1251	0.1001	0.1363	0.1000	0.1522	0.1001
	0.3	0.3530	0.3000	0.3751	0.3000	0.4091	0.3000	0.4564	0.3000
	0.4	0.3529	0.4000	0.3333	0.4001	0.3030	0.4000	0.2609	0.4000
	0.2	0.1764	0.2000	0.1665	0.1999	0.1516	0.2001	0.1305	0.2000
100	0.1	0.1177	0.1001	0.1250	0.1000	0.1364	0.1000	0.1522	0.1000
	0.3	0.3529	0.3000	0.3749	0.2999	0.4089	0.2999	0.4566	0.3000
	0.4	0.3528	0.3999	0.3333	0.3999	0.3030	0.4000	0.2608	0.3999
	0.2	0.1765	0.2001	0.1668	0.2001	0.1516	0.2001	0.1304	0.2000
200	0.1	0.1176	0.1000	0.1250	0.1000	0.1363	0.1000	0.1521	0.1000
	0.3	0.3528	0.2999	0.3750	0.3000	0.4091	0.2999	0.4565	0.3000
	0.4	0.3530	0.4000	0.3333	0.3999	0.3030	0.4000	0.2609	0.4000
	0.2	0.1765	0.2000	0.1667	0.2000	0.1516	0.2001	0.1305	0.2000
500	0.1	0.1177	0.1000	0.1250	0.1000	0.1364	0.1000	0.1522	0.1000
	0.3	0.3529	0.3000	0.3750	0.3000	0.4091	0.3000	0.4565	0.3000
	0.4	0.3529	0.3999	0.3333	0.4000	0.3030	0.4000	0.2609	0.4000
	0.2	0.1765	0.2000	0.1667	0.2000	0.1515	0.2000	0.1305	0.2000

1.2.2 Omissão completamente aleatória

Um caso particular do mecanismo MAR é conhecido por **omissão completamente aleatória** ou **MCAR** (*missing completely at random*) e impõe uma única probabilidade condicional de omissão para cada cenário, *i.e.*,

$$\lambda_{t(ij)} = \alpha_t, \quad (1.11)$$

o que significa que as probabilidades condicionais de omissão não dependem da categoria a que cada unidade pertence, seja ela observada ou não. Essas probabilidades condicionais de omissão estão dispostas na Tabela 1.8. Com as restrições já descritas anteriormente, o modelo correspondente possui apenas 6 parâmetros (3 em α e 3 em θ).

Tabela 1.8: Probabilidades condicionais de omissão sob o mecanismo MCAR

$Y_1 \setminus Y_2$	$W = 1 (Y_1^o, Y_2^o)$		$W = 2 (Y_1^o, Y_2^m)$		$W = 3 (Y_1^m, Y_2^o)$		$W = 4 (Y_1^m, Y_2^m)$	
	1	2	1	2	1	2	1	2
1	α_1	α_1	α_2	α_2	α_3	α_3	α_4	α_4
2	α_1	α_1	α_2	α_2	α_3	α_3	α_4	α_4

$\alpha_1 = 1 - \alpha_2 - \alpha_3 - \alpha_4$

Impondo as restrições do mecanismo MCAR à função de verossimilhança (1.8) obtém-se

$$\begin{aligned} L(\boldsymbol{\theta}, \boldsymbol{\alpha} | \mathbf{N}; \text{MCAR}) &= \prod_{i=1}^2 \prod_{j=1}^2 \theta_{ij}^{n_{1ij}} \prod_{i=1}^2 (\theta_{i1} + \theta_{i2})^{n_{2i+}} \prod_{j=1}^2 (\theta_{1j} + \theta_{2j})^{n_{3+j}} \prod_{t=1}^4 \alpha_t^{n_{t++}} \\ &\equiv L_1(\boldsymbol{\theta} | \mathbf{N}) L_2(\boldsymbol{\alpha} | \mathbf{N}_{++}; \text{MCAR}), \end{aligned} \quad (1.12)$$

em que $\mathbf{N}_{++} = (n_{1++}, n_{2++}, n_{3++}, n_{4++})'$,

$$L_1(\boldsymbol{\theta} | \mathbf{N}) = \prod_{i=1}^2 \prod_{j=1}^2 \theta_{ij}^{n_{1ij}} \prod_{i=1}^2 (\theta_{i1} + \theta_{i2})^{n_{2i+}} \prod_{j=1}^2 (\theta_{1j} + \theta_{2j})^{n_{3+j}}$$

e

$$L_2(\boldsymbol{\alpha} | \mathbf{N}_{++}; \text{MCAR}) = \prod_{t=1}^4 \alpha_t^{n_{t++}}.$$

Nota-se que a verossimilhança previsivelmente se separa numa parte referente a $\boldsymbol{\theta}$ e noutra concernente a $\boldsymbol{\alpha}$. A parcela da verossimilhança referente a $\boldsymbol{\theta}$, representada por $L_1(\boldsymbol{\theta} | \mathbf{N})$, é

igual para os mecanismos MAR e MCAR. Desta forma, as inferências sobre θ baseadas apenas no fator $L_1(\theta | \mathbf{N})$ ⁸ produzem os mesmos resultados sob os dois mecanismos de omissão.

Adicionalmente, percebe-se em (1.12) que \mathbf{N}_{++} é uma estatística suficiente parcial (S-suficiente) para α ⁹, ou equivalentemente, ancilar parcial (S-ancilar) para θ (veja Lindsey, 1996). Portanto, pelo princípio generalizado da condicionalidade (Basu, 1977) pode-se analisar os dados com base na verossimilhança condicional $L_1(\theta | \mathbf{N})$ — núcleo de uma distribuição produto de multinomiais¹⁰: $M_4(n_{1++}, \theta)$, $Bin(n_{2++}, \theta_{1+})$ e $Bin(n_{3++}, \theta_{+1})$ —, que não inclui o parâmetro perturbador (*nuisance*) α , sem incorrer em perda de informação relevante sobre o parâmetro de interesse, θ . Conseqüentemente, sob o mecanismo MCAR, pode-se supor que o total de unidades obtido em cada cenário de omissão foi fixado pelo planejamento amostral e o processo de omissão é ignorável sob o ponto de vista das inferências freqüentistas e verossimilhancistas sobre θ .

Sendo o mecanismo MCAR um caso particular do MAR, o resultado (1.6) permanece válido. Adicionalmente, conclui-se que, sob a validade do mecanismo MCAR, as inferências sobre θ em ACC não serão enviesadas, pois

$$\begin{aligned} P(Y_1 = i, Y_2 = j | W = 1) &= \frac{P(W = 1, Y_1 = i, Y_2 = j)}{P(W = 1)} = \frac{\gamma_{1ij}}{\gamma_{1++}} \\ &= \frac{\gamma_{1ij}}{\sum_{i=1}^2 \sum_{j=1}^2 \gamma_{1ij}} = \frac{\theta_{ij}\alpha_1}{\sum_{i=1}^2 \sum_{j=1}^2 \theta_{ij}\alpha_1} = \theta_{ij} \\ &= P(Y_1 = i, Y_2 = j). \end{aligned} \tag{1.13}$$

Isto demonstra que a tabela completamente categorizada ($W = 1$) constitui uma amostra aleatória da população de interesse e a distribuição marginal de Y_1 [Y_2] do cenário $W = 2$ [$W = 3$] é a mesma da obtida em $W = 1$ ¹¹. Dados obtidos com esta característica foram chamados de *observed at random* (OAR) por Rubin (1976). O conjunto das suposições MAR e OAR refletem o mecanismo MCAR.

⁸Ou seja, as EMV, valores observados das estatísticas de teste de razão de verossimilhanças e, também, estimativas dos erros padrões e valores observados das estatísticas de testes de Wald e *score* de Rao que se utilizem da matriz de informação observada (ao invés da matriz de informação de Fisher).

⁹Suficiente para α para cada valor de θ (*i.e.*, suficiente específica para α) e ancilar para θ para cada valor de α (*i.e.*, ancilar específica para θ).

¹⁰ M_4 denota uma distribuição multinomial com 4 parâmetros, em que 3 são linearmente independentes devido à restrição natural. *Bin* representa uma distribuição binomial.

¹¹De (1.13), tem-se que $P(Y_1 = i | W = 1) = \theta_{i+}$ e $P(Y_2 = j | W = 1) = \theta_{+j}$. Com desenvolvimento análogo a (1.13), chega-se a $P(Y_1 = i | W = 2) = \theta_{i+}$ e $P(Y_2 = j | W = 3) = \theta_{+j}$.

Apesar de as inferências sobre $\boldsymbol{\theta}$ não serem enviesadas com a ACC, ignorar as unidades amostrais dos cenários em que há omissão em alguma(s) variável(eis) implica, em geral, erros padrões mais elevados para os estimadores de $\{\theta_{ij}\}$. A expressão analítica da matriz de covariâncias não ilustra facilmente esse resultado; no entanto, pode-se observar o ganho de informação em θ_{11} , por exemplo, a partir de

$$E\left(-\frac{\partial^2 \ln L_1(\boldsymbol{\theta} | \mathbf{N})}{\partial \theta_{11}^2}\right) = n_{1++} \left[\frac{1}{\theta_{11}} + \frac{1}{1 - \theta_{11} - \theta_{12} - \theta_{21}} \right] + \\ n_{2++} \left[\frac{1}{\theta_{11} + \theta_{12}} + \frac{1 - \theta_{11} - \theta_{12}}{(1 - \theta_{11} + \theta_{12})^2} \right] + \\ n_{3++} \left[\frac{1}{\theta_{11} + \theta_{21}} + \frac{1 - \theta_{11} - \theta_{21}}{(1 - \theta_{11} + \theta_{21})^2} \right], \quad (1.14)$$

notando que os termos de dentro de cada par de colchetes são positivos e o de dentro do primeiro par de colchetes é o que seria obtido com uma ACC.

Os erros padrões dos EMV utilizando apenas as unidades completamente categorizadas também foram comparados com os da estrutura MCAR em amostras finitas por meio de uma simulação. Considerando as probabilidades marginais de interesse $(\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}) = (0.1, 0.4, 0.3, 0.2)$ com apenas os dois primeiros cenários de omissão ($W = 1, 2$) sob o mecanismo MCAR com $\alpha_2 = 0.2, 0.5, 0.8$ e tamanhos de amostra $n_{+++} = 10, 20, 50, 100, 200, 500$, foram geradas um milhão de réplicas de Monte Carlo de distribuições multinomiais com cada uma das combinações de n_{+++} e α_2 com $\{\theta_{ij}\}$. Para cada combinação foram obtidas as estimativas dos $\{\theta_{ij}\}$ sob a ACC e sob o mecanismo MCAR. Os desvios padrões de todas as estimativas sob as duas abordagens estão apresentados na Tabela 1.9 e uma coluna contém a diferença percentual do desvio padrão das estimativas obtidas sob o mecanismo MCAR em relação àquele obtido com uma ACC. Essa coluna indica o quanto o erro padrão é menor quando se utilizam estimadores que incorporam todas as unidades observadas relativamente aos estimadores que ignoram as unidades com omissão em Y_2 e é referida como “ganho”.

O ganho é negativo apenas para o EMV de θ_{11} nas combinações de (n_{+++}, α_2) iguais a $(10, 0.2)$, $(20, 0.5)$ e $(50, 0.8)$, quando¹² $E(n_{111}) \leq 1$ e $E(n_{1ij}) \geq 1.5$, $(i, j) \neq (1, 1)$. O ganho é exagerado, comparativamente a ganhos de (n_{+++}, α_2) “vizinhos”, para todos os θ_{ij} nas

¹²Recordando-se (Seção 1.2.1) que $E(n_{tij}) = n_{+++} \theta_{ij} \lambda_{t(ij)}(\boldsymbol{\alpha})$, tem-se, portanto, para esta simulação,

$$E(n_{1ij}) = n_{+++} \theta_{ij} (1 - \alpha_2).$$

Tabela 1.9: Estimativas de Monte Carlo para os desvios padrões dos estimadores de máxima verossimilhança sob a ACC, sob o mecanismo MCAR e os ganhos percentuais nas diminuições dos desvios padrões sob o mecanismo MCAR em relação àquele obtido sob uma ACC

n_{+++}	$\{\theta_{ij}\}$	$\alpha_2 = 0.2$			$\alpha_2 = 0.5$			$\alpha_2 = 0.8$		
		ACC	MCAR	$1 - \frac{\text{MCAR}}{\text{ACC}}$	ACC	MCAR	$1 - \frac{\text{MCAR}}{\text{ACC}}$	ACC	MCAR	$1 - \frac{\text{MCAR}}{\text{ACC}}$
10	0.1	0.1076	0.1078	-0.2%	0.1437	0.1396	2.8%	0.2199	0.1705	22.5%
	0.4	0.1758	0.1637	6.9%	0.2345	0.1887	19.6%	0.3544	0.2101	40.7%
	0.3	0.1644	0.1581	3.8%	0.2191	0.1903	13.2%	0.3290	0.2115	35.7%
	0.2	0.1435	0.1413	1.6%	0.1914	0.1758	8.1%	0.2878	0.1992	30.8%
20	0.1	0.0755	0.0751	0.4%	0.0977	0.0980	-0.3%	0.1685	0.1515	10.1%
	0.4	0.1232	0.1146	6.9%	0.1596	0.1312	17.8%	0.2749	0.1772	35.5%
	0.3	0.1154	0.1107	4.1%	0.1491	0.1347	9.7%	0.2570	0.1895	26.3%
	0.2	0.1006	0.0986	2.0%	0.1304	0.1251	4.0%	0.2242	0.1820	18.8%
50	0.1	0.0475	0.0471	0.9%	0.0607	0.0596	1.8%	0.0994	0.1003	-0.9%
	0.4	0.0776	0.0721	7.0%	0.0990	0.0809	18.3%	0.1626	0.1146	29.5%
	0.3	0.0726	0.0695	4.3%	0.0926	0.0826	10.8%	0.1519	0.1286	15.4%
	0.2	0.0635	0.0620	2.3%	0.0809	0.0764	5.6%	0.1328	0.1246	6.2%
100	0.1	0.0336	0.0333	1.0%	0.0427	0.0417	2.3%	0.0685	0.0671	2.1%
	0.4	0.0549	0.0511	6.6%	0.0697	0.0569	18.3%	0.1120	0.0775	30.8%
	0.3	0.0513	0.0491	4.3%	0.0650	0.0578	11.1%	0.1048	0.0867	17.3%
	0.2	0.0448	0.0437	2.5%	0.0569	0.0534	6.2%	0.0914	0.0837	8.4%
200	0.1	0.0237	0.0235	1.0%	0.0300	0.0292	2.6%	0.0479	0.0463	3.5%
	0.4	0.0388	0.0361	6.9%	0.0491	0.0401	18.3%	0.0782	0.0537	31.3%
	0.3	0.0362	0.0347	4.4%	0.0459	0.0407	11.2%	0.0733	0.0599	18.3%
	0.2	0.0317	0.0309	2.5%	0.0401	0.0376	6.3%	0.0638	0.0577	9.6%
500	0.1	0.0150	0.0148	1.1%	0.0190	0.0185	2.7%	0.0301	0.0288	4.1%
	0.4	0.0245	0.0228	6.9%	0.0310	0.0253	18.4%	0.0492	0.0336	31.7%
	0.3	0.0229	0.0219	4.4%	0.0290	0.0257	11.3%	0.0460	0.0374	18.6%
	0.2	0.0200	0.0195	2.6%	0.0253	0.0237	6.4%	0.0402	0.0361	10.3%

combinações de (n_{+++}, α_2) iguais a $(10,0.5)$, $(10,0.8)$ e $(20,0.8)$, quando $E(n_{1ij}) \leq 1$ para dois ou mais (i, j) . Para as outras combinações de (n_{+++}, α_2) , quando $E(n_{1ij}) \geq 1.5$ para todos (i, j) , conclui-se que o ganho em se utilizar todas as unidades amostrais em relação a se ignorar as que possuem omissão em Y_2 é maior

- conforme a probabilidade de omissão aumenta (α_2 aumenta e $\alpha_1 = 1 - \alpha_2$ diminui), podendo ocorrer um ganho adicional de até 13 pontos percentuais entre valores sucessivos de α_2 ;
- para probabilidades $\{\theta_{ij}\}$ mais próximas de 0.5 (o que é uma decorrência natural de os erros padrões serem mais elevados para estimadores dessas probabilidades mesmo na ausência de omissão), podendo ocorrer um ganho adicional de até 13 pontos percentuais para valores sucessivos θ_{ij} , em ordem crescente;
- quanto maior o tamanho da amostra, podendo ocorrer um ganho adicional de até 2 pontos percentuais entre valores sucessivos de n_{+++} .

1.2.3 Omissão informativa

Em muitos casos é natural pensar que as probabilidades condicionais de omissão dependem de algum modo das categorias de resposta não observadas, *e.g.*, quando os indivíduos que não responderam a alguma questão em uma sondagem foram os que tinham vergonha de indicar uma das opções de resposta ou pessoas que estão recebendo um tratamento não comparecem à visita médica devido ao seu bom estado de saúde. Este tipo de omissão é conhecido como **omissão informativa**, **omissão não-aleatória** ou **MNAR** (*missing not at random*).

Desenvolvendo (1.4), pode-se notar que quando as probabilidades condicionais de omissão dependem de algum modo das categorias de resposta não-observadas, o processo de omissão não é ignorável sob o ponto de vista das inferências frequentistas ou verossimilhançistas sobre θ .

Na Tabela 1.10 são ilustradas duas estruturas MNAR saturadas para as probabilidades condicionais de omissão. Na primeira, α_1 (α_2) é o incremento na probabilidade condicional de omissão de $Y_1 = 2$ ($Y_2 = 2$) em relação à probabilidade condicional de omissão de $Y_1 = 1$ ($Y_2 = 1$) para os cenários em que há omissão em alguma(s) variável(eis). Na segunda, α_1 (α_2) é o incremento na probabilidade condicional de omissão de $Y_1 = 2$ ($Y_2 = 2$) em relação à probabilidade condicional de omissão de $Y_1 = 1$ ($Y_2 = 1$) para os cenários em que há omissão em Y_1 (Y_2).

Fay (1986) ilustra uma maneira alternativa de propor estruturas MNAR, utilizada, por exemplo, por Little (1985) na análise de dados categorizados longitudinais por meio de modelos

Tabela 1.10: Duas estruturas MNAR para as probabilidades condicionais de omissão

$Y_1 \setminus Y_2$	$W = 2 (Y_1^o, Y_2^m)$		$W = 3 (Y_1^m, Y_2^o)$		$W = 4 (Y_1^m, Y_2^m)$	
	1	2	1	2	1	2
1	α_{20}	$\alpha_{20} + \alpha_2$	α_{30}	$\alpha_{30} + \alpha_2$	α_{40}	$\alpha_{40} + \alpha_2$
2	$\alpha_{20} + \alpha_1$	$\alpha_{20} + \alpha_1 + \alpha_2$	$\alpha_{30} + \alpha_1$	$\alpha_{30} + \alpha_1 + \alpha_2$	$\alpha_{40} + \alpha_1$	$\alpha_{40} + \alpha_1 + \alpha_2$
1	α_{20}	$\alpha_{20} + \alpha_2$	α_{30}	α_{30}	α_{40}	$\alpha_{40} + \alpha_2$
2	α_{20}	$\alpha_{20} + \alpha_2$	$\alpha_{30} + \alpha_1$	$\alpha_{30} + \alpha_1$	$\alpha_{40} + \alpha_1$	$\alpha_{40} + \alpha_1 + \alpha_2$

bayesianos. Denotando a probabilidade de observar Y_1 por $\psi_{1(ij)}$, a de observar Y_2 dado que Y_1 foi observado por $\psi_{21(ij)}$ e a de observar Y_2 dado que Y_1 não foi observado por $\psi_{20(ij)}$, pode-se obter a seguinte reparametrização para as probabilidades condicionais de omissão

$$\begin{aligned}
\lambda_{1(ij)} &= \psi_{1(ij)}\psi_{21(ij)}, \\
\lambda_{2(ij)} &= \psi_{1(ij)}(1 - \psi_{21(ij)}), \\
\lambda_{3(ij)} &= (1 - \psi_{1(ij)})\psi_{20(ij)}, \\
\lambda_{4(ij)} &= (1 - \psi_{1(ij)})(1 - \psi_{20(ij)}).
\end{aligned} \tag{1.15}$$

Com essa parametrização pode-se facilmente propor estruturas MNAR. Dentre uma infinidade delas, uma não sobre-parametrizada é

$$\begin{aligned}
\text{logito}(\psi_{1(ij)}) &= \alpha_{10} + \alpha_1(i - 1) + \alpha_2(j - 1), \\
\text{logito}(\psi_{21(ij)}) &= \alpha_{20} + \alpha_1(i - 1) + \alpha_2(j - 1), \\
\text{logito}(\psi_{20(ij)}) &= \alpha_{30} + \alpha_1(i - 1) + \alpha_2(j - 1).
\end{aligned} \tag{1.16}$$

Sob essa estrutura, α_{10} é a log-chance de se observar Y_1 para $Y_1 = 1, Y_2 = 1$; α_{20} e α_{30} têm interpretações análogas para $\psi_{21(ij)}$ e $\psi_{20(ij)}$; α_1 (α_2) é o incremento comum aos logitos de $\psi_{1(ij)}$, $\psi_{21(ij)}$ e $\psi_{20(ij)}$ para $Y_1 = 2$ ($Y_2 = 2$) em relação a $Y_1 = 1$ ($Y_2 = 1$) independentemente do valor de Y_2 (Y_1). Note que (1.16) é uma reparametrização do modelo MCAR quando $\alpha_1 = \alpha_2 = 0$ ¹³

¹³Contudo, a estrutura MAR saturada não pode ser encaixada em (1.16). Um modelo MAR reduzido, mas não tão reduzido quanto o MCAR, pode ser explicitado por

$$\begin{aligned}
\text{logito}(\psi_{1(ij)}) &= \alpha_{10}, \\
\text{logito}(\psi_{21(ij)}) &= \alpha_{20} + \alpha_1(i - 1), \\
\text{logito}(\psi_{20(ij)}) &= \alpha_{30}.
\end{aligned} \tag{1.17}$$

e a inexistência de associação entre a omissão de Y_1 e Y_2 ocorre quando $\psi_{21(ij)} = \psi_{20(ij)}$ ou, equivalentemente, $\alpha_{20} = \alpha_{30}$.

Elashoff e Elashoff (1974) aparentemente foram os primeiros a estudar o que pode acontecer quando a omissão depende da categoria de resposta da variável de interesse. Stasny (1986) generalizou o trabalho de Chen e Fienberg (1974) permitindo que a probabilidade de não-resposta dependa da categoria não-observada. Choi e Stablein (1988) propuseram testes de homogeneidade marginal para dados binários pareados com omissão informativa. Baker e Laird (1988) reanalisaram os dados de uma pesquisa de intenção de voto (eleição de Truman em 1948), em que uma ACC errou o resultado em 8%, e mostraram que, se tivessem utilizado mecanismos de omissão informativos, poderiam ter errado em apenas 2%.

Uma característica indesejável de estruturas saturadas MNAR é que elas podem não ter ajustes perfeitos e isto está associado à obtenção de estimativas das probabilidades condicionais de omissão na fronteira do espaço paramétrico (Baker e Laird, 1988). Por simplicidade, ilustra-se esse fato para o caso em que se observou apenas os cenários $W = 1, 2$ com a estrutura MNAR da Tabela 1.11.

Tabela 1.11: Estrutura MNAR para as probabilidades condicionais de omissão

$Y_1 \setminus Y_2$	$W = 1 (Y_1^o, Y_2^o)$		$W = 2 (Y_1^o, Y_2^m)$	
	1	2	1	2
1	$1 - \alpha_1$	$1 - \alpha_2$	α_1	α_2
2	$1 - \alpha_1$	$1 - \alpha_2$	α_1	α_2

Sob essa estrutura saturada, α_1 (α_2) representa a probabilidade de não se observar Y_2 , quando $Y_2 = 1$ ($Y_2 = 2$), independentemente dos dois níveis de Y_1 . Para se ter um ajuste perfeito, as frequências observadas devem coincidir com as esperadas estimadas sob o modelo utilizado, o que é traduzido por

$$n_{111} = n_{+++} \hat{\theta}_{11} (1 - \hat{\alpha}_1), \quad n_{112} = n_{+++} \hat{\theta}_{12} (1 - \hat{\alpha}_2), \quad (1.18)$$

$$n_{121} = n_{+++} \hat{\theta}_{21} (1 - \hat{\alpha}_1), \quad n_{122} = n_{+++} \hat{\theta}_{22} (1 - \hat{\alpha}_2), \quad (1.19)$$

$$n_{21+} = n_{+++} (\hat{\theta}_{11} \hat{\alpha}_1 + \hat{\theta}_{12} \hat{\alpha}_2), \quad n_{22+} = n_{+++} (\hat{\theta}_{21} \hat{\alpha}_1 + \hat{\theta}_{22} \hat{\alpha}_2). \quad (1.20)$$

Substituindo as estimativas dos componentes de $\boldsymbol{\theta}$ de (1.18) e (1.19) em (1.20), chega-se ao

sistema

$$\begin{aligned} n_{21+} &= n_{111} \frac{\hat{\alpha}_1}{1 - \hat{\alpha}_1} + n_{112} \frac{\hat{\alpha}_2}{1 - \hat{\alpha}_2}, \\ n_{22+} &= n_{121} \frac{\hat{\alpha}_1}{1 - \hat{\alpha}_1} + n_{122} \frac{\hat{\alpha}_2}{1 - \hat{\alpha}_2}, \end{aligned}$$

ou, equivalentemente em formulação matricial,

$$\begin{pmatrix} n_{21+} \\ n_{22+} \end{pmatrix} = \begin{pmatrix} n_{111} & n_{112} \\ n_{121} & n_{122} \end{pmatrix} \begin{pmatrix} \frac{\hat{\alpha}_1}{1 - \hat{\alpha}_1} \\ \frac{\hat{\alpha}_2}{1 - \hat{\alpha}_2} \end{pmatrix}. \quad (1.21)$$

Para que o sistema (1.21) possua uma única solução para $\left(\frac{\hat{\alpha}_1}{1 - \hat{\alpha}_1}, \frac{\hat{\alpha}_2}{1 - \hat{\alpha}_2}\right)'$, a matriz

$$\begin{pmatrix} n_{111} & n_{112} \\ n_{121} & n_{122} \end{pmatrix}$$

deve ter posto completo (= 2), ou seja, não pode ser singular, o que implica

$$\det \begin{pmatrix} n_{111} & n_{112} \\ n_{121} & n_{122} \end{pmatrix} \neq 0 \implies \frac{n_{111}}{n_{121}} \neq \frac{n_{112}}{n_{122}}.$$

Neste caso, chega-se à solução

$$\hat{\alpha}_1 = 1 - \left(1 + \frac{n_{112}n_{22+} - n_{21+}n_{122}}{n_{112}n_{121} - n_{111}n_{122}}\right)^{-1}, \quad \hat{\alpha}_2 = 1 - \left(1 + \frac{n_{21+}n_{121} - n_{111}n_{22+}}{n_{112}n_{121} - n_{111}n_{122}}\right)^{-1}. \quad (1.22)$$

Estudando as expressões em (1.22), necessárias para se ter um ajuste perfeito, conclui-se que, para $\hat{\alpha}_1$ e $\hat{\alpha}_2$ estarem dentro do espaço paramétrico (entre 0 e 1), n_{21+}/n_{22+} deve estar entre n_{111}/n_{121} e n_{112}/n_{122} . Nessa situação, os EMV dos componentes de α são dados por (1.22) e os de θ são obtidos substituindo (1.22) em (1.18) e (1.19). Em caso contrário, o ajuste não será perfeito, resultando em valores não nulos para as estatísticas de teste de ajuste com zero graus de liberdade!

Analisando tabelas de freqüências em que n_{21+}/n_{22+} não está entre n_{111}/n_{121} e n_{112}/n_{122} , nota-se que pelo menos uma das estimativas de máxima verossimilhança (EMV¹⁴) de (α_1, α_2) obtidas com o algoritmo EM é nula. A verossimilhança com a estrutura da Tabela 1.11 impondo $\alpha_1 = 0$ permite a obtenção dos EMV analiticamente, que são dados por

$$\hat{\theta}_{11} = \frac{n_{111}}{n_{+++}}, \quad \hat{\theta}_{12} = \frac{n_{112} + n_{21+}}{n_{+++}}, \quad \hat{\theta}_{21} = \frac{n_{121}}{n_{+++}}, \quad \hat{\theta}_{22} = \frac{n_{122} + n_{22+}}{n_{+++}}, \quad (1.23)$$

¹⁴Note que a mesma abreviatura EMV foi utilizada para estimadores e estimativas de máxima verossimilhança, cabendo ao leitor fazer a distinção entre as duas conforme o contexto.

$$\hat{\alpha}_2 = 1 - \left(1 + \frac{n_{21+} + n_{22+}}{n_{112} + n_{122}} \right)^{-1}, \quad (1.24)$$

ilustrando que todas as unidades com omissão em Y_2 são alocadas em $Y_2 = 2$. Realizando o mesmo procedimento, com a imposição de $\alpha_2 = 0$, obtém-se

$$\hat{\theta}_{11} = \frac{n_{111} + n_{21+}}{n_{+++}}, \quad \hat{\theta}_{12} = \frac{n_{112}}{n_{+++}}, \quad \hat{\theta}_{21} = \frac{n_{121} + n_{22+}}{n_{+++}}, \quad \hat{\theta}_{22} = \frac{n_{122}}{n_{+++}}, \quad (1.25)$$

$$\hat{\alpha}_1 = 1 - \left(1 + \frac{n_{22+} + n_{21+}}{n_{121} + n_{111}} \right)^{-1}, \quad (1.26)$$

indicando que todas as unidades com omissão em Y_2 são alocadas em $Y_2 = 1$.

Quando $n_{111}/n_{121} = n_{112}/n_{122} = n_{21+}/n_{22+}$, o sistema (1.21) possui infinitas soluções. Neste caso, $\hat{\alpha}_2$ pode ser obtido em função de $\hat{\alpha}_1$,

$$\hat{\alpha}_2(\hat{\alpha}_1) = \left(1 - \frac{n_{122}}{n_{121} \frac{\hat{\alpha}_1}{1-\hat{\alpha}_1} - n_{22+}} \right)^{-1}, \quad 0 \leq \hat{\alpha}_1 \leq \left(1 + \frac{n_{121}}{n_{22+}} \right)^{-1}, \quad (1.27)$$

ou, *vice-versa*,

$$\hat{\alpha}_1(\hat{\alpha}_2) = \left(1 - \frac{n_{121}}{n_{122} \frac{\hat{\alpha}_2}{1-\hat{\alpha}_2} - n_{22+}} \right)^{-1}, \quad 0 \leq \hat{\alpha}_2 \leq \left(1 + \frac{n_{122}}{n_{22+}} \right)^{-1}, \quad (1.28)$$

resultando em uma faixa de valores para $(\boldsymbol{\theta}, \boldsymbol{\alpha})$ que maximizam a função de verossimilhança¹⁵. Os EMV dos componentes de $\boldsymbol{\alpha}$ são dados por (1.27) [ou (1.28)] e os de $\boldsymbol{\theta}$ são obtidos substituindo (1.27) [ou (1.28)] em (1.18) e (1.19).

Analisando-se os resultados obtidos com o algoritmo EM, também se nota o seguinte, quando $n_{111}/n_{121} = n_{112}/n_{122} \neq n_{21+}/n_{22+}$, caso em que o sistema (1.21) não teria solução: se $n_{111} + n_{121} \neq n_{112} + n_{122}$, os EMV dos componentes de $(\boldsymbol{\theta}, \boldsymbol{\alpha})$ são dados por (1.23)–(1.24) ou (1.25)–(1.26), ou seja, a função de verossimilhança possui um único máximo global com $\alpha_1 = 0$ ou $\alpha_2 = 0$; contudo, se $n_{111} + n_{121} = n_{112} + n_{122}$, tanto (1.23)–(1.24) quanto (1.25)–(1.26) são EMV dos componentes de $(\boldsymbol{\theta}, \boldsymbol{\alpha})$, isto é, há dois máximos globais. Neste último caso, ainda é interessante ressaltar que se as proporções observadas na tabela completamente categorizada forem utilizadas como valores iniciais para os componentes de $\boldsymbol{\theta}$ e os valores iniciais para α_1 e α_2 forem os mesmos no algoritmo EM, o algoritmo converge em duas iterações e não se obtém

¹⁵A faixa de variação de $\hat{\alpha}_1$ [$\hat{\alpha}_2$] em (1.27) [(1.28)] foi obtida estudando os casos em que o correspondente resultado de $\hat{\alpha}_2(\hat{\alpha}_1)$ [$\hat{\alpha}_1(\hat{\alpha}_2)$] assume valores dentro do espaço paramétrico.

os EMV apresentados em (1.23)–(1.24) ou em (1.25)–(1.26). O motivo da deficiência desses valores iniciais no algoritmo EM pode ser compreendido estudando as expressões do passo-E

$$\hat{n}_{2ij}^{(k)} = n_{2i+} \frac{\hat{\theta}_{ij}^{(k)} \hat{\alpha}_j^{(k)}}{\hat{\theta}_{i1}^{(k)} \hat{\alpha}_1^{(k)} + \hat{\theta}_{i2}^{(k)} \hat{\alpha}_2^{(k)}}, \quad i, j = 1, 2, \quad (1.29)$$

e do passo-M

$$\hat{\theta}_{ij}^{(k+1)} = \frac{n_{1ij} + n_{2ij}^{(k)}}{n_{+++}}, \quad i, j = 1, 2, \quad (1.30)$$

$$\hat{\alpha}_j^{(k+1)} = \frac{n_{21j}^{(k)} + n_{22j}^{(k)}}{n_{11j} + n_{12j} + n_{21j}^{(k)} + n_{22j}^{(k)}}, \quad j = 1, 2 \quad (1.31)$$

sob $n_{111}/n_{121} = n_{112}/n_{122}$ e $n_{111} + n_{121} = n_{112} + n_{122}$, que em conjunto implicam $n_{1i1} = n_{1i2}$, $i = 1, 2$, em que $\{\hat{n}_{2ij}^{(k)}\}$, $\{\hat{\theta}_{ij}^{(k)}\}$ e $\{\hat{\alpha}_j^{(k)}\}$ são as estimativas de $\{n_{2ij}\}$, $\{\theta_{ij}\}$ e $\{\alpha_j\}$ na k -ésima iteração do algoritmo EM. Os valores iniciais $\hat{\theta}_{ij}^{(0)} = n_{1ij}/n_{1+++}$, sob $n_{1i1} = n_{1i2}$, resultam em $\hat{\theta}_{i1}^{(0)} = \hat{\theta}_{i2}^{(0)}$ e, juntamente com $\hat{\alpha}_1^{(0)} = \hat{\alpha}_2^{(0)}$, levam a $\hat{n}_{2i1}^{(0)} = \hat{n}_{2i2}^{(0)} (= \frac{n_{2i+}}{2})$. Em seguida, como $n_{1i1} = n_{1i2}$ e $\hat{n}_{2i1}^{(0)} = \hat{n}_{2i2}^{(0)}$, obtém-se $\hat{\theta}_{i1}^{(1)} = \hat{\theta}_{i2}^{(1)}$ e $\hat{\alpha}_1^{(1)} = \hat{\alpha}_2^{(1)}$ e, por conseguinte, $\hat{n}_{2ij}^{(1)} = \hat{n}_{2ij}^{(0)}$, implicando $\hat{\theta}_{ij}^{(2)} = \hat{\theta}_{ij}^{(1)} = (n_{1ij} + \frac{n_{2i+}}{2})/n_{+++}$ e $\hat{\alpha}_j^{(2)} = \hat{\alpha}_j^{(1)} = \frac{n_{21+} + n_{22+}}{2n_{11j} + 2n_{12j} + n_{21+} + n_{22+}}$, que não são expressões equivalentes a (1.23)–(1.24) ou a (1.25)–(1.26). Outros valores iniciais que satisfaçam $\hat{\theta}_{i1}^{(0)} \hat{\alpha}_1^{(0)} = \hat{\theta}_{i2}^{(0)} \hat{\alpha}_2^{(0)}$, $i = 1, 2$, também provocam o mesmo problema de convergência do algoritmo EM.

Comparou-se as EMV de $(\boldsymbol{\theta}, \alpha_1, \alpha_2)$ de todos os casos apresentados com as obtidas pelo algoritmo EM em uma parte do estudo de simulação que será descrito a seguir e notou-se que as diferenças são desprezáveis, diminuindo conforme se aumenta o número de iterações do algoritmo EM. Um resumo dos casos descritos é apresentado na Tabela 1.12, destacando as características especiais, como número de conjuntos de $(\boldsymbol{\theta}, \alpha_1, \alpha_2)$ que maximizam a verossimilhança, se o ajuste é perfeito e se a EMV de α_1 ou α_2 está na fronteira do espaço paramétrico.

A condição para que se tenha um ajuste perfeito não deve ser encarada como surpreendente, uma vez que está embutida nas funções paramétricas da estrutura do modelo MNAR associadas a n_{111}/n_{121} , n_{21+}/n_{22+} e n_{112}/n_{122} . Pode-se verificar isso, notando que

$$\frac{(1 - \alpha_1)\theta_{11}}{(1 - \alpha_1)\theta_{21}} < \frac{\alpha_1\theta_{11} + \alpha_2\theta_{12}}{\alpha_1\theta_{21} + \alpha_2\theta_{22}} \implies \frac{\theta_{11}\theta_{22}}{\theta_{21}\theta_{12}} < 1, \quad \frac{(1 - \alpha_2)\theta_{12}}{(1 - \alpha_2)\theta_{22}} < \frac{\alpha_1\theta_{11} + \alpha_2\theta_{12}}{\alpha_1\theta_{21} + \alpha_2\theta_{22}} \implies \frac{\theta_{11}\theta_{22}}{\theta_{21}\theta_{12}} > 1$$

e que ambas desigualdades não podem ser satisfeitas simultaneamente.

Tabela 1.12: EMV de $(\theta, \alpha_1, \alpha_2)$ sob a estrutura MNAR da Tabela 1.11 e características correspondentes

Caso	Condição	Nº máx.	Qualidade do ajuste	$\hat{\alpha}_1=0$ ou $\hat{\alpha}_2=0$	EMV
I	$\frac{n_{111}}{n_{121}} < \frac{n_{21+}}{n_{22+}} < \frac{n_{112}}{n_{122}}$	1	perfeito	não	(1.22)–(1.18)–(1.19)
II	$\frac{n_{111}}{n_{121}} \neq \frac{n_{112}}{n_{122}}$ e $\frac{n_{111}}{n_{121}} = \frac{n_{21+}}{n_{22+}}$ ou $\frac{n_{112}}{n_{122}} = \frac{n_{21+}}{n_{22+}}$	1	perfeito	sim	(1.23)–(1.24) ou (1.25)–(1.26)
III	$\frac{n_{111}}{n_{121}} < \frac{n_{112}}{n_{122}} < \frac{n_{21+}}{n_{22+}}$ ou $\frac{n_{21+}}{n_{22+}} < \frac{n_{111}}{n_{121}} < \frac{n_{112}}{n_{122}}$	1	imperfeito	sim	(1.23)–(1.24) ou (1.25)–(1.26)
IV	$\frac{n_{111}}{n_{121}} = \frac{n_{112}}{n_{122}} \neq \frac{n_{21+}}{n_{22+}}$ e $n_{111} + n_{121} \neq n_{112} + n_{122}$	1	imperfeito	sim	(1.23)–(1.24) ou (1.25)–(1.26)
V	$\frac{n_{111}}{n_{121}} = \frac{n_{112}}{n_{122}} = \frac{n_{21+}}{n_{22+}}$	∞	perfeito	não/sim	(1.27/1.28)–(1.18)–(1.19)
VI	$\frac{n_{111}}{n_{121}} = \frac{n_{112}}{n_{122}} \neq \frac{n_{21+}}{n_{22+}}$ e $n_{111} + n_{121} = n_{112} + n_{122}$	2	imperfeito	sim	(1.23)–(1.24) e (1.25)–(1.26)

Nº máx.: número de conjuntos de $(\theta, \alpha_1, \alpha_2)$ que maximizam a verossimilhança

$\frac{n_{111}}{n_{121}} < \frac{n_{21+}}{n_{22+}} < \frac{n_{112}}{n_{122}}$ denota “ $\frac{n_{111}}{n_{121}} < \frac{n_{21+}}{n_{22+}} < \frac{n_{112}}{n_{122}}$ ou $\frac{n_{111}}{n_{121}} > \frac{n_{21+}}{n_{22+}} > \frac{n_{112}}{n_{122}}$ ”

Quando os EMV das probabilidades condicionais de omissão estiverem na fronteira do espaço paramétrico, procedimentos iterativos como Newton-Raphson e *scoring* de Fisher podem convergir para estimativas destas probabilidades que são < 0 ou > 1 a não ser que sejam utilizadas (i) restrições de desigualdades nos métodos iterativos, (ii) formulações próprias para probabilidades, como as ligações logito ou probito ou (iii) o algoritmo EM. Apesar de sua estabilidade e fácil implementação, o algoritmo EM tem uma baixa taxa de convergência e não fornece prontamente estimativas para as variâncias dos estimadores. Propostas de soluções para estes problemas foram feitas por Louis (1982), Meilijson (1989), Meng e Rubin (1991, 1993), Liu e Rubin (1994) e Liu, Rubin e Wu (1998), para problemas gerais, e por Baker (1992), especificamente para dados categorizados. Em nossa avaliação, o método de aceleração de Louis (1982), conhecido por *Louis' Turbo EM*, teve uma taxa de convergência compatível com os procedimentos Newton-Raphson e *scoring* de Fisher, mas reteve também os proble-

mas de convergência para valores localizados fora do espaço paramétrico. Isso demonstra que embora o algoritmo EM garanta, em geral, a convergência das estimativas para valores localizados dentro ou na fronteira do espaço paramétrico quando utilizado em sua formulação original (Dempster *et al.*, 1977), ele pode perder esta característica ao ser adaptado, como ocorre no caso do método de aceleração de Louis. O comportamento das outras propostas não foi estudado.

Park e Brown (1994) propuseram a utilização de distribuições *a priori* Dirichlet, com hiperparâmetros obtidos empiricamente (similarmente ao procedimento conhecido como Bayes empírico, *empirical Bayes*), proporcionais às frequências observadas na tabela completamente categorizada, de tal forma que as estimativas obtidas sob mecanismos MNAR fiquem dentro do espaço paramétrico. Por meio de um estudo de simulação, eles demonstraram que os erros quadráticos médios desses estimadores são, em alguns casos, bem menores que os obtidos pelo método da máxima verossimilhança. Park (1998) sugeriu modificações deste método que produzem estimadores com menores erros quadráticos médios. Garcia (2003) apresentou análises na mesma linha de Park e Brown (1994) para tabelas com dados esparsos, que provocam ainda maior instabilidade nos estimadores.

Baker, Rosenberger e DerSimonian (1992) apresentaram EMV em forma fechada para probabilidades em tabelas 2×2 sob diversas estruturas MNAR para o mecanismo de omissão, explicitando também casos com soluções na fronteira do espaço paramétrico. Outras análises considerando modelos estruturais não-ignoráveis para o mecanismo de omissão na mesma linha de Baker e Laird (1988) foram apresentadas por Conaway, Waternaux, Alfred, Bellinger e Levinton (1992), Conaway (1992, 1994), Chambers e Welsh (1993), Baker (1995a, 1996), Rubin, Stern e Vehovar (1995) e Raab e Donnelly (1999).

Conaway (1993), Liu, Waternaux e Petkova (1999), Deltour, Richardson e Hesran (1999) e Albert (2000) utilizaram modelos de transição para modelar a dependência temporal em dados categorizados longitudinais, com omissão não-aleatória, estendendo os trabalhos de Stram, Wei e Ware (1988) e Chen (1988) que fizeram o mesmo com omissão aleatória.

Laird (1988), Heyting, Tolboom e Essers (1992), Shih (1992), Diggle e Shih (1993), Little (1995), Goetghebeur e Molenberghs (1996), Goetghebeur e Shapiro (1996), Robins (1997), Robins e Gill (1997), Goetghebeur, Molenberghs e Katz (1998) e Birmingham, Rotnitzky e Fitzmaurice (2003) discutiram aspectos inferenciais de dados omissos particulares aos estudos longitudinais, sob o contexto de respostas categorizadas e/ou contínuas.

Michiels e Molenberghs (1997) utilizaram a idéia de estimadores protetores (*protective estimators*), propostos por Brown (1990) no contexto de dados multivariados normais com omissão, em dados categorizados longitudinais. Esta classe de estimadores é apropriada quando o abandono do estudo depende apenas das respostas não-observadas. Condições necessárias e suficientes para se obter estimativas dentro do espaço paramétrico foram explicitadas. Um exemplo de estrutura MNAR que fornece essa classe de estimadores é apresentado na Tabela 1.11.

Outra característica inconveniente de estruturas MNAR é que, mesmo quando não sobre-saturadas, podem não ter seus parâmetros identificáveis. Sob a estrutura MNAR saturada da Tabela 1.11, isso significa que existiriam dois conjuntos de parâmetros $(\boldsymbol{\theta}, \alpha_1, \alpha_2)$ com valores diferentes, *i.e.*, $(\boldsymbol{\theta}^0, \alpha_1^0, \alpha_2^0) \neq (\boldsymbol{\theta}^1, \alpha_1^1, \alpha_2^1)$, tal que

$$P(\mathbf{N} | \boldsymbol{\theta}^0, \alpha_1^0, \alpha_2^0) = P(\mathbf{N} | \boldsymbol{\theta}^1, \alpha_1^1, \alpha_2^1), \quad (1.32)$$

para todos os valores das frequências observáveis \mathbf{N} . Uma vez que o modelo multinomial parametrizado por $(\gamma_{111}, \gamma_{112}, \gamma_{121}, \gamma_{21+}, \gamma_{22+})$ é claramente identificável, deve-se questionar se os parâmetros $(\theta_{11}, \theta_{12}, \theta_{21}, \alpha_1, \alpha_2)$ podem ser unicamente determinados a partir de $(\gamma_{111}, \gamma_{112}, \gamma_{121}, \gamma_{21+}, \gamma_{22+})$ segundo a reparametrização adotada

$$\gamma_{111} = \theta_{11}(1 - \alpha_1), \quad \gamma_{112} = \theta_{12}(1 - \alpha_2), \quad (1.33)$$

$$\gamma_{121} = \theta_{21}(1 - \alpha_1), \quad \gamma_{122} = \theta_{22}(1 - \alpha_2), \quad (1.34)$$

$$\gamma_{21+} = \theta_{11}\alpha_1 + \theta_{12}\alpha_2, \quad \gamma_{22+} = \theta_{21}\alpha_1 + \theta_{22}\alpha_2. \quad (1.35)$$

Isolando os componentes de $\boldsymbol{\theta}$ em (1.33) e (1.34) e substituindo em (1.35), chega-se ao sistema

$$\begin{pmatrix} \gamma_{21+} \\ \gamma_{22+} \end{pmatrix} = \begin{pmatrix} \gamma_{111} & \gamma_{112} \\ \gamma_{121} & \gamma_{122} \end{pmatrix} \begin{pmatrix} \frac{\alpha_1}{1-\alpha_1} \\ \frac{\alpha_2}{1-\alpha_2} \end{pmatrix}, \quad (1.36)$$

que deve satisfazer a condição $\gamma_{111}/\gamma_{121} \neq \gamma_{112}/\gamma_{122}$ para ter uma única solução relativamente a $\left(\frac{\alpha_1}{1-\alpha_1}, \frac{\alpha_2}{1-\alpha_2}\right)'$. Portanto, não há um mapeamento biunívoco entre $(\gamma_{111}, \gamma_{112}, \gamma_{121}, \gamma_{21+}, \gamma_{22+})$ e $(\theta_{11}, \theta_{12}, \theta_{21}, \alpha_1, \alpha_2)$ quando $\gamma_{111}/\gamma_{121} = \gamma_{112}/\gamma_{122}$ ¹⁶ ou, equivalentemente, quando

$$\frac{\theta_{11}\theta_{22}}{\theta_{21}\theta_{12}} = 1, \quad (1.37)$$

¹⁶Para exemplificar, verifique que tanto $(\theta_{11}, \theta_{12}, \theta_{21}, \alpha_1, \alpha_2)$ igual a $(0.1, 0.1, 0.4, 0.8, 0.3)$, como $(1/30, 1/6, 2/15, 0.4, 0.58)$, implicam $(\gamma_{111}, \gamma_{112}, \gamma_{121}, \gamma_{21+}, \gamma_{22+}) = (0.02, 0.07, 0.08, 0.11, 0.44)$.

caso em que a suposição de identificabilidade, que é uma das usuais condições de regularidade assumidas (veja, *e.g.*, Lehmann, 1998, pp.456-457), não é satisfeita e, conseqüentemente, os resultados assintóticos do EMV comumente utilizados não permanecem necessariamente válidos.

Apesar da semelhança entre as expressões (1.18)–(1.21) e (1.33)–(1.36), em que as primeiras podem ser obtidas das últimas após substituição de $(\gamma_{111}, \gamma_{112}, \gamma_{121}, \gamma_{122}, \gamma_{21+}, \gamma_{22+})$ e $(\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}, \alpha_1, \alpha_2)$ por seus EMV $\frac{1}{n_{+++}}(n_{111}, n_{112}, n_{121}, n_{122}, n_{21+}, n_{22+})$ e $(\hat{\theta}_{11}, \hat{\theta}_{12}, \hat{\theta}_{21}, \hat{\theta}_{22}, \hat{\alpha}_1, \hat{\alpha}_2)$, deve-se salientar que em (1.18)–(1.21) estudou-se as condições necessárias para se ter um ajuste perfeito, enquanto que em (1.33)–(1.36), para que $(\theta, \alpha_1, \alpha_2)$ sejam identificáveis. Como o verdadeiro valor dos parâmetros é desconhecido na prática, é um equívoco acreditar que seria possível concluir se os parâmetros $(\theta, \alpha_1, \alpha_2)$ são identificáveis com base em determinadas freqüências observáveis, pois um modelo probabilístico sob $\gamma_{111}/\gamma_{121} \neq \gamma_{112}/\gamma_{122}$ pode originar freqüências observáveis $n_{111}/n_{121} = n_{112}/n_{122}$ e outro modelo probabilístico sob $\gamma_{111}/\gamma_{121} = \gamma_{112}/\gamma_{122}$ pode originar freqüências observáveis $n_{111}/n_{121} \neq n_{112}/n_{122}$.

Rothenberg (1971) faz uma distinção entre identificabilidade global e local de $(\theta^0, \alpha_1^0, \alpha_2^0)$. A primeira é definida por não existir $(\theta^1, \alpha_1^1, \alpha_2^1) \neq (\theta^0, \alpha_1^0, \alpha_2^0)$ que implique (1.32) em todo o espaço paramétrico e, a última, numa vizinhança aberta de $(\theta^0, \alpha_1^0, \alpha_2^0)$. Em seguida, ele demonstra que, sob certas condições de regularidade, se a matriz de informação de Fisher for não singular, então os parâmetros do modelo estrutural são localmente identificáveis. Contudo, se a matriz de informação de Fisher for singular, então os parâmetros são globalmente não identificáveis, pois identificabilidade local é uma condição necessária para identificabilidade global. A matriz de informação de Fisher é uma medida da quantidade de informação sobre os parâmetros desconhecidos disponível na amostra. Deste modo, a inidentificabilidade corresponde à falta de informação suficiente para distinguir entre estruturas subjacentes alternativas.

Como a matriz de informação de Fisher é uma matriz de covariâncias do vetor (aleatório) *score*, ela é semidefinida positiva. Se, adicionalmente, os componentes do vetor *score* forem linearmente independentes, ela será positiva definida¹⁷ (Lehmann e Casella, 1998, p.125). Caso os componentes do vetor *score* sejam linearmente dependentes, a sua distribuição será degenerada e a matriz de informação de Fisher, singular. Para exemplificar esta questão no contexto da estrutura MNAR da Tabela 1.11, considere o logaritmo da função de verossimilhança cor-

¹⁷Veja, *e.g.*, as definições de matriz semidefinida e definida positiva em Anderson (2003, p.628).

respondente

$$\begin{aligned} \ln L(\boldsymbol{\tau} | \mathbf{N}) \propto & n_{111} \ln(\theta_{11}[1 - \alpha_1]) + n_{112} \ln(\theta_{12}[1 - \alpha_2]) + \\ & n_{121} \ln(\theta_{21}[1 - \alpha_1]) + n_{122} \ln([1 - \theta_{11} - \theta_{12} - \theta_{21}][1 - \alpha_2]) + \\ & n_{21+} \ln(\theta_{11}\alpha_1 + \theta_{12}\alpha_2) + n_{22+} \ln(\theta_{21}\alpha_1 + [1 - \theta_{11} - \theta_{12} - \theta_{21}]\alpha_2) \end{aligned} \quad (1.38)$$

e o seu vetor *score*

$$\mathbf{S}(\boldsymbol{\tau}) = \begin{bmatrix} \frac{n_{111}}{\theta_{11}} - \frac{n_{122}}{1 - \theta_{11} - \theta_{12} - \theta_{21}} + \frac{n_{21+\alpha_1}}{\theta_{11}\alpha_1 + \theta_{12}\alpha_2} - \frac{n_{22+\alpha_2}}{\theta_{21}\alpha_1 + (1 - \theta_{11} - \theta_{12} - \theta_{21})\alpha_2} \\ \frac{n_{112}}{\theta_{12}} - \frac{n_{122}}{1 - \theta_{11} - \theta_{12} - \theta_{21}} + \frac{n_{21+\alpha_2}}{\theta_{11}\alpha_1 + \theta_{12}\alpha_2} - \frac{n_{22+\alpha_2}}{\theta_{21}\alpha_1 + (1 - \theta_{11} - \theta_{12} - \theta_{21})\alpha_2} \\ \frac{n_{121}}{\theta_{21}} - \frac{n_{122}}{1 - \theta_{11} - \theta_{12} - \theta_{21}} + \frac{n_{22+(\alpha_1 - \alpha_2)}}{\theta_{21}\alpha_1 + (1 - \theta_{11} - \theta_{12} - \theta_{21})\alpha_2} \\ - \frac{n_{111} + n_{121}}{1 - \alpha_1} + \frac{n_{21+\theta_{11}}}{\theta_{11}\alpha_1 + \theta_{12}\alpha_2} + \frac{n_{22+\theta_{21}}}{\theta_{21}\alpha_1 + (1 - \theta_{11} - \theta_{12} - \theta_{21})\alpha_2} \\ - \frac{n_{112} + n_{122}}{1 - \alpha_2} + \frac{n_{21+\theta_{12}}}{\theta_{11}\alpha_1 + \theta_{12}\alpha_2} + \frac{n_{22+(1 - \theta_{11} - \theta_{12} - \theta_{21})}}{\theta_{21}\alpha_1 + (1 - \theta_{11} - \theta_{12} - \theta_{21})\alpha_2} \end{bmatrix}, \quad (1.39)$$

em que $\boldsymbol{\tau} = (\theta_{11}, \theta_{12}, \theta_{21}, \alpha_1, \alpha_2)'$. Ao avaliar (1.39) sob (1.37)¹⁸, *i.e.*, após substituir θ_{12} por $\theta_{11}\xi$ e, conseqüentemente, $1 - \theta_{11} - \theta_{12} - \theta_{21}$ por $\theta_{21}\xi$, em que $\xi = \frac{1 - \theta_{11} - \theta_{21}}{\theta_{11} + \theta_{21}}$, chega-se ao vetor *score*

$$\mathbf{S}(\boldsymbol{\tau}^1) = \begin{bmatrix} \frac{\partial \ln L}{\partial \theta_{11}}(\boldsymbol{\tau}^1) \\ \frac{\partial \ln L}{\partial \theta_{12}}(\boldsymbol{\tau}^1) \\ \frac{\partial \ln L}{\partial \theta_{21}}(\boldsymbol{\tau}^1) \\ \frac{\partial \ln L}{\partial \alpha_1}(\boldsymbol{\tau}^1) \\ \frac{\partial \ln L}{\partial \alpha_2}(\boldsymbol{\tau}^1) \end{bmatrix} = \begin{bmatrix} \frac{n_{111}}{\theta_{11}} - \frac{n_{122}}{\theta_{21}\xi} + \frac{n_{21+\alpha_1}}{\theta_{11}(\alpha_1 + \xi\alpha_2)} - \frac{n_{22+\alpha_2}}{\theta_{21}(\alpha_1 + \xi\alpha_2)} \\ \frac{n_{112}}{\theta_{11}\xi} - \frac{n_{122}}{\theta_{21}\xi} + \frac{n_{21+\alpha_2}}{\theta_{11}(\alpha_1 + \xi\alpha_2)} - \frac{n_{22+\alpha_2}}{\theta_{21}(\alpha_1 + \xi\alpha_2)} \\ \frac{n_{121}}{\theta_{21}} - \frac{n_{122}}{\theta_{21}\xi} + \frac{n_{22+(\alpha_1 - \alpha_2)}}{\theta_{21}(\alpha_1 + \xi\alpha_2)} \\ - \frac{n_{111} + n_{121}}{1 - \alpha_1} + \frac{n_{21+} + n_{22+}}{\alpha_1 + \xi\alpha_2} \\ - \frac{n_{112} + n_{122}}{1 - \alpha_2} + \frac{\xi(n_{21+} + n_{22+})}{\alpha_1 + \xi\alpha_2} \end{bmatrix}, \quad (1.40)$$

¹⁸Substituindo a restrição $\theta_{22} = 1 - \theta_{11} - \theta_{12} - \theta_{21}$ em $\frac{\theta_{11}\theta_{22}}{\theta_{21}\theta_{12}} = 1$, chega-se a

$$\frac{\theta_{11}(1 - \theta_{11} - \theta_{12} - \theta_{21})}{\theta_{21}\theta_{12}} = 1 \implies \theta_{12} = \theta_{11} \frac{1 - \theta_{11} - \theta_{21}}{\theta_{11} + \theta_{21}}.$$

Logo,

$$1 - \theta_{11} - \theta_{12} - \theta_{21} = 1 - \theta_{11} - \theta_{11} \frac{1 - \theta_{11} - \theta_{21}}{\theta_{11} + \theta_{21}} - \theta_{21} = \theta_{21} \frac{1 - \theta_{11} - \theta_{21}}{\theta_{11} + \theta_{21}}.$$

em que $\boldsymbol{\tau}^1 = \left(\theta_{11}, \theta_{12} = \theta_{11} \frac{1-\theta_{11}-\theta_{21}}{\theta_{11}+\theta_{21}}, \theta_{21}, \alpha_1, \alpha_2 \right)'$. Estudando (1.40), verifica-se que, *e.g.*, o último componente pode ser obtido em função dos demais de acordo com

$$\frac{\theta_{11}\xi}{1-\alpha_2} \frac{\partial \ln L}{\partial \theta_{11}}(\boldsymbol{\tau}^1) - \frac{\theta_{11}\xi}{1-\alpha_2} \frac{\partial \ln L}{\partial \theta_{12}}(\boldsymbol{\tau}^1) + \frac{\theta_{21}\xi}{1-\alpha_2} \frac{\partial \ln L}{\partial \theta_{21}}(\boldsymbol{\tau}^1) + \xi \frac{1-\alpha_1}{1-\alpha_2} \frac{\partial \ln L}{\partial \alpha_1}(\boldsymbol{\tau}^1) = \frac{\partial \ln L}{\partial \alpha_2}(\boldsymbol{\tau}^1),$$

indicando que no ponto $\boldsymbol{\tau}^1$ a distribuição do vetor *score* (1.39) é degenerada, a matriz de informação de Fisher correspondente é singular e os parâmetros $(\boldsymbol{\theta}, \alpha_1, \alpha_2)$ são inidentificáveis. Para uma revisão e discussão mais abrangentes sobre a questão de identificabilidade, veja Paulino e Pereira (1994).

Fitzmaurice, Laird e Zahner (1996) classificaram estruturas MNAR como identificáveis ou inidentificáveis com base numa única avaliação da matriz de informação de Fisher. Glonek (1999) ilustrou condições necessárias de identificabilidade dos parâmetros em análises de dados binários com omissão não-aleatória, similarmente ao que foi realizado em (1.33)–(1.36), chamando a atenção de que para 5 das 8 estruturas consideradas por Fitzmaurice, Laird e Zahner (1996), a questão de identificabilidade não pode ser resolvida de maneira tão direta como os autores fizeram, pois depende dos valores particulares dos parâmetros escolhidos.

Molenberghs, Goetghebeur, Lipsitz e Kenward (1999) apresentaram diversos modelos para a análise de dados categorizados com omissão, exemplificando os problemas de estimativas fora do espaço paramétrico e a falta de identificabilidade de parâmetros. Bonetti, Cole e Gelber (1999) analisaram dados categorizados de qualidade de vida com omissão informativa e também discutiram questões de identificabilidade. Smith, Skinner e Clarke (1999) utilizaram modelos informativos para a análise dos dados obtidos na pesquisa de eleição britânica de 1992, mostrando que diversos modelos não saturados MNAR também podem ter estimativas na fronteira no espaço paramétrico; propuseram uma interpretação geométrica dos EMV e sugeriram que esse comportamento extremo não deve ser encarado como uma evidência contra o mecanismo de omissão assumido. Também mostraram que sob alguns modelos, quando a verossimilhança perfilada se torna muito plana, intervalos de confiança muito grandes podem ser reduzidos impondo limites plausíveis para as chances de respostas.

Clarke (2002) estendeu os resultados de Glonek (1999) e mostrou que as estimativas obtidas na fronteira do espaço paramétrico não devem ser interpretadas como indicador de que a estrutura para o mecanismo de omissão esteja incorreta. Clarke e Smith (2004) analisaram, através de simulações, a cobertura de intervalos de confiança do tipo Wald, *bootstrap* (Efron e Tibshirani, 1994; Davison e Hinkley, 1997) e da verossimilhança perfilada para EMV quando

estimativas dos parâmetros do mecanismo de omissão MNAR estão na fronteira do espaço paramétrico. Os três métodos produzem uma cobertura indesejável para parâmetros do mecanismo de omissão, mas bom desempenho para os outros parâmetros de interesse. Esses autores também comentaram que, quando o modelo para o mecanismo de omissão informativo for verdadeiro, a probabilidade de se ter uma solução na fronteira do espaço paramétrico é assintoticamente nula, mas positiva em amostras finitas. Nestes casos, o EMV tem uma distribuição de probabilidade discreta que não é bem aproximada pela distribuição normal.

Rotnitzky, Cox, Bottai e Robins (2000) forneceram uma teoria unificada para derivar a distribuição assintótica do EMV e da estatística de teste de razão de verossimilhanças (TRV) sob a hipótese nula de que os parâmetros assumem o valor em que a matriz de informação de Fisher é singular. No contexto da estrutura MNAR da Tabela 1.11, devido a (1.37), tal incluiria a testar-se a ausência de associação entre Y_1 e Y_2 . Os autores apresentam um exemplo de mecanismo não-ignorável em respostas contínuas. Nessa mesma linha, Hens (2005) reinterpreta a teoria e questiona a dificuldade de sua aplicação, contrastando com resultados de estudos de simulação da estatística de TRV e *bootstrap*. Bottai (2003) examina o comportamento assintótico de regiões de confiança em modelos paramétricos unidimensionais quando a informação de Fisher é igual a zero. Contudo, todos os autores assumem identificabilidade como uma das condições de regularidade.

Com o objetivo de avaliar a frequência de ocorrência de soluções em cada um dos casos da Tabela 1.12 em amostras finitas, realizou-se um estudo de simulação gerando cem mil réplicas de Monte Carlo de distribuições multinomiais com a estrutura MNAR da Tabela 1.11, com $(\alpha_1, \alpha_2) = (0.8, 0.3)$, combinadas com as estruturas para $\{\theta_{ij}\}$ definidas na Tabela 1.13 e $n_{+++} = 50, 100, 200, 500, 1\ 000, 2\ 000, 5\ 000, 10\ 000, 20\ 000, 50\ 000, 100\ 000$.

As porcentagens de réplicas em cada um dos casos explicitados na Tabela 1.12 são apresentadas nas Tabelas 1.15 e 1.16. Dentre as estruturas A1–D1 (Tabela 1.15), que não sofrem problemas de identificabilidade, a ordem crescente das estruturas com maior número de réplicas do caso I (ajuste perfeito, EMV dentro do espaço paramétrico, um máximo global) é a mesma das que possuem as funções paramétricas associadas a n_{21+}/n_{22+} mais próximas daquelas associadas a n_{111}/n_{121} ou a n_{112}/n_{122} ($D1 < C1 < A1 < B1$), conforme ilustrado na Tabela 1.14, concordando com as condições previamente estudadas para esse caso. O fato de o número de soluções fronteira (casos II a VI) ser positivo em amostras finitas, mas decrescente, tendendo a zero assintoticamente, coincide com os comentários de Clarke e Smith

Tabela 1.13: Estruturas para as probabilidades marginais $\{\theta_{ij}\}$

$Y_1 \setminus Y_2$	Estrutura A1		Estrutura B1		Estrutura C1		Estrutura D1	
	1	2	1	2	1	2	1	2
1	0.30	0.20	0.20	0.10	0.30	0.20	0.30	0.10
2	0.20	0.30	0.20	0.50	0.40	0.10	0.40	0.20

$Y_1 \setminus Y_2$	Estrutura A2		Estrutura B2		Estrutura C2		Estrutura D2	
	1	2	1	2	1	2	1	2
1	0.25	0.25	0.20	0.20	0.15	0.15	0.10	0.10
2	0.25	0.25	0.30	0.30	0.35	0.35	0.40	0.40

Tabela 1.14: Funções paramétricas associadas a $\frac{n_{111}}{n_{121}}$, $\frac{n_{112}}{n_{122}}$ e $\frac{n_{21+}}{n_{22+}}$

Função paramétrica	Estrutura								
	A1	B1	C1	D1	A2	B2	C2	D2	
$\frac{\gamma_{111}}{\gamma_{121}} = \frac{(1 - \alpha_1)\theta_{11}}{(1 - \alpha_1)\theta_{21}}$	1.50	1.00	0.75	0.75	1.00	0.67	0.43	0.25	
$\frac{\gamma_{112}}{\gamma_{122}} = \frac{(1 - \alpha_2)\theta_{12}}{(1 - \alpha_2)\theta_{22}}$	0.67	0.20	2.00	0.50	1.00	0.67	0.43	0.25	
$\frac{\gamma_{21+}}{\gamma_{22+}} = \frac{\alpha_1\theta_{11} + \alpha_2\theta_{12}}{\alpha_1\theta_{21} + \alpha_2\theta_{22}}$	1.20	0.61	0.86	0.71	1.00	0.67	0.43	0.25	

(2004). O mesmo comportamento não é observado para as estruturas A2–D2 (Tabela 1.16), que apresentam problemas de identificabilidade. Independentemente dos valores de $\{\theta_{ij}\}$ na classe das estruturas A2–D2 (Tabela 1.13), as proporções de réplicas com soluções para todos os parâmetros dentro do espaço paramétrico (caso I) permanecem com a mesma magnitude (cerca de 41 a 43%) para todos os tamanhos de amostra considerados.

Adicionalmente, em cada réplica de Monte Carlo, computou-se a EMV de $\theta_{+1} - \theta_{1+}$ sob a ACC e sob o mecanismo MNAR verdadeiro. Para cada condição realizou-se o TRV para a homogeneidade marginal, rejeitando a hipótese com um nível de 5% de significância. As médias de todas as estimativas e as porcentagens de rejeições estão ilustradas, respectivamente, nas Tabelas 1.17–1.18 e 1.19–1.20, apenas para os casos I e III, pois concentram a maior parte das réplicas.

Tabela 1.15: Proporções de réplicas (%) com soluções em cada um dos casos (I a VI) da Tabela 1.12

n_{++++}	Estrutura A1						Estrutura B1						Estrutura C1						Estrutura D1					
	I	II	III	IV	V	VI	I	II	III	IV	V	VI	I	II	III	IV	V	VI	I	II	III	IV	V	VI
50	50.1	2.4	45.4	1.9	0.2	0.0	62.4	2.0	35.1	0.5	0.0	*	50.5	2.1	44.8	1.6	0.1	0.8	43.9	2.3	50.1	2.3	0.1	1.2
100	57.3	1.0	41.0	0.7	0.0	*	71.6	0.7	27.5	0.2	*	*	56.5	0.9	42.0	0.5	0.0	0.2	46.0	1.0	51.6	1.0	0.0	0.4
200	65.2	0.4	34.2	0.2	0.0	*	80.9	0.3	18.8	0.0	*	*	62.0	0.3	37.6	0.1	0.0	0.0	48.9	0.3	50.3	0.4	0.0	0.1
500	76.6	0.1	23.4	0.0	0.0	*	91.9	0.0	8.1	*	*	*	69.4	0.1	30.5	*	*	*	54.7	0.1	45.1	0.1	*	0.0
1 000	84.4	0.0	15.6	*	*	*	97.7	0.0	2.3	*	*	*	76.4	0.0	23.6	*	*	*	60.5	0.0	39.5	0.0	*	*
2 000	92.6	0.0	7.4	*	*	*	99.8	*	0.2	*	*	*	84.2	0.0	15.8	*	*	*	65.6	0.0	34.4	0.0	*	*
5 000	98.9	0.0	1.1	*	*	*	100.0	*	0.0	*	*	*	94.5	0.0	5.5	*	*	*	73.7	0.0	26.3	*	*	*
10 000	99.9	*	0.1	*	*	*	100.0	*	*	*	*	*	98.8	*	1.2	*	*	*	81.6	*	18.4	*	*	*
20 000	100.0	*	*	*	*	*	100.0	*	*	*	*	*	99.9	*	0.1	*	*	*	89.9	*	10.1	*	*	*
50 000	100.0	*	*	*	*	*	100.0	*	*	*	*	*	100.0	*	*	*	*	*	97.8	*	2.2	*	*	*
100 000	100.0	*	*	*	*	*	100.0	*	*	*	*	*	100.0	*	*	*	*	*	99.8	*	0.2	*	*	*

Obs.: 0.0% [*] indica que um número desprezável (*i.e.*, entre 0 e 50) [exatamente 0] de réplicas de Monte Carlo (das 100 000) apresentaram determinado caso.

Tabela 1.16: Proporções de réplicas (%) com soluções em cada um dos casos (I a VI) da Tabela 1.12

n_{+++}	Estrutura A2						Estrutura B2						Estrutura C2						Estrutura D2					
	I	II	III	IV	V	VI	I	II	III	IV	V	VI	I	II	III	IV	V	VI	I	II	III	IV	V	VI
50	41.2	2.8	53.2	2.6	0.3	0.0	41.3	2.6	53.5	2.5	0.1	0.0	41.5	2.6	53.6	2.2	0.0	0.0	42.1	2.3	53.8	1.7	0.1	0.0
100	41.7	1.4	55.5	1.3	0.1	0.0	41.8	1.2	55.7	1.2	0.0	*	41.9	1.2	55.7	1.1	0.0	0.0	41.9	1.2	55.9	1.1	0.0	*
200	41.8	0.7	56.8	0.7	0.0	*	42.0	0.4	57.1	0.5	0.0	*	41.8	0.6	57.1	0.6	0.0	*	41.7	0.6	57.2	0.5	0.0	*
500	42.1	0.3	57.3	0.2	0.0	*	41.9	0.1	57.8	0.2	*	*	42.0	0.2	57.6	0.2	0.0	*	42.3	0.2	57.4	0.1	0.0	*
1 000	42.2	0.1	57.6	0.1	0.0	*	42.2	0.1	57.7	0.0	0.0	*	42.0	0.1	57.9	0.1	*	*	42.2	0.1	57.6	0.1	0.0	*
2 000	42.4	0.1	57.5	0.1	*	*	42.3	0.0	57.6	0.0	*	*	42.2	0.0	57.8	0.0	*	*	42.0	0.0	57.9	0.0	0.0	*
5 000	42.1	0.0	57.8	0.0	*	*	42.0	0.0	58.0	0.0	*	*	42.3	0.0	57.7	0.0	*	*	42.2	0.0	57.8	0.0	0.0	*
10 000	42.2	0.0	57.8	0.0	0.0	*	42.3	0.0	57.7	0.0	*	*	42.2	0.0	57.8	0.0	*	*	42.1	0.0	57.9	0.0	*	*
20 000	42.0	0.0	57.9	0.0	*	*	42.0	0.0	58.0	0.0	*	*	42.2	0.0	57.8	0.0	*	*	42.4	0.0	57.6	*	*	*
50 000	42.3	0.0	57.7	0.0	*	*	41.8	*	58.2	*	*	*	42.2	*	57.8	0.0	*	*	42.2	0.0	57.8	*	*	*
100 000	42.0	*	58.0	*	*	*	42.2	0.0	57.8	*	*	*	42.7	*	57.3	*	*	*	42.2	*	57.8	*	*	*

Obs.: 0.0% [*] indica que um número desprezável (*i.e.*, entre 0 e 50) [exatamente 0] de réplicas de Monte Carlo (das 100 000) apresentaram determinado caso.

Tabela 1.17: Estimativas de Monte Carlo dos valores esperados dos EMV de $\theta_{+1} - \theta_{1+}$ sob a ACC e sob o mecanismo MNAR das estruturas A1-D1 (Tabela 1.13) para os casos I e III (Tabela 1.12)

n_{+++}	ACC						MNAR									
	A1		B1		C1		D1		A1		B1		C1		D1	
	I	III	I	III	I	III	I	III	I	III	I	III	I	III	I	III
50	-0.232	-0.210	-0.073	-0.035	-0.140	-0.208	0.010	0.044	-0.149	-0.072	0.012	0.170	-0.002	0.082	0.050	0.074
100	-0.233	-0.207	-0.069	-0.036	-0.145	-0.207	0.013	0.043	-0.122	0.001	0.043	0.248	0.034	0.171	0.063	0.106
200	-0.231	-0.205	-0.065	-0.037	-0.151	-0.205	0.015	0.042	-0.091	0.082	0.073	0.278	0.079	0.254	0.085	0.152
500	-0.227	-0.205	-0.062	-0.041	-0.160	-0.197	0.017	0.043	-0.047	0.145	0.097	0.280	0.133	0.289	0.125	0.254
1 000	-0.225	-0.208	-0.060	-0.043	-0.165	-0.192	0.020	0.042	-0.021	0.150	0.104	0.280	0.162	0.290	0.167	0.337
2 000	-0.223	-0.210	-0.060	-0.045	-0.168	-0.189	0.023	0.040	-0.005	0.150	0.104	0.281	0.182	0.290	0.210	0.383
5 000	-0.222	-0.212	-0.060	-0.034	-0.171	-0.185	0.026	0.037	0.003	0.150	0.102	0.286	0.196	0.290	0.255	0.390
10 000	-0.222	-0.211	-0.060	*	-0.171	-0.185	0.027	0.035	0.002	0.150	0.101	*	0.200	0.290	0.277	0.390
20 000	-0.222	*	-0.060	*	-0.171	-0.183	0.028	0.034	0.001	*	0.100	*	0.201	0.290	0.291	0.390
50 000	-0.222	*	-0.060	*	-0.171	*	0.028	0.034	0.000	*	0.100	*	0.200	*	0.299	0.390
100 000	-0.222	*	-0.060	*	-0.171	*	0.029	0.033	0.000	*	0.100	*	0.200	*	0.301	0.390
Teórico†	-0.222		-0.060		-0.171		0.029		0.000		0.100		0.200		0.300	

*Nenhuma das réplicas de Monte Carlo (das 100 000) apresentaram determinado caso (Tabela 1.15).

†Sob a ACC, o valor teórico esperado é obtido por $P(Y_2 = 1|W = 1) - P(Y_1 = 1|W = 1)$ e, sob a estrutura MNAR, por $\theta_{+1} - \theta_{1+}$.

Tabela 1.18: Estimativas de Monte Carlo dos valores esperados dos EMV de $\theta_{+1} - \theta_{1+}$ sob a ACC e sob o mecanismo MNAR das estruturas A2-D2 (Tabela 1.13) para os casos I e III (Tabela 1.12)

n_{+++}	ACC						MNAR									
	A2		B2		C2		D2		A2		B2		C2		D2	
	I	III	I	III	I	III	I	III	I	III	I	III	I	III	I	III
50	-0.279	-0.276	-0.181	-0.174	-0.084	-0.073	0.011	0.029	-0.182	-0.148	-0.079	-0.049	0.018	0.049	0.118	0.139
100	-0.280	-0.276	-0.180	-0.177	-0.080	-0.076	0.018	0.025	-0.176	-0.137	-0.075	-0.036	0.025	0.063	0.125	0.161
200	-0.278	-0.278	-0.179	-0.177	-0.079	-0.077	0.021	0.024	-0.172	-0.131	-0.071	-0.030	0.027	0.070	0.128	0.170
500	-0.278	-0.278	-0.178	-0.177	-0.078	-0.077	0.022	0.023	-0.170	-0.128	-0.071	-0.024	0.029	0.073	0.129	0.173
1 000	-0.278	-0.278	-0.178	-0.177	-0.078	-0.077	0.022	0.023	-0.170	-0.126	-0.070	-0.025	0.031	0.073	0.130	0.174
2 000	-0.278	-0.278	-0.178	-0.178	-0.078	-0.078	0.022	0.022	-0.171	-0.124	-0.071	-0.024	0.030	0.076	0.130	0.175
5 000	-0.278	-0.278	-0.178	-0.178	-0.078	-0.078	0.022	0.022	-0.172	-0.126	-0.070	-0.026	0.030	0.074	0.130	0.176
10 000	-0.278	-0.278	-0.178	-0.178	-0.078	-0.078	0.022	0.022	-0.169	-0.125	-0.072	-0.026	0.031	0.074	0.130	0.176
20 000	-0.278	-0.278	-0.178	-0.178	-0.078	-0.078	0.022	0.022	-0.171	-0.127	-0.069	-0.025	0.030	0.075	0.129	0.173
50 000	-0.278	-0.278	-0.178	-0.178	-0.078	-0.078	0.022	0.022	-0.171	-0.125	-0.069	-0.025	0.030	0.076	0.130	0.176
100 000	-0.278	-0.278	-0.178	-0.178	-0.078	-0.078	0.022	0.022	-0.170	-0.129	-0.070	-0.026	0.030	0.074	0.131	0.174
Teórico†	-0.278		-0.178		-0.078		0.022		0.000		0.100		0.200		0.300	

†Sob a ACC, o valor teórico esperado é obtido por $P(Y_2 = 1|W = 1) - P(Y_1 = 1|W = 1)$ e, sob a estrutura MNAR, por $\theta_{+1} - \theta_{1+}$.

Pela Tabela 1.13, percebe-se que $\theta_{+1} - \theta_{1+}$ é igual a 0.000, 0.100, 0.200 e 0.300 para as estruturas Ai , Bi , Ci e Di , $i = 1, 2$. Analisando as Tabelas 1.17 e 1.18, nota-se que sob a ACC essas quantidades são estimadas enviesadamente, quer no caso I, quer no caso III, pois $P(Y_2 = 1|W = 1) - P(Y_1 = 1|W = 1)$ é igual a $-0.222/-0.278$, $-0.060/-0.178$, $-0.171/-0.078$ e $0.029/0.022$, respectivamente, para as estruturas A1/A2–D1/D2. Assim, os EMV sob a ACC estão estimando quantidades diferentes daquelas pretendidas.

Especificamente quando não há problemas de identificabilidade (Tabela 1.17), os EMV sob o mecanismo MNAR são enviesados (com viés entre 0.1% e 25.0%) para pequenas amostras ($n_{+++} = 50, 100$), na mesma direção dos vieses dos EMV obtidos sob a ACC, mas com magnitudes menores. Quando a solução das equações de verossimilhança está dentro do espaço paramétrico (caso I), estes vieses diminuem conforme o tamanho da amostra cresce, com a mesma velocidade da ordem das estruturas A1–D1 obtida anteriormente, indicando que quanto maior é a probabilidade de obter uma solução dentro do espaço paramétrico, menor é o tamanho da amostra necessário para se diminuir o viés. Apesar de o número de soluções das equações de verossimilhança com α_1 ou α_2 na fronteira do espaço paramétrico (caso III) tender a zero, assintoticamente, os EMV são enviesados (com viés entre 8.6% e 15.0%) mesmo para grandes tamanhos de amostra ($n_{+++} \geq 5\,000$).

Quando há problemas de identificabilidade, como previamente discutido, o EMV deixa de ser consistente e isso é refletido nos resultados obtidos sob a estrutura MNAR da Tabela 1.18. No entanto, é interessante notar que, embora esse estimador seja assintoticamente enviesado, o respectivo viés ainda assim é sempre menor que o obtido sob a ACC.

As estruturas Ai (Bi , Ci e Di), $i = 1, 2$, estão associadas a uma diferença nula (positiva) entre θ_{+1} e θ_{1+} , indicando que as probabilidades de se rejeitar a hipótese de homogeneidade marginal nas Tabelas 1.19 e 1.20 refletem os níveis de significância (poderes) empíricos dos TRV.

Quando não há problemas de identificabilidade (Tabela 1.19), se a solução das equações de verossimilhança estiver dentro do espaço paramétrico (caso I), apenas a partir do tamanho de amostra 10 000 se verifica que os níveis de significância empíricos da estrutura A1 atingem o nível de 5% sob o mecanismo MNAR; da mesma forma, somente a partir dos tamanhos de amostra entre 500 e 1 000 se averigua que os poderes empíricos das estruturas B1–D1 sob o mecanismo MNAR superam o nível de significância adotado. Tendo em vista que as estimativas dos valores esperados dos EMV sob o mecanismo MNAR são sempre maiores no

Tabela 1.19: Estimativas de Monte Carlo das probabilidades de rejeições à homogeneidade marginal sob a ACC e sob o mecanismo MNAR das estruturas A1–D1 (Tabela 1.13) para os casos I e III (Tabela 1.12)

n_{+++}	ACC						MNAR									
	A1		B1		C1		D1		A1		B1		C1		D1	
	I	III	I	III	I	III	I	III	I	III	I	III	I	III	I	III
50	0.552	0.374	0.190	0.088	0.095	0.273	0.058	0.106	0.038	0.049	0.023	0.103	0.010	0.071	0.015	0.060
100	0.796	0.573	0.217	0.076	0.163	0.383	0.044	0.089	0.037	0.032	0.024	0.203	0.011	0.118	0.010	0.064
200	0.970	0.880	0.294	0.095	0.338	0.629	0.040	0.098	0.034	0.021	0.045	0.483	0.021	0.267	0.012	0.095
500	1.000	0.999	0.558	0.237	0.773	0.937	0.049	0.140	0.033	0.037	0.209	0.953	0.123	0.719	0.036	0.223
1 000	1.000	1.000	0.832	0.518	0.979	0.997	0.077	0.216	0.030	1.000	0.459	1.000	0.428	0.978	0.104	0.454
2 000	1.000	1.000	0.984	0.824	1.000	1.000	0.140	0.349	0.029	1.000	0.759	1.000	0.819	1.000	0.305	1.000
5 000	1.000	1.000	1.000	1.000	1.000	1.000	0.371	0.662	0.038	1.000	0.987	1.000	0.997	1.000	0.811	1.000
10 000	1.000	1.000	1.000	*	1.000	1.000	0.693	0.907	0.050	1.000	1.000	*	1.000	1.000	0.988	1.000
20 000	1.000	*	1.000	*	1.000	1.000	0.951	0.996	0.050	*	1.000	*	1.000	1.000	1.000	1.000
50 000	1.000	*	1.000	*	1.000	*	1.000	1.000	0.049	*	1.000	*	1.000	*	1.000	1.000
100 000	1.000	*	1.000	*	1.000	*	1.000	1.000	0.050	*	1.000	*	1.000	*	1.000	1.000

*Nenhuma das réplicas de Monte Carlo (das 100 000) apresentaram determinado caso (Tabela 1.15).

Tabela 1.20: Estimativas de Monte Carlo das probabilidades de rejeições à homogeneidade marginal sob a ACC e sob o mecanismo MNAR das estruturas A2-D2 (Tabela 1.13) para os casos I e III (Tabela 1.12)

n_{+++}	ACC						MNAR									
	A2		B2		C2		D2		A2		B2		C2		D2	
	I	III	I	III	I	III	I	III	I	III	I	III	I	III	I	III
50	0.548	0.530	0.281	0.301	0.108	0.137	0.064	0.100	0.028	0.055	0.011	0.044	0.007	0.033	0.022	0.041
100	0.829	0.764	0.457	0.450	0.121	0.158	0.048	0.082	0.022	0.044	0.009	0.039	0.003	0.035	0.013	0.034
200	0.987	0.968	0.766	0.714	0.201	0.238	0.048	0.084	0.017	0.040	0.006	0.033	0.002	0.032	0.011	0.034
500	1.000	1.000	0.992	0.978	0.471	0.467	0.070	0.108	0.015	0.043	0.005	0.032	0.001	0.030	0.009	0.037
1 000	1.000	1.000	1.000	1.000	0.789	0.735	0.107	0.150	0.015	0.536	0.005	0.494	0.001	0.114	0.009	0.496
2 000	1.000	1.000	1.000	1.000	0.978	0.955	0.188	0.229	0.015	0.999	0.004	1.000	0.001	1.000	0.010	1.000
5 000	1.000	1.000	1.000	1.000	1.000	1.000	0.437	0.451	0.015	1.000	0.004	1.000	0.001	1.000	0.009	1.000
10 000	1.000	1.000	1.000	1.000	1.000	1.000	0.752	0.723	0.015	1.000	0.004	1.000	0.001	1.000	0.009	1.000
20 000	1.000	1.000	1.000	1.000	1.000	1.000	0.967	0.945	0.015	1.000	0.004	1.000	0.000	1.000	0.009	1.000
50 000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.015	1.000	0.004	1.000	0.001	1.000	0.010	1.000
100 000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.014	1.000	0.004	1.000	0.000	1.000	0.009	1.000

caso III do que no caso I (veja a Tabela 1.17), é natural que os poderes também sejam maiores no caso III para as estruturas B1–D1 e que a estrutura A1 sempre leve à rejeição da hipótese de homogeneidade marginal para grandes tamanhos de amostra ($n_{+++} = 1\,000$ a $10\,000$), quando isto deveria ocorrer apenas em cerca de 5% dos casos. Pela mesma razão de diferença entre os valores esperados e estimados, a comparação entre as probabilidades de rejeição da homogeneidade marginal sob a ACC e sob o mecanismo MNAR é dificultada. Contudo, os valores esperados e estimados de $\theta_{+1} - \theta_{1+}$ para $n_{+++} = 50$ da estrutura D1 (Tabela 1.17) nos casos I e III sob a ACC são iguais a 0.010 e 0.044 e, sob o mecanismo MNAR, são iguais a 0.050 e 0.074, mas, mesmo os valores sob a ACC estando mais próximos de zero do que aqueles obtidos sob o mecanismo MNAR, os poderes alcançados sob a ACC são maiores do que os obtidos sob o mecanismo MNAR — os poderes nos casos I e III (Tabela 1.19) são iguais a 0.058 e 0.106 sob a ACC e 0.015 e 0.060 sob o mecanismo MNAR. Apenas a partir de $n_{+++} = 1\,000$ os poderes sob o mecanismo MNAR se tornam maiores que aqueles obtidos sob a ACC para o caso I da estrutura D1. Isto ilustra a enorme incerteza adicionada aos estimadores dos parâmetros com a suposição de que a omissão depende de categorias não-observadas.

Nota-se, na Tabela 1.20, que quando há problemas de identificabilidade, a obtenção de soluções das equações de verossimilhança dentro do espaço paramétrico (caso I), em geral, implica a não rejeição da hipótese de homogeneidade marginal sob o mecanismo MNAR. Se α_1 , ou α_2 , estiver na fronteira do espaço paramétrico (caso III), o mesmo acontece para $n_{+++} \leq 500$, enquanto que quase sempre se rejeita a hipótese para $n_{+++} \geq 2\,000$.

1.3 Seleção da estrutura para o mecanismo de omissão

Murray e Findlay (1988) descreveram um estudo de hipertensão, em que por motivos éticos, quando a pressão sanguínea de um paciente excedia um limite determinado no protocolo por duas ou mais semanas, ele era retirado do estudo e o tratamento era modificado para atender às suas necessidades. Sendo a saída do estudo uma decorrência dos valores previamente obtidos, a adoção do mecanismo MAR se justifica pelo planejamento experimental. No entanto, nem sempre o desenho do estudo justifica a adoção de um mecanismo de omissão.

Quando o ajuste de estruturas saturadas MNAR é perfeito, as correspondentes frequências estimadas diferem daquelas obtidas sob a estrutura MAR apenas com relação às frequências ampliadas não-observadas (as interrogações da Tabela 1.4). Isso mostra que não é possível

realizar comparações de ajuste entre modelos saturados. Pode-se apenas realizar testes de ajuste de modelos reduzidos condicionalmente ao ajuste dos modelos saturados, como testar se um mecanismo MCAR se ajusta bem dado um modelo MAR ou MNAR. Mas, como o valor máximo da verossimilhança obtido sob uma estrutura saturada MNAR, que possui o ajuste perfeito, corresponde àquele obtido sob a estrutura MAR, os testes de ajuste do mecanismo MCAR condicionalmente a estas duas estruturas terão o mesmo resultado e, novamente, nenhuma das duas será preferida. Os testes terão resultados diferentes apenas se o ajuste da estrutura saturada MNAR não for perfeito (Seção 1.2.3). Neste caso, o valor máximo da verossimilhança obtido sob a estrutura MAR será maior que o da MNAR e, assim, pode ser que a estrutura MCAR condicionalmente à MAR seja rejeitada, mas a MCAR condicionalmente à MNAR, não, o que favoreceria a estrutura MAR em detrimento da MNAR. Portanto, a não ser que uma amostra das unidades com omissão seja investigada — prática que eliminaria parte da falta de informação sobre o mecanismo de omissão, veja, *e.g.*, Glynn, Laird e Rubin (1986) e Lyles e Allen (2003) —, a escolha da estrutura para o mecanismo de omissão, em geral, baseia-se em suposições inverificáveis.

O acompanhamento do estudo e a tentativa de investigar o motivo da ocorrência da omissão podem ajudar na proposição de modelos “mais adequados” para o problema. Para esses modelos sugere-se realizar uma análise de sensibilidade verificando:

1. a estabilidade das inferências de interesse e
2. a plausibilidade dos valores esperados estimados para as frequências ampliadas,

$$E(\widehat{n_{tij}}) = n_{+++} \hat{\theta}_{ij} \lambda_{t(ij)}(\hat{\alpha}). \quad (1.41)$$

Baker, Ko e Graubard (2003) revisaram diversos artigos em que esta análise de sensibilidade é adotada e concluíram que, para pequenos conjuntos de dados com poucas variáveis, os intervalos de confiança para os parâmetros de interesse sob diferentes estruturas do mecanismo de omissão costumam ter um alto grau de sobreposição, indicando que a análise de sensibilidade não traz informação adicional nestes casos. O tamanho da amostra, o número de variáveis e o grau de sobreposição não foram quantificados, embora os autores tenham apresentado uma tabela comparativa com estas informações para os artigos analisados.

As Seções 1.2.1 e 1.2.2 mostraram que a ACC: (i) sob o mecanismo MAR [MCAR], [não] leva, em geral, a inferências enviesadas sobre θ ; (ii) sob o mecanismo MCAR, implica erros padrões mais elevados para os estimadores de θ quando todos os valores esperados das

frequências observáveis da tabela completamente categorizada forem maiores que 1.5. Os resultados sugerem que, mesmo para tamanhos amostrais superiores a 10 unidades, as complicações adicionais nos planos teórico e computacional são compensadoras. Assim, análises sob as estruturas MAR/MCAR protegem os resultados de possíveis vieses que poderiam ocorrer em ACC se o mecanismo verdadeiro for o MAR e ainda produzem estimadores mais eficientes se o mecanismo for o MCAR.

Os casos particulares utilizados nas simulações da Seção 1.2.3 sob o mecanismo MNAR verdadeiro sugerem que, dependendo da estrutura utilizada, B1 ou D1, o tamanho da amostra necessário para se ter um viés menor que 1% na estimação de $\theta_{+1} - \theta_{1+}$ pode variar de 500 a 5 000 (Tabela 1.17) se os EMV estiverem dentro do espaço paramétrico e os parâmetros forem identificáveis, situações em que os TRV aparentam se comportar assintoticamente “bem” (atingir o nível de significância desejado e aumentar o poder, tendendo a 100%). O estudo de simulação também mostra que quando a estimativa de uma das probabilidades condicionais de omissão estiver na fronteira do espaço paramétrico e/ou os parâmetros não forem identificáveis, os EMV são enviesados para qualquer tamanho de amostra, embora sempre com viés menor que os correspondentes obtidos sob a ACC, e os TRV têm comportamentos indesejáveis, principalmente no caso de estruturas com parâmetros inidentificáveis.

Quando não há motivos explícitos para se preferir uma das estruturas de omissão ajustadas e ainda existem dúvidas de que o mecanismo de omissão possa ter uma estrutura mais complexa do que aquela passível de estimação com os dados disponíveis, alguns autores, dentre os quais destaca-se Rosenbaum e Rubin (1983), Nordheim (1984), Phillips (1993) e Copas e Eguchi (2001), preferem realizar outro tipo de análise de sensibilidade, mais tarde formalizada por Kenward, Goetghebeur e Molenberghs (2001) e Molenberghs, Kenward e Goetghebeur (2001) e que se descreve a seguir. Nesse contexto, os autores distinguiram dois tipos de **incertezas estatísticas**: a **imprecisão estatística**, devido à amostragem, e a **ignorância estatística**, causada pela omissão. Quando o tamanho da amostra tende ao infinito, a imprecisão estatística das estimativas dos parâmetros desaparece e o que resta é a ignorância com relação à distribuição das unidades com omissão. A imprecisão estatística pode ser capturada por erros padrões e regiões de confiança. Os autores sugerem que a ignorância estatística deve ser avaliada por **regiões de ignorância** e **regiões de incerteza**. As primeiras contêm a região de valores plausíveis para as estimativas pontuais sob o ajuste de diversas estruturas para o mecanismo de omissão e as demais atuam como uma região maior capturando os efeitos combinados de imprecisão e ignorância. Vansteelandt, Goetghebeur, Kenward e Molenberghs

(2003) estudaram propriedades de tais regiões.

Regiões de ignorância e incerteza são obtidas por meio de modelos sobre-parametrizados. Primeiramente, escolhe-se um grupo de parâmetros dentre os associados às probabilidades marginais de categorização e às probabilidades condicionais de omissão. A seguir, particiona-se esse grupo de parâmetros em $(\boldsymbol{\mu}, \boldsymbol{\tau})$, de tal forma que a dimensão de $\boldsymbol{\mu}$ seja igual a do número de graus de liberdade dos dados observados e $\boldsymbol{\tau}$ tenha um ou mais parâmetros selecionados dentre os restantes. Cada valor fixado de $\boldsymbol{\tau}$ produz uma estimativa $\hat{\boldsymbol{\mu}}(\boldsymbol{\tau})$ de $\boldsymbol{\mu}$ e uma região de $100(1 - \alpha)\%$ confiança. A união destas estimativas pontuais e regiões de confiança, obtidas quando se varia $\boldsymbol{\tau}$, gera, respectivamente, a região de ignorância e de $100(1 - \alpha)\%$ incerteza para $\boldsymbol{\mu}$. $\boldsymbol{\tau}$ é chamado de **parâmetro de sensibilidade** e $\boldsymbol{\mu}$, de **parâmetro estimável**. Existindo o interesse marginal em um único parâmetro, o termo região é substituído por intervalo, como de costume.

Os autores ainda sugerem que $\boldsymbol{\mu}$ seja selecionado de algum modelo saturado e que se especifique um ou mais parâmetros extras em $\boldsymbol{\tau}$, desde que o modelo não esteja mal ajustado ou com estimativas na fronteira do espaço paramétrico uma vez que se deseja dar atenção à ignorância e não à imprecisão. Na Seção 1.2.3, viu-se que, mesmo quando uma estrutura MNAR saturada é verdadeira, a probabilidade de se obter estimativas na fronteira do espaço paramétrico não é desprezável. Sendo assim, rejeitar estas estruturas não parece ser uma prática adequada. De qualquer forma, como a imprecisão estatística é capturada apenas para os parâmetros estimáveis, sugere-se que $\boldsymbol{\mu}$ contenha pelo menos os parâmetros de interesse, ou seja, os associados às probabilidades marginais de categorização. Kenward, Goetghebeur e Molenberghs (2001) estudaram o caso teórico da Tabela 1.3 e exemplificaram que apenas 1 ou 2 parâmetros em $\boldsymbol{\tau}$ já são suficientes para se obter informação adicional sobre os parâmetros/inferências de interesse, apesar de que para esse problema, $\boldsymbol{\tau}$ poderia conter até 7 parâmetros. Análises de sensibilidade sob esses moldes também foram apresentadas por Vansteelandt e Goetghebeur (2001, 2004) e Verzilli e Carpenter (2002).

Kenward, Goetghebeur e Molenberghs (2001) e Molenberghs, Kenward e Goetghebeur (2001) ainda ressaltaram que o intervalo para o melhor-pior caso (*best-worst case interval*), alocando as unidades omissas em categorias que produzam casos extremos para as inferências de interesse, apesar de ter grande amplitude, é um método simples, bastante informativo e um ponto de partida honesto para uma modelagem cautelosa. No entanto, as regiões de ignorância devem ter amplitudes menores que este intervalo nos casos em que se têm ajustes

perfeitos. Para ilustrar o uso deste intervalo no estudo apresentado no Exemplo 1, pode-se alocar as unidades omissas da Tabela 1.1 conforme a situação extrema A (B) ilustrada na Tabela 1.21 que produz a menor (maior) estimativa possível para a razão de chances (o quociente entre a chance de recém-nascidos de mães não-fumantes terem pesos ≥ 2.5 kg e a chance correspondente para as mães fumantes). O intervalo para o melhor-pior caso para a razão de chances é (0.82; 2.50).

Tabela 1.21: Alocações extremas para as unidades omissas da Tabela 1.1

Alo- cação	MFum PesoRN	$W = 1$		$W = 2$		$W = 3$		$W = 4$		Total	
		< 2.5	≥ 2.5	< 2.5	≥ 2.5	< 2.5	≥ 2.5	< 2.5	≥ 2.5	< 2.5	≥ 2.5
A	sim	4 512	21 009	0	1 049	0	464	0	0	4 512	22 522
	não	3 394	24 132	1 135	0	142	0	1 224	0	5 895	24 132
B	sim	4 512	21 009	1 049	0	142	0	1 224	0	6 927	21 009
	não	3 394	24 132	0	1 135	0	464	0	0	3 394	25 731

MFum: Mãe fumante, PesoRN: Peso do recém-nascido (kg).

1.4 Exemplo de análise

Os dados apresentados no Exemplo 1 (pp.1-2), Tabela 1.1, são aqui analisados com o propósito de ilustrar as idéias apresentadas neste capítulo. Por conveniência, a notação das seções anteriores será mantida, sendo Y_1 (Y_2) a variável representativa do hábito de fumo da mãe (peso do recém-nascido), com valor 1 para fumante (< 2.5 kg) ou 2 para não-fumante (≥ 2.5 kg).

Devido à falta de informação, que pudesse auxiliar a proposição de estruturas plausíveis para o mecanismo de omissão, considera-se, arbitrariamente, as estruturas MCAR (Seção 1.2.2), MAR_{red} (1.17), MAR (Seção 1.2.1), MNAR1 (1.16)¹⁹ e MNAR2, definida por

$$\begin{aligned}
 \text{logito}(\psi_{1(ij)}) &= \alpha_{10} + \alpha_1(i-1) + \alpha_2(j-1) + \alpha_3(i-1)(j-1), \\
 \text{logito}(\psi_{21(ij)}) &= \alpha_{20} + \alpha_1(i-1) + \alpha_2(j-1) + \alpha_3(i-1)(j-1), \\
 \text{logito}(\psi_{20(ij)}) &= \alpha_{20} + \alpha_1(i-1) + \alpha_2(j-1) + \alpha_3(i-1)(j-1).
 \end{aligned}
 \tag{1.42}$$

¹⁹Veja a função de verossimilhança sob a estrutura MNAR1 em (2.91), p.87.

Note que a estrutura MNAR2 é obtida da MNAR1 impondo a restrição $\alpha_{30} = \alpha_{20}$ e adicionando o termo $\alpha_3(i-1)(j-1)$.

A Tabela 1.22 contém as EMV para a razão de chances e intervalos de 95% de confiança tanto sob uma ACC quanto sob as estruturas para o mecanismo de omissão mencionadas. Os intervalos de confiança do tipo Wald foram construídos com os erros padrões obtidos da estimativa da matriz de informação observada. A Tabela 1.23 ilustra os dados observados e as freqüências ampliadas esperadas estimadas sob cada uma dessas estruturas dos mecanismos de omissão.

Tabela 1.22: EMV e intervalos de 95% de confiança para a razão de chances

ACC* / MCAR / MAR _{red} / MAR	1.53	(1.46; 1.60)
MNAR1	1.50	(1.42; 1.57)
MNAR2	0.83	(0.79; 0.86)

*Em geral, os resultados sob a ACC não coincidem com os obtidos sob mecanismos MAR.

Pode-se observar que:

- As diferenças entre MAR e MAR_{red} são maiores sob os padrões de omissão $W = 1$ e $W = 3$, sendo uma decorrência de que sob a estrutura MAR_{red} as probabilidades condicionais $\{\lambda_{3(ij)}\}$ são constantes e $\{\lambda_{1(ij)}\}$ dependem apenas de i . Da mesma forma, como as probabilidades condicionais $\{\lambda_{1(ij)}\}$ e $\{\lambda_{2(ij)}\}$ (não) dependem de i sob a estrutura MAR_{red} (MCAR), é de se esperar que as duas estruturas tenham maiores discordâncias nos padrões de omissão $W = 1$ e $W = 2$. As estruturas MCAR e MAR_{red} (únicas não saturadas) não se ajustam aos dados ($p < 0.001$) segundo o TRV. Apesar dessas diferenças, note que as somas das freqüências estimadas dos quatro padrões de omissão coincidem sob as estruturas MCAR, MAR_{red} e MAR, sendo uma conseqüência de que as EMV de θ sob qualquer mecanismo mais restritivo que o MAR sempre são iguais, conforme evidenciado nas fatorações explicitadas em (1.8) e (1.12).
- Como era de se esperar, as estruturas saturadas MAR e MNAR1 geram freqüências estimadas exatamente iguais às observadas, discordando apenas onde houve omissão. A estrutura MNAR1 aloca as unidades com omissão em Y_2 mais nas celas correspondentes

Tabela 1.23: Dados observados e EMV das frequências ampliadas esperadas*

Modelo	MFum	$W = 1 (Y_1^o, Y_2^o)$		$W = 2 (Y_1^o, Y_2^m)$		$W = 3 (Y_1^m, Y_2^o)$		$W = 4 (Y_1^m, Y_2^m)$		Total		
		< 2.5	≥ 2.5	< 2.5	≥ 2.5	< 2.5	≥ 2.5	< 2.5	≥ 2.5		< 2.5	≥ 2.5
Dados	sim	4 512	21 009	?	? 1 049	?	?	?	?	?	?	?
obser-	não	3 394	24 132	?	? 1 135	?	?	?	?	?	?	?
vados	total					142	464			1 224		57 061
MCAR	sim	4 541	20 984	187	864 1 051	52	240	105	484	4 884	22 571	27 456
	não	3 416	24 106	141	992 1 133	39	275	79	556	3 675	25 930	29 605
	total					91	515			1 224		57 061
MAR _{red}	sim	4 541	20 985	187	863 1 049	52	240	105	484	4 884	22 571	27 456
	não	3 416	24 105	141	994 1 135	39	275	79	556	3 675	25 930	29 605
	total					91	515			1 224		57 061
MAR	sim	4 512	21 009	187	862 1 049	81	216	105	484	4 884	22 571	27 456
	não	3 394	24 132	141	994 1 135	61	248	79	556	3 675	25 930	29 605
	total					142	464			1 224		57 061
MNAR1	sim	4 512	21 009	525	524 1 049	80	208	379	212	5 496	21 954	27 449
	não	3 394	24 132	450	685 1 135	62	256	336	296	4 242	25 369	29 612
	total					142	464			1 224		57 061
MNAR2	sim	4 512	20 977	0	1 088 1 088	0	489	0	25	4 512	22 579	27 091
	não	2 744	24 132	1 778	0 1 778	799	0	517	0	5 838	24 132	29 970
	total					799	489			543		57 061

MFum: Mãe fumante, PesoRN: Peso do recém-nascido (kg).

*Alguns totais não coincidem com as somas das frequências respectivas devido aos arredondamentos.

a $Y_2 = 1$ do que naquelas correspondentes a $Y_2 = 2$ em relação à estrutura MAR, corroborando com a especulação (3) descrita no Exemplo 1. Em compensação, não há grandes diferenças entre as duas estruturas onde houve omissão em Y_1 .

- A estrutura MNAR2, apesar de ser saturada, não gerou frequências estimadas exatamente iguais às observadas e teve algumas estimativas dos $\{\lambda_{t(ij)}\}$ na fronteira do espaço paramétrico, conforme se pode perceber pelas frequências estimadas nulas. Conseqüentemente, o valor da estatística de razão de verossimilhanças do ajuste do mecanismo é de 1 863.77 apesar de ela ter zero graus de liberdade. As frequências estimadas têm uma certa semelhança com a alocação extrema A da Tabela 1.21. Apesar de se saber (Seção 1.2.3) que uma estrutura MNAR verdadeira poderia ter estimativas na fronteira do espaço paramétrico, o tamanho de amostra grande ($>50\ 000$) indica que isso não deveria ocorrer facilmente. Além disso, a conclusão obtida por essa estrutura MNAR, de que recém-nascidos de mães fumantes têm maior probabilidade de nascerem com ≥ 2.5 kg do que de mães não-fumantes, contradiz o esperado, indicando que esse mecanismo de omissão não deve ter gerado os dados omissos.

Considere agora a estrutura sobre-parametrizada

$$\begin{aligned}\text{logito}(\psi_{1(ij)}) &= \alpha_{10} + \alpha_1(i-1) + \alpha_2(j-1) + \alpha_3(i-1)(j-1), \\ \text{logito}(\psi_{21(ij)}) &= \alpha_{20} + \alpha_1(i-1) + \alpha_2(j-1) + \alpha_3(i-1)(j-1), \\ \text{logito}(\psi_{20(ij)}) &= \alpha_{30} + \alpha_1(i-1) + \alpha_2(j-1) + \alpha_3(i-1)(j-1)\end{aligned}\tag{1.43}$$

para o mecanismo de omissão, com $\boldsymbol{\mu} = (\boldsymbol{\theta}, \alpha_{10}, \alpha_{20}, \alpha_{30}, \alpha_1, \alpha_2)$ representando o parâmetro estimável e $\tau = \alpha_3$ denotando o parâmetro de sensibilidade. Aqui, α_3 é o acréscimo nos logitos das probabilidades de se observar Y_1, Y_2 dado que Y_1 foi observado e Y_2 dado que Y_1 foi omissos quando $Y_1 = 2$ e $Y_2 = 2$. Portanto, quanto maior é α_3 , maiores são as chances de se observar Y_1 e Y_2 quando tiverem conjuntamente o valor 2. Conseqüentemente, menos unidades com omissão são alocadas em $(Y_1 = 2, Y_2 = 2)$ e menores são as razões de chance obtidas. As EMV e intervalos de 95% de confiança para a razão de chances variando α_3 de -5 a 5 estão ilustradas na Figura 1.1.

O intervalo de ignorância para a razão de chances é $(0.94; 2.23)$, e o de 95% de incerteza, construído por meio da união dos intervalos de 95% de confiança de cada valor de α_3 , é $(0.90; 2.34)$.

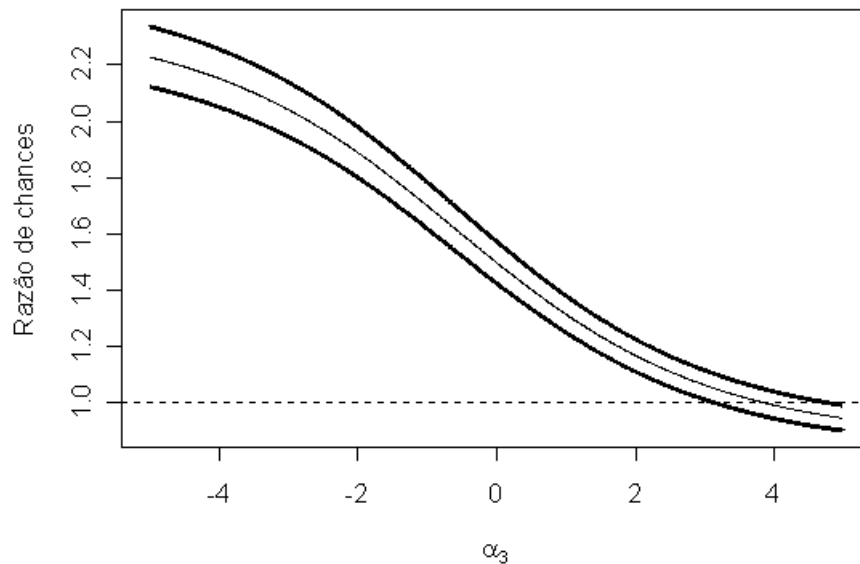


Figura 1.1: EMV e intervalos de 95% de confiança para a razão de chances variando α_3

Apesar de uma razão de chances menor que 1 ser improvável, conforme se discutiu previamente, a ignorância provocada pela omissão não nos permite inferir que recém-nascidos de mães não-fumantes têm maior probabilidade de nascerem com ≥ 2.5 kg do que de mães fumantes.

Baker *et al.* (1992) analisaram esse conjunto de dados ajustando diversas estruturas para o mecanismo de omissão, incluindo várias estruturas MNAR, e concluíram que a análise de sensibilidade aumentou suas confianças no resultado de que há uma associação inversa entre mães fumantes e um maior peso dos recém-nascidos e que isto poderia ser uma consequência da pequena fração da amostra com omissão, apenas 7% dos 57 061 pares de mães/recém-nascidos. Com este exemplo, mostra-se que mesmo com um tamanho de amostra “grande” ($> 50\,000$) com “pouca” omissão (7%) e quando as inferências de diferentes estruturas para o mecanismo de omissão têm conclusões concordantes, a ignorância sobre o mecanismo de omissão, em que o mecanismo de omissão poderia ter uma estrutura sobre-parametrizada, pode criar dúvidas sobre tais inferências. No entanto, se existisse um maior conhecimento sobre o mecanismo de omissão, a ignorância poderia ser diminuída.

1.5 Outras abordagens

Schaffer (1997), Kenward e Molenberghs (1999), Lindsey (1999), Allison (2001), Schafer e Graham (2002), Little e Rubin (2002) e Diggle, Heagerty, Liang e Zeger (2002) apresentaram revisões de modelos e/ou métodos utilizados para analisar dados com omissão, tanto com respostas categorizadas quanto contínuas. Fleiss, Levin e Paik (2003), Molenberghs e Verbeke (2005) e Paulino e Singer (2006) fizeram o mesmo, exclusivamente para respostas categorizadas, e Verbeke e Molenberghs (2000), para respostas contínuas. No contexto específico de não-resposta em pesquisas, destacam-se os livros editados por Groves, Dillman, Eltinge e Little (2002), Madow e Olkin (1983), Madow, Nisselson e Olkin (1983) e Madow, Olkin e Rubin (1983).

A metodologia GSK (Grizzle, Starmer e Koch, 1969) foi estendida por Koch, Imrey e Reinfurt (1972) e Woolson e Clarke (1984) para incorporar dados categorizados com omissão, utilizando estimadores de mínimos quadrados generalizados em dois estágios. Os primeiros assumiram que as unidades amostrais com omissão foram fixadas pelo planejamento amostral, enquanto que os últimos adicionaram uma categoria na variável resposta representando a omissão. Ambas propostas, que são válidas apenas sob o mecanismo MCAR, foram contrastadas por Lipsitz, Laird e Harrington (1994) com uma terceira abordagem, antes sugerida por Paulino (1991), que consiste em se obter no primeiro estágio EMV das probabilidades de categorização sob o mecanismo MAR e então utilizá-las juntamente com uma estimativa de sua matriz de covariâncias assintótica na etapa seguinte com a metodologia GSK. Ao comparar, em um estudo de simulação, vieses e erros quadráticos médios entre as três abordagens, concluíram que os EMV devem ser preferidos no primeiro estágio se o mecanismo de omissão depende do que foi observado e há mais de 15% de omissão, para evitar vieses; caso contrário, se o mecanismo de omissão não depende do que foi ou não observado, os EMV são tão eficientes quanto os estimadores obtidos pelo método de Koch, Imrey e Reinfurt (1972), e ambos são mais eficientes que os estimadores obtidos pelo método de Woolson e Clarke (1984).

Neste trabalho considera-se a modelagem da distribuição conjunta de respostas categorizadas. Se o interesse for apenas comparar distribuições marginais, pode ser mais atrativo utilizar a abordagem de equações de estimação generalizadas descrita em Liang e Zeger (1986), Venezuela (2003) e Artes e Botter (2005) para dados completos. Mark e Gail (1994), Fitzmaurice, Laird e Lipsitz (1994), Kenward, Lesaffre e Molenberghs (1994), Rotnitzky e Wypij (1994), Robins, Rotnitzky e Zhao (1994, 1995) e Robins e Rotnitzky (1995) mostraram que a ACC com métodos baseados em equações de estimação generalizadas têm resultados enviesados se o

mecanismo de omissão for MAR e propuseram correções. Chen e Little (1999) desenvolveram um teste Wald para avaliar a plausibilidade do mecanismo MCAR quando se deseja analisar os dados com equações de estimação generalizadas. Outras análises e propostas baseadas em equações de estimação sob mecanismos de omissão não-ignoráveis foram consideradas por Fitzmaurice, Molenberghs e Lipsitz (1995), Zhao, Lipsitz e Lew (1996), Rotnitzky e Robins (1997), Troxel, Lipsitz e Brennan (1997), Paik (1997), Wang (1999), Lipsitz, Ibrahim e Zhao (1999), Preisser, Galecki, Lohman e Wagenknecht (2000), Miller, Ten Have, Reboussin, Lohman e Rejeski (2001), FitzGerald (2002) e Ten Have, Joffe e Cary (2003). Comparações dos vieses de alguns desses métodos para dados binários são estudados em Fitzmaurice, Lipsitz, Molenberghs e Ibrahim (2001) e Preisser, Lohman e Rathouz (2002). Cook, Zeng e Yi (2004) compararam algumas dessas propostas com a imputação do último valor observado em estudos longitudinais (*last observation carried forward*, LOCF), mostrando que esta prática, em geral, produz estimadores enviesados. Baker (1995b), Fitzmaurice, Laird e Zahner (1996) e Fitzmaurice, Heath e Clifford (1996) utilizaram modelos marginais para dados binários com medidas repetidas, estimados por máxima verossimilhança, com mecanismos de omissão ignoráveis e não-ignoráveis.

Ten Have, Kunselman, Pulkstenis e Landis (1998) utilizaram um modelo logístico com efeitos aleatórios para analisar dados binários com omissão informativa e Ten Have, Miller, Reboussin e James (2000) fizeram o mesmo com dados ordinais. Lesaffre, Molenberghs e Dewulf (1996), Molenberghs, Kenward e Lesaffre (1997), Sheiner, Beal e Dunne (1997) e Toledano e Gatsonis (1999) realizaram análises de dados ordinais com medidas repetidas com mecanismos de omissão aleatória e informativa.

Neste capítulo não se utilizou variáveis explicativas ou covariáveis (\mathbf{X}). No próximo, elas são introduzidas sob a suposição de que são sempre observadas, uma vez que as inferências passam a se basear na distribuição das variáveis respostas condicional em \mathbf{X} , sugerindo que o vetor \mathbf{X} é conhecido ou fixado de antemão para cada uma das unidades experimentais. Para permitir que as covariáveis tenham omissão deve-se especificar pelo menos parte de sua distribuição marginal, além de se modelar o correspondente mecanismo de omissão, se este não for ignorável.

Schluchter e Jackson (1989) utilizaram modelos log-lineares para analisar dados de sobrevivência (Klein e Moeschberger, 2003; Colosimo e Giolo, 2006) com censura e covariáveis categorizadas parcialmente observadas com omissão aleatória e Baker (1994b) estendeu a análise considerando omissão informativa. Ibrahim (1990), Lipsitz e Ibrahim (1996) e Horton e Laird

(1998) ilustraram a utilização de covariáveis categorizadas com omissão em modelos lineares generalizados (Nelder e Wedderburn, 1972; McCullagh e Nelder, 1989; Paula, 2004; Cordeiro e Lima Neto, 2004) com omissão ignorável. Ibrahim e Lipsitz (1996) estenderam a metodologia de Ibrahim (1990) analisando dados binários com omissão não-ignorável. Ibrahim, Lipsitz e Chen (1999), Lipsitz, Ibrahim, Chen e Peterson (1999) e Ibrahim, Lipsitz e Horton (2001) fizeram o mesmo com a classe de modelos lineares generalizados. Lipsitz, Ibrahim e Fitzmaurice (1999) estudaram o caso em que há omissão simultaneamente na variável resposta e em covariáveis categorizadas. Blackhurst e Schluchter (1989), Vach e Schumacher (1993), Vach (1994, 1997), Vach e Blettner (1995) e Horton e Laird (2001) estudaram a incorporação de covariáveis omissas em regressão logística e Gibbons e Hosmer (1991) e Lipsitz, Parzen e Ewell (1998), em regressão logística condicional. Satten e Carroll (2000) e Paik (2004) sugeriram outros modelos para a análise de dados categorizados com covariáveis com omissão.

Uma outra forma de analisar dados omissos é imputar os valores ausentes e proceder a análise como se os dados estivessem completos. No entanto, essa prática ignora a incerteza que se tem sobre os valores desconhecidos. Uma maneira de contornar este inconveniente é imputar um conjunto de valores plausíveis para cada valor omissos, proceder as análises de cada banco de dados e, ao final, resumir as inferências obtidas por meio de ajustes especiais que incorporam a variabilidade associada às diversas imputações. Rubin (1978) propôs este método e chamou-o de imputação múltipla. Mais detalhes podem ser encontrados em Rubin (1987, 1996), Schafer (1997, 1999), Lipsitz, Zhao e Molenberghs (1998), King, Honaker, Joseph e Scheve (2001), Little e Rubin (2002), Schafer e Graham (2002) e West e Dawson (2002).

Basu e Pereira (1982), Gunel (1984), Dickey, Jyang e Kadane (1987), Kadane (1993), Forster e Smith (1998), Green e Park (2003) e Oleson e He (2004) ilustraram aplicações de métodos bayesianos na análise de dados categorizados com omissão. Ainda sob o enfoque bayesiano, Paulino e Pereira (1992, 1995), Walker (1996) e Soares e Paulino (2001) estudaram métodos apropriados para as inferências de interesse sem a imposição de restrições de identificabilidade nas probabilidades condicionais de censura. Soares (2004) generalizou essa abordagem e desenvolveu métodos de comparação bayesiana de mecanismos para o processo de omissão, considerando também o problema de erros de classificação (*misclassification*).

Vermunt (1996), Hagnaars e McCutcheon (2002), Reboussin, Miller, Lohman e Ten Have (2002) e Lin, McCulloch e Rosenheck (2004) estudaram dados categorizados com omissão em análises de classes latentes (Andersen, 1994), Holman e Glas (2005), em modelos de teoria de resposta ao item (Lindern e Hambleton, 1996; Andrade, Tavares e Valle, 2000; Azevedo, 2003)

e, Hancock e Mueller (2006), em modelagens de equações estruturais (Bollen, 1989; Hoyle, 1995; Latif, 2000; Melhado, 2004; Kline, 2005; Sanches, 2005). Outras propostas variadas de análises de dados categorizados com omissão foram feitas por Liang e Qin (2000), Baker (2000), Paik e Sacco (2000), Paik, Sacco e Lin (2000), Lin e Paik (2001), Qin, Leung e Shao (2002), Horton e Fitzmaurice (2002), Albert, Follmann, Wang e Suh (2002) e Scharfstein e Irizarry (2003).

Outro tipo particular de dados com omissão é encontrado rotineiramente em aplicações de escore de crédito (*credit scoring*, veja Rosa, 2000; Mays, 2001, 2003; Thomas, Edelman e Crook, 2002; Pereira, 2004). As instituições financeiras geralmente utilizam modelos estatísticos que mensuram o risco de concessão de crédito aos seus clientes, negando o benefício no caso de o cliente possuir uma pontuação indicativa de um alto risco de inadimplência. Os modelos são constantemente avaliados e, quando se detecta que não estão mais discriminando satisfatoriamente os clientes adimplentes dos inadimplentes, que é a variável resposta, são reformulados. Os clientes que tiveram o crédito negado, devido ao modelo estatístico ou a outras políticas da instituição, não podem ser classificados em alguma das duas características de interesse e, portanto, são caracterizados como clientes com dados omissos. Várias técnicas têm sido criadas para tentar inferir sobre a reputação destes clientes e costumam ser englobadas no que se chama de inferência sobre rejeitados (*reject inference*). Uma breve revisão dos métodos mais utilizados é apresentada em Thomas, Edelman e Crook (2004, pp.133-145), embora não se considere explicitamente mecanismos que geram os dados omissos, o que já é feito em Feelders (2000). Outros problemas semelhantes desta área, em que não se obtém amostras aleatórias da população-alvo, resultam de campanhas de *marketing*, que contatam determinados grupos por meio de mala-direta ou telefone — por vezes selecionados a partir de algum modelo estatístico com o objetivo de maximizar as respostas positivas —, e clientes que recebem um limite para o cartão de crédito, por exemplo, mas nem mesmo chegam a desbloquear o cartão para o seu uso. Cada padrão de omissão tem sua característica peculiar. Em relação à população, os clientes potenciais que aceitam mais facilmente às ofertas para adquirir produtos de crédito (empréstimo pessoal, limite de cheque especial ou cartão de crédito) costumam ser exatamente os mais necessitados e, por conseguinte, de maior risco de inadimplência. Em compensação, especula-se que os clientes que recebem o cartão de crédito, mas não o desbloqueiam, sejam majoritariamente clientes “bons”, por pagarem suas dívidas ou pelo fato de não quererem criá-las. Em contrapartida, os clientes que têm o crédito rejeitado devem ser os que provavelmente não honrarão as suas dívidas.

CAPÍTULO 2

Formulação geral

No Capítulo 1 considerou-se a modelagem de dados categorizados com omissão para o caso particular de duas respostas dicotômicas (Y_1, Y_2) , sem variáveis explicativas. Neste capítulo estende-se essa modelagem para casos que incluem um vetor aleatório (\mathbf{Y}) de variáveis respostas, apresentado em Paulino (1988, 1991) e Paulino e Singer (2006), e um vetor fixado $(\mathbf{X} = \mathbf{x})$ de variáveis explicativas, que não foi considerado previamente na literatura.

Outras modelagens e processos de estimação para dados categorizados com omissão foram apresentados por Paulino e Soares (2003), Baker (1994a) e Molenberghs e Goetghebeur (1997). Paulino e Soares (2003) estenderam os resultados apresentados em Paulino (1991) sob a distribuição multinomial para o produto de distribuições de Poisson. Baker (1994a) e Molenberghs e Goetghebeur (1997) descreveram algoritmos que permitem a obtenção de estimativas de máxima verossimilhança de parâmetros da distribuição multinomial sob mecanismos de omissão aleatória ou informativa.

Na Seção 2.1 descrevem-se o problema e a notação. Na Seção 2.2 apresentam-se o modelo probabilístico e os mecanismos de omissão. Na Seção 2.3 descrevem-se processos de estimação pelas metodologias de máxima verossimilhança (MV) e mínimos quadrados generalizados (MQG) sem impor restrições às probabilidades de categorização. Na Seção 2.4 apresentam-se processos de estimação que permitem a imposição de restrições por meio de modelos lineares ou log-lineares para a metodologia de MV e por meio de modelos funcionais lineares para a metodologia de MQG.

2.1 Descrição do problema e notação

Por simplicidade, admite-se que o vetor aleatório \mathbf{Y} pode assumir R possíveis valores \mathbf{y} correspondentes às combinações dos níveis de suas componentes $(Y_1, Y_2, \dots, Y_k)'$. Por exemplo, no caso trivariado $\mathbf{Y} = (Y_1, Y_2, Y_3)'$, com Y_1 , Y_2 e Y_3 podendo assumir, respectivamente, 2, 3 e 5 valores, temos $R = 2 \times 3 \times 5 = 30$. Da mesma forma, assume-se que o vetor \mathbf{X} pode assumir S possíveis valores \mathbf{x} , indicando que S corresponde às combinações dos níveis de suas componentes $(X_1, X_2, \dots, X_q)'$. As R categorias de respostas (S subpopulações) são indexadas por r (s).

Supõe-se que cada uma de n_{s++} unidades amostrais selecionadas aleatoriamente da s -ésima subpopulação pode ser (mas não necessariamente é) classificada na r -ésima categoria de resposta independentemente e com a mesma probabilidade $\theta_{r(s)}$, $r = 1, \dots, R$, $s = 1, \dots, S$. Isto indica que as $n_{+++} = \sum_{s=1}^S n_{s++}$ unidades seguem uma amostragem aleatória estratificada¹ com alocações distribuídas pelas subpopulações (estratos) conforme o vetor $\mathbf{N}_{++} = (n_{1++}, \dots, n_{S++})'$.

Por diversos motivos, nem sempre é possível se observar completamente as respostas de todas as variáveis integrantes de \mathbf{Y} . Nesses casos, apenas uma parte das n_{s++} unidades amostrais é classificada em uma das R categorias de respostas originalmente definidas e as demais estão associadas a algum tipo de omissão. Sendo assim, para cada subpopulação s , $s = 1, \dots, S$, definem-se T_s cenários de omissão, indexados pelo índice t , da seguinte forma. O cenário de ausência de omissão ou classificação completa é representado por $t = 1$ e aqueles com alguma omissão, por $t = 2, \dots, T_s$. Admite-se que as unidades correspondentes ao t -ésimo cenário de omissão, $t = 2, \dots, T_s$, são registradas em classes \mathcal{C}_{stc} , $c = 1, \dots, R_{st}$, constituídas por pelo menos duas das R categorias, com $\mathcal{C}_{stc} \cap \mathcal{C}_{std} = \emptyset$, $c \neq d$ e $\cup_{c=1}^{R_{st}} \mathcal{C}_{stc} = \{1, \dots, R\}$. Portanto, cada um dos $t = 2, \dots, T_s$ cenários de omissão formam partições $\mathcal{P}_{st} = \{\mathcal{C}_{stc}, c = 1, \dots, R_{st}\}$ do cenário de ausência de omissão $\mathcal{P}_{s1} = \mathcal{P}_1 = \{\{r\}, r = 1, \dots, R\}$ e R_{st} representa o número de classes de respostas existentes no t -ésimo cenário de omissão para a s -ésima subpopulação. Para consistência da notação, supõe-se que o cenário de ausência de omissão tem classes equivalentes às R categorias de respostas, isto é, $\mathcal{C}_{s1r} = \mathcal{C}_{1r} = \{r\}$, $r = 1, \dots, R$ e $R_{s1} = R_1 = R$. Representa-se o número de classes existentes nos $T_s - 1$ cenários com alguma omissão da s -ésima subpopulação por $l_s = \sum_{t=2}^{T_s} R_{st}$.

¹A inexistência de variáveis explicativas é um caso particular do exposto, quando $S = 1$, indicando uma amostragem aleatória simples.

Por conveniência computacional, constroem-se vetores \mathbf{z}_{stc} , de dimensão $R \times 1$, com elementos iguais a 1 associados às categorias pertencentes à classe \mathcal{C}_{stc} , e com demais elementos nulos; a matriz $\mathbf{Z}_{st} = [\mathbf{z}_{stc}, c = 1, \dots, R_{st}]$, de dimensão $R \times R_{st}$, contém os vetores indicadores de todas as classes do t -ésimo cenário de omissão para a s -ésima subpopulação e a matriz $\mathbf{Z}_s = [\mathbf{Z}_{st}, t = 1, \dots, T_s]$, de dimensão $R \times (R + l_s)$, inclui os vetores indicadores de todas as classes de todos os cenários de omissão da s -ésima subpopulação. Note que, $\mathbf{Z}_{s1} = \mathbf{I}_R$ (matriz identidade de ordem R), $s = 1, \dots, S$. As freqüências observáveis, $\{n_{stc}\}$, indicam as unidades da s -ésima subpopulação com o t -ésimo padrão de omissão classificadas na c -ésima classe, $s = 1, \dots, S$, $t = 1, \dots, T_s$, $c = 1, \dots, R_{st}$. O vetor $\mathbf{N}_{st} = (n_{stc}, c = 1, \dots, R_{st})'$ empilha as freqüências observáveis do t -ésimo cenário da s -ésima subpopulação, $\mathbf{N}_s = (\mathbf{N}'_{st}, t = 1, \dots, T_s)'$ contém todas as freqüências observáveis da s -ésima subpopulação, $\mathbf{N} = (\mathbf{N}'_s, s = 1, \dots, S)'$ inclui todas as freqüências observáveis e $n_{st+} = \sum_{c=1}^{R_{st}} n_{stc}$ indica o total de unidades selecionadas da s -ésima subpopulação com o t -ésimo cenário de omissão.

Assume-se que uma unidade amostral selecionada da s -ésima subpopulação com a r -ésima categoria de resposta é classificada no t -ésimo cenário de omissão com probabilidade $\lambda_{t(rs)}$, $r = 1, \dots, R$, $s = 1, \dots, S$, $t = 1, \dots, T_s$. O conjunto $\{\lambda_{t(rs)}\}$ contém as probabilidades condicionais de omissão. Como os valores das variáveis explicativas são conhecidos ou fixados de antemão, definindo as S subpopulações, supõe-se a inexistência de omissão em \mathbf{X} .

Com o intuito de esclarecer esta notação, consideram-se alguns exemplos.

Exemplo 1 (Continuação, p.1)

Devido à inexistência de variáveis explicativas, elimina-se o índice s . Associa-se a categoria de resposta 1 às mães fumantes com recém-nascidos pesando < 2.5 kg, a categoria 2 às mães fumantes com recém-nascidos pesando ≥ 2.5 kg, a categoria 3 às mães não-fumantes com recém-nascidos pesando < 2.5 kg e a categoria 4 às mães não-fumantes com recém-nascidos pesando ≥ 2.5 kg. Os cenários de omissão indexados por $t = 1, 2, 3, 4$ são aqueles ilustrados na Tabela 1.2.

No cenário de ausência de omissão ($t = 1$), em que as unidades são completamente categorizadas em uma das $R_1 = R = 4$ classes/categorias $\mathcal{C}_{1r} = \{r\}$, $r = 1, \dots, 4$, tem-se $\mathcal{P}_1 = \{\{1\}, \{2\}, \{3\}, \{4\}\}$, $\mathbf{z}_{11} = (1, 0, 0, 0)'$, $\mathbf{z}_{12} = (0, 1, 0, 0)'$, $\mathbf{z}_{13} = (0, 0, 1, 0)'$, $\mathbf{z}_{14} = (0, 0, 0, 1)'$, $\mathbf{N}_1 = (4\ 512, 21\ 009, 3\ 394, 24\ 132)'$ e $n_{1+} = 53\ 047$.

No cenário de omissão indexado por $t = 2$, em que não é observado o peso dos recém-

-nascidos, $n_{21} = 1\,049$ mães são classificadas na primeira das suas $R_2 = 2$ classes, que corresponde a $\mathcal{C}_{21} = \{1, 2\}$ e indica que essas unidades estão associadas às categorias 1 e 2 por meio de $\mathbf{z}_{21} = (1, 1, 0, 0)'$; analogamente, as outras $n_{22} = 1\,135$ mães são classificadas na segunda classe deste cenário, que corresponde a $\mathcal{C}_{22} = \{3, 4\}$ e indica que essas unidades estão associadas às categorias 3 e 4 por meio de $\mathbf{z}_{22} = (0, 0, 1, 1)'$. Assim, $\mathcal{P}_2 = \{\{1, 2\}, \{3, 4\}\}$, $\mathbf{N}_2 = (1\,049, 1\,135)'$ e $n_{2+} = 2\,184$.

Da mesma forma, no cenário $t = 3$ obtém-se $R_3 = 2$, $\mathcal{C}_{31} = \{1, 3\}$, $\mathcal{C}_{32} = \{2, 4\}$, $\mathcal{P}_3 = \{\{1, 3\}, \{2, 4\}\}$, $\mathbf{z}_{31} = (1, 0, 1, 0)'$, $\mathbf{z}_{32} = (0, 1, 0, 1)'$, $\mathbf{N}_3 = (142, 464)'$, $n_{3+} = 606$ e no cenário $t = 4$, $R_4 = 1$, $\mathcal{P}_4 = \{\mathcal{C}_{41}\} = \{\{1, 2, 3, 4\}\}$, $\mathbf{z}_{41} = (1, 1, 1, 1)'$, $\mathbf{N}_4 = n_{4+} = 1\,224$.

Por fim, tem-se $l = R_2 + R_3 + R_4 = 5$, $\mathbf{N} = (\mathbf{N}'_1, \mathbf{N}'_2, \mathbf{N}'_3, \mathbf{N}'_4)'$, $n_{++} = 57\,061$ e $\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3, \mathbf{Z}_4]$, em que $\mathbf{Z}_1 = [\mathbf{z}_{11}, \mathbf{z}_{12}, \mathbf{z}_{13}, \mathbf{z}_{14}] = \mathbf{I}_4$,

$$\mathbf{Z}_2 = [\mathbf{z}_{21}, \mathbf{z}_{22}] = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} = \mathbf{I}_2 \otimes \mathbf{1}_2, \quad \mathbf{Z}_3 = [\mathbf{z}_{31}, \mathbf{z}_{32}] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{1}_2 \otimes \mathbf{I}_2,$$

$\mathbf{Z}_4 = \mathbf{z}_{41} = (1, 1, 1, 1)' = \mathbf{1}_4$, \otimes denota o produto de Kronecker (à direita) e o vetor $\mathbf{1}_k$, de dimensão $k \times 1$, possui todos os elementos iguais a 1. ■

Exemplo 2 (Lipsitz e Fitzmaurice, 1996)

Em um estudo realizado nos Estados Unidos, 661 e 477 pares de mães/filhos das cidades de Kingston-Harriman (Tennessee) e Portage (Wisconsin) foram, respectivamente, classificados de acordo com o hábito de fumo da mãe (nenhum, moderado e forte) e o estado respiratório do filho (normal, ofegante com resfriado e ofegante sem resfriado) aos 10 anos de idade. O objetivo do estudo era avaliar a influência da fumaça do cigarro no estado respiratório das crianças das duas cidades. Os dados estão expostos na Tabela 2.1.

Por motivos desconhecidos, somente 44% (50%) dos pares de mães/filhos da cidade de Kingston-Harriman (Portage) foram completamente categorizados. A informação sobre o estado respiratório das crianças e o hábito de fumo das mães não foi observada para, respectivamente, 49% (39%) e 8%² (11%).

²A soma das 3 porcentagens de Kingston-Harriman ultrapassa os 100% devido aos arredondamentos.

Tabela 2.1: Freqüências observadas

Cidade de residência	Hábito de fumo da mãe	Estado respiratório da criança			omisso
		normal	ofegante com resfriado	ofegante sem resfriado	
Kingston-Harriman	nenhum	167	17	19	176
	moderado	10	1	3	24
	forte	52	10	11	121
	omisso	28	10	12	
Portage	nenhum	120	22	19	103
	moderado	8	5	1	3
	forte	39	12	12	80
	omisso	31	8	14	

Associa-se o índice $s = 1$ ($s = 2$) à subpopulação de Kingston-Harriman (Portage). Para cada uma das $S = 2$ subpopulações, as categorias de respostas, indexadas por r , terão valores correspondentes aos apresentados na Tabela 2.2. Representa-se o cenário de omissão em que o estado respiratório das crianças (hábito de fumo das mães) não é observado por $t = 2$ ($t = 3$), igualmente para as duas cidades.

Tabela 2.2: Categorias de respostas e correspondentes valores associados ao índice r

Hábito de fumo da mãe	Estado respiratório da criança		
	normal	ofegante com resfriado	ofegante sem resfriado
nenhum	1	2	3
moderado	4	5	6
forte	7	8	9

Nos cenários de ausência de omissão ($t = 1$), em que as unidades são completamente categorizadas em uma das $R_{s1} = R = 9$ classes/categorias, têm-se $\mathcal{P}_{s1} = \{\{1\}, \{2\}, \dots, \{9\}\}$, $\mathbf{Z}_{s1} = \mathbf{I}_9$, $s = 1, 2$, $\mathbf{N}_{11} = (167, 17, 19, 10, 1, 3, 52, 10, 11)'$, $n_{11+} = 290$, $\mathbf{N}_{21} = (120, 22, 19, 8, 5, 1, 39, 12, 12)'$ e $n_{21+} = 238$.

Para as duas cidades, nos cenários $t = 2$ têm-se $R_{s2} = 3$, $\mathcal{P}_{s2} = \{\{1, 2, 3\}, \{4, 5, 6\}, \{7, 8, 9\}\}$, $\mathbf{Z}_{s2} = \mathbf{I}_3 \otimes \mathbf{1}_3$, $s = 1, 2$, $\mathbf{N}_{12} = (176, 24, 121)'$, $n_{12+} = 321$, $\mathbf{N}_{22} = (103, 3, 80)'$, $n_{22+} = 186$

e nos cenários $t = 3$, $R_{s3} = 3$, $\mathcal{P}_{s3} = \{\{1, 4, 7\}, \{2, 5, 8\}, \{3, 6, 9\}\}$, $\mathbf{Z}_{s3} = \mathbf{1}_3 \otimes \mathbf{I}_3$, $s = 1, 2$, $\mathbf{N}_{13} = (28, 10, 12)'$, $n_{13+} = 50$, $\mathbf{N}_{23} = (31, 8, 14)'$, $n_{23+} = 53$.

Adicionalmente, obtém-se $l_s = R_{s2} + R_{s3} = 6$, $\mathbf{N}_s = (\mathbf{N}'_{s1}, \mathbf{N}'_{s2}, \mathbf{N}'_{s3})'$, $\mathbf{Z}_s = [\mathbf{Z}_{s1}, \mathbf{Z}_{s2}, \mathbf{Z}_{s3}]$, $s = 1, 2$, $\mathbf{N}_{++} = (n_{1++}, n_{2++})' = (661, 477)'$, $n_{+++} = 1\ 138$ e $\mathbf{N} = (\mathbf{N}'_1, \mathbf{N}'_2)'$.

Note que, neste exemplo, os cenários de omissão observados para as duas cidades foram os mesmos. Caso isto não tivesse ocorrido, as quantidades R_{st} , \mathcal{C}_{stc} , \mathcal{P}_{st} , l_s , \mathbf{z}_{stc} , \mathbf{Z}_{st} e \mathbf{Z}_s não seriam necessariamente iguais para $s = 1, 2$. ■

Exemplo 3 (André, Neves e Tseng, 1990)

Uma amostra de 97 crianças em idade escolar com dentição permanente foi avaliada por meio de duas técnicas com relação à predisposição de se contrair cáries dentárias. Um método (convencional) de difícil aplicação em grande escala e de custos elevados, é o da contagem de bactérias *Lactobacillus* em amostras salivares. De acordo com o maior ou menor número destas bactérias, o grau de suscetibilidade à cárie foi classificado em alto, médio ou baixo. Um método (simplificado) de fácil aplicação e baixos custos utilizou esta mesma classificação de acordo com as colorações azul, violeta e rosa, obtidas com a reação da amostra de saliva com resazurina. O estudo teve o objetivo de comparar as distribuições marginais dos graus de suscetibilidade à cárie dentária obtidas pelos dois métodos e obter uma medida de concordância entre as duas classificações. As frequências observadas, apresentadas na Tabela 2.3, ilustram que 46 crianças não puderam ser classificadas de acordo com a escala de interesse devido à ocorrência de cores intermediárias no teste baseado na resazurina. Este conjunto de dados também foi analisado por Paulino e Pereira (1995), Rodrigues (1996), Soares e Paulino (2001), Soares (2004) e Paulino e Singer (2006).

Novamente suprime-se o índice s , devido à inexistência de variáveis explicativas. As categorias de respostas, representadas pelo índice r , têm os valores da Tabela 2.2 com correspondência adequada à Tabela 2.3. Representa-se o cenário de omissão em que não existe distinção entre as categorias alto e médio (médio e baixo) para o método simplificado por $t = 2$ ($t = 3$).

No cenário de ausência de omissão ($t = 1$), em que as unidades são completamente categorizadas em uma das $R_1 = R = 9$ classes/categorias, $\mathcal{P}_1 = \{\{r\}, r = 1, \dots, 9\}$, $\mathbf{Z}_1 = \mathbf{I}_9$, $\mathbf{N}_1 = (7, 11, 2, 3, 9, 5, 0, 10, 4)'$ e $n_{1+} = 51$.

O cenário $t = 2$ pode ser encaixado no contexto de partições associando as classes $\mathcal{C}_{21} = \{1, 4\}$, $\mathcal{C}_{22} = \{2, 5\}$, $\mathcal{C}_{23} = \{3, 6\}$ e $\mathcal{C}_{24} = \{7, 8, 9\}$ às frequências $n_{21} = 8$, $n_{22} = 7$, $n_{23} = 3$

Tabela 2.3: Freqüências observadas dos graus de suscetibilidade à cárie dentária

Método simplificado	Método convencional		
	alto	médio	baixo
alto	7	11	2
médio	3	9	5
baixo	0	10	4
alto / médio	8	7	3
médio / baixo	7	14	7

e $n_{24} = 0$. Note que a definição da última classe é um artifício para, juntamente com as outras, formar uma partição do conjunto de categorias de respostas. Portanto, tem-se $R_2 = 4$, $\mathcal{P}_2 = \{\{1, 4\}, \{2, 5\}, \{3, 6\}, \{7, 8, 9\}\}$,

$$\mathbf{Z}_2 = [\mathbf{z}_{21}, \mathbf{z}_{22}, \mathbf{z}_{23}, \mathbf{z}_{24}] = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{array} \right] = \begin{bmatrix} \mathbf{1}_2 \otimes \mathbf{I}_3 & \mathbf{0}_6 \\ \mathbf{0}_{3,3} & \mathbf{1}_3 \end{bmatrix},$$

$\mathbf{N}_2 = (8, 7, 3, 0)'$ e $n_{2+} = 18$, em que $\mathbf{0}_k$ denota o vetor, de dimensão $k \times 1$, com todos os elementos iguais a 0 e $\mathbf{0}_{j,k}$, a matriz, de dimensão $j \times k$, com todos os elementos nulos. Da mesma forma, no cenário $t = 3$, obtém-se $R_3 = 4$, $\mathcal{P}_3 = \{\{1, 2, 3\}, \{4, 7\}, \{5, 8\}, \{6, 9\}\}$,

$$\mathbf{Z}_3 = [\mathbf{z}_{31}, \mathbf{z}_{32}, \mathbf{z}_{33}, \mathbf{z}_{34}] = \left[\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right] = \begin{bmatrix} \mathbf{1}_3 & \mathbf{0}_{3,3} \\ \mathbf{0}_6 & \mathbf{1}_2 \otimes \mathbf{I}_3 \end{bmatrix},$$

$\mathbf{N}_3 = (0, 7, 14, 7)'$ e $n_{3+} = 28$.

Obtém-se também $l = R_2 + R_3 = 8$, $\mathbf{N} = (\mathbf{N}'_1, \mathbf{N}'_2, \mathbf{N}'_3)'$, $n_{++} = 97$ e $\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3]$. ■

Por meio destes exemplos, note-se que as condições $\mathcal{C}_{stc} \cap \mathcal{C}_{std} = \emptyset$, $c \neq d$ e $\cup_{c=1}^{R_{st}} \mathcal{C}_{stc} = \{1, \dots, R\}$ podem ser verificadas observando que \mathbf{Z}_{st} tem um elemento igual a 1 em exatamente uma única coluna para cada linha, $s = 1, \dots, S$, $t = 1, \dots, T_s$.

2.2 Modelo probabilístico e mecanismos de omissão

Assume-se que as frequências observáveis \mathbf{N} seguem uma distribuição produto de multinomiais especificada pela função de probabilidade

$$P(\mathbf{N} | \boldsymbol{\theta}, \{\lambda_{t(rs)}\}, \mathbf{N}_{++}) = \prod_{s=1}^S \frac{n_{s++}!}{\prod_{t=1}^{T_s} \prod_{c=1}^{R_{st}} n_{stc}!} \prod_{r=1}^R (\theta_{r(s)} \lambda_{1(rs)})^{n_{s1r}} \prod_{t=2}^{T_s} \prod_{c=1}^{R_{st}} \left(\sum_{r \in \mathcal{C}_{stc}} \theta_{r(s)} \lambda_{t(rs)} \right)^{n_{stc}}, \quad (2.1)$$

em que $\theta_{r(s)}$ é a probabilidade de uma unidade amostral da s -ésima subpopulação ser classificada na r -ésima categoria de resposta; $\lambda_{t(rs)}$ é a probabilidade de uma unidade amostral com a r -ésima categoria de resposta selecionada da s -ésima subpopulação ser classificada no t -ésimo cenário de omissão, $\boldsymbol{\theta} = (\boldsymbol{\theta}'_s, s = 1, \dots, S)'$, $\boldsymbol{\theta}_s = (\theta_{r(s)}, r = 1, \dots, R)'$, $\sum_{r=1}^R \theta_{r(s)} = 1$, $s = 1, \dots, S$ e $\sum_{t=1}^{T_s} \lambda_{t(rs)} = 1$, $r = 1, \dots, R$, $s = 1, \dots, S$.

Caso fosse possível identificar a hipotética categorização de resposta nos vários padrões de omissão, y_{str} indicaria o número hipotético de unidades amostrais da s -ésima subpopulação com o t -ésimo cenário de omissão classificadas na r -ésima categoria de resposta, $s = 1, \dots, S$, $t = 1, \dots, T_s$, $r = 1, \dots, R$. Deste modo, $\{y_{str}\}$ denotam as frequências ampliadas, observadas apenas para o cenário de ausência de omissão, pelo que $n_{s1r} = y_{s1r}$, $s = 1, \dots, S$, $r = 1, \dots, R$. Nos cenários com alguma omissão, estas frequências não são observáveis e conhece-se somente as frequências associadas às classes de respostas \mathcal{C}_{stc} , indicadas por meio de

$$n_{stc} = \sum_{r \in \mathcal{C}_{stc}} y_{str}, \quad s = 1, \dots, S, \quad t = 2, \dots, T_s, \quad c = 1, \dots, R_{st}. \quad (2.2)$$

Para cada subpopulação, há $R - 1$ parâmetros $\{\theta_{r(s)}\}$ e $R(T_s - 1)$ parâmetros $\{\lambda_{t(rs)}\}$, linearmente independentes, totalizando $RT_s - 1$. Da mesma forma, há R frequências observadas

no cenário de ausência de omissão e l_s delas nos cenários com alguma omissão. Considerando que n_{s++} foi fixado previamente, há um total de $R - 1 + l_s$ frequências observadas linearmente independentes em cada subpopulação. Assim, os $R \sum_{s=1}^S T_s - S$ parâmetros $\{\theta_{r(s)}, \lambda_{t(rs)}\}$ linearmente independentes, associados às frequências ampliadas $\{y_{str}\}$, quando confrontados com as $S(R - 1) + \sum_{s=1}^S l_s$ frequências observáveis $\{n_{stc}\}$ linearmente independentes, que estão associadas aos parâmetros $\{\sum_{r \in \mathcal{C}_{stc}} \theta_{r(s)} \lambda_{t(rs)}\}$, evidenciam uma sobre-parametrização de (2.1) com $\sum_{s=1}^S [R(T_s - 1) - l_s]$ parâmetros não identificáveis.

Uma vez que o interesse se concentra geralmente nas probabilidades de categorização $\{\theta_{r(s)}\}$, estruturas mais reduzidas para as probabilidades condicionais de omissão $\{\lambda_{t(rs)}\}$ são consideradas para tornar o modelo identificável. O mecanismo de omissão aleatória ou não-informativa é explicitado por

$$\mathbf{MAR} : \lambda_{t(rs)} = \alpha_{t(cs)}, \quad s = 1, \dots, S, \quad t = 1, \dots, T_s, \quad c = 1, \dots, R_{st}, \quad r \in \mathcal{C}_{stc}, \quad (2.3)$$

indicando que as probabilidades condicionais de omissão dependem apenas das classes observadas. O modelo estatístico sob o mecanismo MAR é saturado, com função de verossimilhança dada por

$$\begin{aligned} L(\boldsymbol{\theta}, \{\alpha_{t(cs)}\} | \mathbf{N}; \mathbf{MAR}) &\propto \prod_{s=1}^S \prod_{r=1}^R (\theta_{r(s)} \alpha_{1(rs)})^{n_{s1r}} \prod_{t=2}^{T_s} \prod_{c=1}^{R_{st}} \left(\sum_{r \in \mathcal{C}_{stc}} \theta_{r(s)} \alpha_{t(cs)} \right)^{n_{stc}} \\ &= \prod_{s=1}^S \prod_{r=1}^R \theta_{r(s)}^{n_{s1r}} \prod_{t=2}^{T_s} \prod_{c=1}^{R_{st}} \left(\sum_{r \in \mathcal{C}_{stc}} \theta_{r(s)} \right)^{n_{stc}} \prod_{s=1}^S \prod_{t=1}^{T_s} \prod_{c=1}^{R_{st}} \alpha_{t(cs)}^{n_{stc}} \\ &= \prod_{s=1}^S \prod_{r=1}^R \theta_{r(s)}^{n_{s1r}} \prod_{t=2}^{T_s} \prod_{c=1}^{R_{st}} (\mathbf{z}'_{stc} \boldsymbol{\theta}_s)^{n_{stc}} \prod_{s=1}^S \prod_{t=1}^{T_s} \prod_{c=1}^{R_{st}} \alpha_{t(cs)}^{n_{stc}} \\ &\equiv L_1(\boldsymbol{\theta} | \mathbf{N}) L_2(\{\alpha_{t(cs)}\} | \mathbf{N}; \mathbf{MAR}), \end{aligned} \quad (2.4)$$

em que

$$L_1(\boldsymbol{\theta} | \mathbf{N}) = \prod_{s=1}^S \prod_{r=1}^R \theta_{r(s)}^{n_{s1r}} \prod_{t=2}^{T_s} \prod_{c=1}^{R_{st}} (\mathbf{z}'_{stc} \boldsymbol{\theta}_s)^{n_{stc}}$$

e

$$L_2(\{\alpha_{t(cs)}\} | \mathbf{N}; \mathbf{MAR}) = \prod_{s=1}^S \prod_{t=1}^{T_s} \prod_{c=1}^{R_{st}} \alpha_{t(cs)}^{n_{stc}}.$$

O mecanismo de omissão completamente aleatória é um caso particular do MAR definido por

$$\mathbf{MCAR} : \lambda_{t(rs)} = \alpha_{t(s)}, \quad s = 1, \dots, S, \quad t = 1, \dots, T_s, \quad r = 1, \dots, R, \quad (2.5)$$

indicando que as probabilidades condicionais de omissão independem das categorias de respostas, sendo ou não parcialmente observadas. O modelo estatístico sob o mecanismo MCAR tem $S(R-2) + \sum_{s=1}^S T_s$ parâmetros linearmente independentes, pois há $T_s - 1$ parâmetros $\{\alpha_{t(s)}\}$ em cada subpopulação, que juntamente com os $R - 1$ parâmetros $\{\theta_{r(s)}\}$, totalizam $R - 2 + T_s$ parâmetros em cada subpopulação. Subtraindo os $S(R-2) + \sum_{s=1}^S T_s$ parâmetros das $S(R-1) + \sum_{s=1}^S l_s$ frequências observáveis, há, sob este mecanismo de omissão, $S + \sum_{s=1}^S (l_s - T_s)$ graus de liberdade na função de verossimilhança, que é explicitada por

$$L(\boldsymbol{\theta}, \{\alpha_{t(s)}\} | \mathbf{N}; \mathbf{MCAR}) \propto L_1(\boldsymbol{\theta} | \mathbf{N}) L_2(\{\alpha_{t(s)}\} | \{n_{st+}\}; \mathbf{MCAR}), \quad (2.6)$$

em que $L_1(\boldsymbol{\theta} | \mathbf{N})$ tem a mesma definição de (2.4) e

$$L_2(\{\alpha_{t(s)}\} | \{n_{st+}\}; \mathbf{MCAR}) = \prod_{s=1}^S \prod_{t=1}^{T_s} \alpha_{t(s)}^{n_{st+}}.$$

Note que as probabilidades condicionais de omissão definidas nos mecanismos MAR e MCAR dependem das variáveis explicativas. Little (1995) acredita que o termo “MCAR” deve ser reservado para o caso em que a omissão não depende das variáveis respostas e explicativas, *i.e.*, quando $\lambda_{t(rs)} = \alpha_{t(s)}$. Ele também sugere que se use a expressão “omissão dependente de covariáveis” — a tradução de *covariate-dependent dropout* foi adaptada, uma vez que o artigo trata apenas de modelagens de abandonos de estudos — quando o mecanismo de omissão não depender das respostas observadas ou omissas, mas depender das variáveis explicativas. Utiliza-se, nesta dissertação, a definição $\lambda_{t(rs)} = \alpha_{t(s)}$ para o mecanismo MCAR por ser a generalização mais imediata quando se passa da distribuição multinomial para a distribuição produto de multinomiais e porque é mais abrangente do que o MCAR de Little, que pode ser visto como caso especial (MCAR reduzido). Estruturas mais parcimoniosas podem ser estudadas sob os dois mecanismos, permitindo que as probabilidades condicionais de omissão não variem para algumas ou para todas as subpopulações. Desconsideram-se estas restrições adicionais, pois elas não alteram as estimativas de $\boldsymbol{\theta}$, de interesse principal, devido à fatoração das verossimilhanças sob os mecanismos MAR e MCAR numa parte referente a $\boldsymbol{\theta}$, denotada por $L_1(\boldsymbol{\theta} | \mathbf{N})$, e noutra concernente às probabilidades condicionais de omissão, representada por $L_2(\{\alpha_{t(cs)}\} | \mathbf{N}; \mathbf{MAR})$ no caso MAR ou $L_2(\{\alpha_{t(s)}\} | \{n_{st+}\}; \mathbf{MCAR})$ no caso MCAR.

Mecanismos de omissão não-aleatória, informativa ou MNAR podem ser explicitados fazendo com que pelo menos duas das probabilidades condicionais de omissão de respostas pertencentes à mesma classe não sejam iguais, ou seja, $\{a, b\} \in \mathcal{C}_{stc}$ e $\lambda_{t(as)} \neq \lambda_{t(bs)}$. Ainda assim, é necessário especificar pelo menos $\sum_{s=1}^S [R(T_s - 1) - l_s]$ restrições paramétricas para obter uma estrutura identificável.

2.3 Estimação das probabilidades de categorização para modelos estruturais saturados

Na Seção 2.3.1 apresenta-se o processo de estimação de $\boldsymbol{\theta}$ pela metodologia de MV sob os mecanismos MAR e MCAR e na Seção 2.3.2, pela metodologia de MQG sob o mecanismo MCAR. Na estimação por MV (MQG) utilizam-se as estatísticas de razão de verossimilhanças, Pearson e Neyman (apenas Neyman) para testar o ajuste do mecanismo MCAR condicionalmente à suposição do mecanismo MAR.

Conforme indicado na Seção 1.2.1, num cenário de omissão total, representado por $\mathcal{P}_{st} = \{\mathcal{C}_{st1}\} = \{\{1, \dots, R\}\}$, as unidades amostrais não trazem qualquer informação para a estimação de $\boldsymbol{\theta}$ quer sob o mecanismo MAR, quer sob o mecanismo mais restritivo MCAR. Como o interesse concentra-se exatamente nesse parâmetro, ignoram-se estes cenários de omissão, redefinindo T_s como o número de cenários de omissão parcial e n_{s++} como o número de unidades que sofreram algum tipo de categorização. Com isto, os termos definidos nos Exemplos 2 e 3 não sofrem alterações, enquanto que os do Exemplo 1 necessitam algumas modificações ao se desconsiderar o cenário $t = 4$. Portanto, redefine-se $l = R_2 + R_3 = 4$, $\mathbf{N} = (\mathbf{N}'_1, \mathbf{N}'_2, \mathbf{N}'_3)'$, $n_{++} = 55\,837$ e $\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3]$.

Para simplificar a obtenção de alguns resultados por meio de operações matriciais, utiliza-se a seguinte notação adicional:

- $\bar{\boldsymbol{\theta}}_s = [\mathbf{I}_{R-1}, \mathbf{0}_{R-1}] \boldsymbol{\theta}_s = (\theta_{r(s)}, r = 1, \dots, R - 1)'$ contém as $R - 1$ primeiras componentes de $\boldsymbol{\theta}_s$, $s = 1, \dots, S$;
- $\bar{\boldsymbol{\theta}} = (\mathbf{I}_S \otimes [\mathbf{I}_{R-1}, \mathbf{0}_{R-1}]) \boldsymbol{\theta} = (\bar{\boldsymbol{\theta}}'_s, s = 1, \dots, S)'$;
- a matriz $\bar{\mathbf{Z}}_{st}$, de dimensão $(R - 1) \times (R_{st} - 1)$, é obtida de \mathbf{Z}_{st} por remoção da última linha e coluna, $s = 1, \dots, S$, $t = 1, \dots, T_s$;

- $\bar{\mathbf{Z}}_s = (\bar{\mathbf{Z}}_{st}, t = 1, \dots, T_s), s = 1, \dots, S;$
- $\bar{\boldsymbol{\theta}}_{st} = \bar{\mathbf{Z}}'_{st} \bar{\boldsymbol{\theta}}_s = (\theta_{c(st)}, c = 1, \dots, R_{st} - 1)'$ contém os parâmetros $\{\theta_{r(s)}\}$ associados às $R_{st} - 1$ primeiras classes do t -ésimo cenário de omissão da s -ésima subpopulação, em que $\theta_{c(st)} = \sum_{r \in \mathcal{C}_{stc}} \theta_{r(s)}, s = 1, \dots, S, t = 1, \dots, T_s;$
- $\mathbf{p}_{st} = \mathbf{N}_{st}/n_{st+} = (p_{c(st)}, c = 1, \dots, R_{st})'$ são as proporções observadas no t -ésimo cenário de omissão da s -ésima subpopulação, $s = 1, \dots, S, t = 1, \dots, T_s;$
- $\mathbf{p}_s = (\mathbf{p}'_{st}, t = 1, \dots, T_s)', s = 1, \dots, S;$
- $\bar{\mathbf{N}}_{st} = [\mathbf{I}_{R_{st}-1}, \mathbf{0}_{R_{st}-1}] \mathbf{N}_{st} = (n_{stc}, c = 1, \dots, R_{st} - 1)', s = 1, \dots, S, t = 1, \dots, T_s;$
- $\bar{\mathbf{p}}_{st} = \bar{\mathbf{N}}_{st}/n_{st+}, s = 1, \dots, S, t = 1, \dots, T_s;$
- $\bar{\mathbf{p}}_s = (\bar{\mathbf{p}}'_{st}, t = 1, \dots, T_s)', s = 1, \dots, S.$

Sempre que for necessário, obtém-se $\boldsymbol{\theta}_s$ e $\boldsymbol{\theta}$ a partir de $\bar{\boldsymbol{\theta}}_s$ e $\bar{\boldsymbol{\theta}}$ por meio das relações

$$\boldsymbol{\theta}_s = \begin{pmatrix} \mathbf{0}_{R-1} \\ 1 \end{pmatrix} + \begin{pmatrix} \mathbf{I}_{R-1} \\ -\mathbf{1}'_{R-1} \end{pmatrix} \bar{\boldsymbol{\theta}}_s \equiv \mathbf{b}_s + \mathbf{B}_s \bar{\boldsymbol{\theta}}_s, \quad (2.7)$$

$$\boldsymbol{\theta} = \mathbf{1}_S \otimes \begin{pmatrix} \mathbf{0}_{R-1} \\ 1 \end{pmatrix} + \left[\mathbf{I}_S \otimes \begin{pmatrix} \mathbf{I}_{R-1} \\ -\mathbf{1}'_{R-1} \end{pmatrix} \right] \bar{\boldsymbol{\theta}} \equiv \mathbf{b} + \mathbf{B} \bar{\boldsymbol{\theta}}, \quad (2.8)$$

em que $\mathbf{b}_s = (\mathbf{0}'_{R-1}, 1)'$, $\mathbf{B}_s = (\mathbf{I}_{R-1}, -\mathbf{1}_{R-1})'$, $\mathbf{b} = \mathbf{1}_S \otimes (\mathbf{0}'_{R-1}, 1)'$ e $\mathbf{B} = \mathbf{I}_S \otimes (\mathbf{I}_{R-1}, -\mathbf{1}_{R-1})'$.

2.3.1 Estimação por máxima verossimilhança para modelos estruturais saturados sob os mecanismos MAR e MCAR

A estimação de $\boldsymbol{\theta}$ por máxima verossimilhança pode ser feita maximizando apenas o fator $L_1(\boldsymbol{\theta} | \mathbf{N})$ de (2.4), ou equivalentemente,

$$\ln L_1(\boldsymbol{\theta} | \mathbf{N}) = \sum_{s=1}^S \left[\sum_{r=1}^R n_{s1r} \ln \theta_{r(s)} + \sum_{t=2}^{T_s} \sum_{c=1}^{R_{st}} n_{stc} \ln (\mathbf{z}'_{stc} \boldsymbol{\theta}_s) \right]. \quad (2.9)$$

Com exceção do padrão de omissão monótono, exposto na Seção 1.2.1, a equação $\partial \ln L_1(\boldsymbol{\theta} | \mathbf{N}) / \partial \boldsymbol{\theta} = \mathbf{0}$, em geral não tem solução explícita, o que implica que os EMV de $\boldsymbol{\theta}$ devem ser obtidos por meio de métodos iterativos. O algoritmo EM pode ser explicitado com

$$\hat{\boldsymbol{\theta}}_s^{(i+1)} = \frac{1}{n_{s++}} \left(\mathbf{N}_{s1} + \sum_{t=2}^{T_s} \mathbf{D}_{\hat{\boldsymbol{\theta}}_s^{(i)}} \mathbf{Z}_{st} \mathbf{D}_{\mathbf{z}'_{st} \hat{\boldsymbol{\theta}}_s^{(i)}}^{-1} \mathbf{N}_{st} \right), \quad s = 1, \dots, S, \quad i = 1, \dots, \quad (2.10)$$

em que $\mathbf{D}_{\hat{\boldsymbol{\theta}}_s^{(i)}}$ representa uma matriz diagonal com elementos $\hat{\boldsymbol{\theta}}_s^{(i)}$ na diagonal principal e $\hat{\boldsymbol{\theta}}_s^{(i)}$ é a estimativa do ponto de máximo local obtida na i -ésima iteração. Dempster, Laird e Rubin (1977) mostraram que a log-verossimilhança da $(i + 1)$ -ésima iteração é sempre maior ou igual àquela obtida na i -ésima iteração e, assim, à medida que $i \rightarrow \infty$, $\hat{\boldsymbol{\theta}}_s^{(i)}$ convergirá para um ponto de máximo local. Se este ponto também for o máximo global, como em geral ocorre para distribuições da família exponencial, então $\hat{\boldsymbol{\theta}}_s = \hat{\boldsymbol{\theta}}_s^{(\infty)}$ será a estimativa de máxima verossimilhança de $\boldsymbol{\theta}_s$. Apresentam-se alguns detalhes sobre a obtenção de (2.10) no Apêndice A.1.

O processo iterativo pode ser inicializado com, por exemplo, as proporções observadas de unidades completamente categorizadas, ou seja, $\hat{\boldsymbol{\theta}}_s^{(0)} = \mathbf{p}_{s1} = \mathbf{N}_{s1}/n_{s1+}$. Neste caso é importante substituir qualquer frequência eventualmente nula do cenário de ausência de omissão por um valor pequeno, *e.g.*, $(R n_{s1+})^{-1}$ ou 10^{-6} , pois um valor nulo de $\hat{\boldsymbol{\theta}}_{r(s)}^{(0)}$ faz com que suas estimativas correspondentes em qualquer iteração continuem nulas, não permitindo que se incorporem informações referentes a $\theta_{r(s)}$ obtidas dos cenários com alguma omissão.

A lentidão do algoritmo EM pode ser contornada pelo uso dos algoritmos de Newton-Raphson ou *scoring* de Fisher. Estes algoritmos necessitam a determinação do vetor gradiente e da matriz hessiana, ou da matriz de informação de Fisher, de (2.9). O vetor *score* de $\ln L_1(\boldsymbol{\theta} | \mathbf{N}_s)$, com dimensão $S(R - 1) \times 1$, pode ser representado por

$$\mathbf{S}_1(\bar{\boldsymbol{\theta}}) = (\mathbf{S}'_{1s}, s = 1, \dots, S)', \quad (2.11)$$

em que

$$\mathbf{S}_{1s} = \sum_{t=1}^{T_s} \bar{\mathbf{Z}}_{st} [\boldsymbol{\Sigma}(\bar{\boldsymbol{\theta}}_{st})]^{-1} (\bar{\mathbf{p}}_{st} - \bar{\boldsymbol{\theta}}_{st}), \quad s = 1, \dots, S \quad (2.12)$$

e $\boldsymbol{\Sigma}(\bar{\boldsymbol{\theta}}_{st}) = \frac{1}{n_{st+}} (\mathbf{D}_{\bar{\boldsymbol{\theta}}_{st}} - \bar{\boldsymbol{\theta}}_{st} \bar{\boldsymbol{\theta}}'_{st})$. A matriz hessiana de $\ln L_1(\boldsymbol{\theta} | \mathbf{N})$, com dimensão $S(R - 1) \times S(R - 1)$, pode ser expressa por

$$\mathbf{H}_1(\bar{\boldsymbol{\theta}}) = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{0}_{R-1,R-1} & \cdots & \mathbf{0}_{R-1,R-1} \\ \mathbf{0}_{R-1,R-1} & \mathbf{H}_{12} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_{R-1,R-1} \\ \mathbf{0}_{R-1,R-1} & \cdots & \mathbf{0}_{R-1,R-1} & \mathbf{H}_{1S} \end{bmatrix}, \quad (2.13)$$

em que

$$\mathbf{H}_{1s} = - \sum_{t=1}^{T_s} \bar{\mathbf{Z}}_{st} \left[\mathbf{D}_{\bar{\mathbf{N}}_{st}} \mathbf{D}_{\bar{\boldsymbol{\theta}}_{st}}^{-2} + \frac{n_{stR_{st}}}{(1 - \mathbf{1}'_{R_{st-1}} \bar{\boldsymbol{\theta}}_{st})^2} \mathbf{1}_{R_{st-1}} \mathbf{1}'_{R_{st-1}} \right] \bar{\mathbf{Z}}'_{st}, \quad s = 1, \dots, S \quad (2.14)$$

e $\mathbf{D}_{\hat{\boldsymbol{\theta}}_{st}}^{-2} = \mathbf{D}_{\hat{\boldsymbol{\theta}}_{st}}^{-1} \mathbf{D}_{\hat{\boldsymbol{\theta}}_{st}}^{-1}$. Apresentam-se alguns detalhes da obtenção de (2.12) e (2.14) no Apêndice A.2.

O algoritmo *scoring* de Fisher exige a estimação adicional das probabilidades condicionais de omissão $\{\alpha_{t(cs)}\}$ sob o mecanismo MAR ou $\{\alpha_{t(s)}\}$ sob o mecanismo MCAR, pois

$$E(n_{stc} | \mathbf{N}_{++}, \boldsymbol{\theta}, \{\boldsymbol{\alpha}_{st}^{\text{MAR}}\}) = n_{s++} \mathbf{z}'_{stc} \boldsymbol{\theta}_s \alpha_{t(cs)}, \quad (2.15)$$

$$E(n_{stc} | \mathbf{N}_{++}, \boldsymbol{\theta}, \{\boldsymbol{\alpha}_{st}^{\text{MCAR}}\}) = n_{s++} \mathbf{z}'_{stc} \boldsymbol{\theta}_s \alpha_{t(s)}, \quad (2.16)$$

$s = 1, \dots, S$, $t = 1, \dots, T_s$, $c = 1, \dots, R_{ts}$, em que $\boldsymbol{\alpha}_{st}^{\text{MAR}} = (\alpha_{t(cs)}, c = 1, \dots, R_{st})'$ engloba as probabilidades condicionais de omissão do t -ésimo cenário da s -ésima subpopulação sob o mecanismo MAR e $\boldsymbol{\alpha}_{st}^{\text{MCAR}} = \alpha_{t(s)}$ abrange a probabilidade correspondente sob o mecanismo MCAR. Como o modelo estatístico sob o mecanismo MAR é saturado, após a obtenção das EMV $\{\hat{\boldsymbol{\theta}}_s\}$ de $\{\boldsymbol{\theta}_s\}$, pode-se obter as EMV das probabilidades condicionais de omissão pela propriedade da invariância (veja os detalhes no Apêndice A.3), resultando em

$$\hat{\boldsymbol{\alpha}}_{st}^{\text{MAR}} = \frac{1}{n_{s++}} \mathbf{D}_{\mathbf{z}'_{st} \hat{\boldsymbol{\theta}}_s}^{-1} \mathbf{N}_{st}, \quad s = 1, \dots, S, \quad t = 1, \dots, T_s. \quad (2.17)$$

O fator $L_2(\{\alpha_{t(s)}\} | \{n_{st+}\}; \text{MCAR})$ conduz diretamente aos EMV das probabilidades condicionais de omissão sob o mecanismo MCAR

$$\hat{\boldsymbol{\alpha}}_{st}^{\text{MCAR}} = \hat{\alpha}_{t(s)} = \frac{n_{st+}}{n_{s++}}, \quad s = 1, \dots, S, \quad t = 1, \dots, T_s. \quad (2.18)$$

A matriz de informação de Fisher referente a $\bar{\boldsymbol{\theta}}$ sob o mecanismo MAR é

$$\mathcal{I}_1(\bar{\boldsymbol{\theta}}, \{\boldsymbol{\alpha}_{st}^{\text{MAR}}\}) = \begin{bmatrix} \mathcal{I}_{11}^{\text{MAR}} & \mathbf{0}_{R-1, R-1} & \cdots & \mathbf{0}_{R-1, R-1} \\ \mathbf{0}_{R-1, R-1} & \mathcal{I}_{12}^{\text{MAR}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_{R-1, R-1} \\ \mathbf{0}_{R-1, R-1} & \cdots & \mathbf{0}_{R-1, R-1} & \mathcal{I}_{1S}^{\text{MAR}} \end{bmatrix}, \quad (2.19)$$

em que

$$\mathcal{I}_{1s}^{\text{MAR}} = n_{s++} \sum_{t=1}^{T_s} \bar{\mathbf{Z}}_{st} \left[\mathbf{D}_{\hat{\boldsymbol{\alpha}}_{st}^{\text{MAR}}} \mathbf{D}_{\hat{\boldsymbol{\theta}}_{st}}^{-1} + \frac{\alpha_{t(sR_{ts})}}{1 - \mathbf{1}'_{R_{st}-1} \hat{\boldsymbol{\theta}}_{st}} \mathbf{1}_{R_{st}-1} \mathbf{1}'_{R_{st}-1} \right] \bar{\mathbf{Z}}'_{st}, \quad s = 1, \dots, S$$

e $\bar{\boldsymbol{\alpha}}_{st}^{\text{MAR}} = [\mathbf{I}_{R_{st}-1}, \mathbf{0}_{R_{st}-1}] \boldsymbol{\alpha}_{st}^{\text{MAR}} = (\alpha_{t(cs)}, c = 1, \dots, R_{st} - 1)'$, $s = 1, \dots, S$, $t = 1, \dots, T_s$.

A matriz de informação de Fisher referente a $\bar{\boldsymbol{\theta}}$ sob o mecanismo MCAR é

$$\mathcal{I}_1(\bar{\boldsymbol{\theta}}, \{\boldsymbol{\alpha}_{st}^{\text{MCAR}}\}) = \begin{bmatrix} \mathcal{I}_{11}^{\text{MCAR}} & \mathbf{0}_{R-1, R-1} & \cdots & \mathbf{0}_{R-1, R-1} \\ \mathbf{0}_{R-1, R-1} & \mathcal{I}_{12}^{\text{MCAR}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_{R-1, R-1} \\ \mathbf{0}_{R-1, R-1} & \cdots & \mathbf{0}_{R-1, R-1} & \mathcal{I}_{1S}^{\text{MCAR}} \end{bmatrix}, \quad (2.20)$$

em que

$$\mathcal{I}_{1s}^{\text{MCAR}} = n_{s++} \sum_{t=1}^{T_s} \alpha_{t(s)} \bar{\mathbf{Z}}_{st} \left[\mathbf{D}_{\hat{\boldsymbol{\theta}}_{st}}^{-1} + \frac{1}{1 - \mathbf{1}'_{R_{st-1}} \bar{\boldsymbol{\theta}}_{st}} \mathbf{1}_{R_{st-1}} \mathbf{1}'_{R_{st-1}} \right] \bar{\mathbf{Z}}'_{st}, \quad s = 1, \dots, S.$$

Utilizando-se a matriz de informação de Fisher como medida de precisão, $\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}}^{\text{MAR}} = \left[\mathcal{I}_1 \left(\hat{\boldsymbol{\theta}}, \{ \hat{\boldsymbol{\alpha}}_{st}^{\text{MAR}} \} \right) \right]^{-1}$ e $\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}}^{\text{MCAR}} = \left[\mathcal{I}_1 \left(\hat{\boldsymbol{\theta}}, \{ \hat{\boldsymbol{\alpha}}_{st}^{\text{MCAR}} \} \right) \right]^{-1}$ são estimativas das matrizes de covariâncias assintóticas de $\hat{\boldsymbol{\theta}}$ sob os mecanismos MAR e MCAR, respectivamente. Empregando-se a matriz de informação observada $-\mathbf{H}_1(\hat{\boldsymbol{\theta}})$ como medida de precisão, $\left[-\mathbf{H}_1(\hat{\boldsymbol{\theta}}) \right]^{-1}$ é uma estimativa da matriz de covariâncias assintótica de $\hat{\boldsymbol{\theta}}$ tanto sob o mecanismo MAR quanto sob o MCAR. Interessante é o fato de que a estimativa da matriz de informação de Fisher de $\hat{\boldsymbol{\theta}}$ sob o mecanismo MAR é igual à estimativa da matriz de informação observada de $\hat{\boldsymbol{\theta}}$ sob os mecanismos MAR e MCAR, isto é,

$$\mathcal{I}_1 \left(\hat{\boldsymbol{\theta}}, \{ \hat{\boldsymbol{\alpha}}_{st}^{\text{MAR}} \} \right) = -\mathbf{H}_1(\hat{\boldsymbol{\theta}}), \quad (2.21)$$

conforme se ilustra no Apêndice A.4. Portanto, efetivamente há 3 processos iterativos diferenciados para a obtenção das EMV $\hat{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}$: (a) EM, (b) *scoring* de Fisher sob MCAR e (c) *scoring* de Fisher sob MAR ou Newton-Raphson sob MAR ou MCAR. Como o EMV de $\boldsymbol{\theta}$ é o mesmo sob os mecanismos MAR e MCAR, pode-se utilizar o processo iterativo (b) mesmo que se suponha o mecanismo MAR, desde que após a obtenção de $\hat{\boldsymbol{\theta}}$ se utilize a estimativa da matriz de covariâncias assintótica sob o mecanismo MAR.

Usando (2.8) e o método *delta*, pode-se obter as estimativas das matrizes de covariâncias assintóticas de $\hat{\boldsymbol{\theta}}$ sob os mecanismos MAR e MCAR por meio de $\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}}^{\text{MAR}} = \mathbf{B} \hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}}^{\text{MAR}} \mathbf{B}'$ e $\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}}^{\text{MCAR}} = \mathbf{B} \hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}}^{\text{MCAR}} \mathbf{B}'$.

O teste de ajuste do mecanismo MCAR condicionalmente ao MAR pode ser realizado com a estatística da razão de verossimilhanças de Wilks

$$\begin{aligned} Q_V(\text{MCAR}|\text{MAR}) &= -2 \ln \frac{L_2(\{ \hat{\alpha}_{t(s)} \} | \{ n_{st+} \}; \text{MCAR})}{L_2(\{ \hat{\alpha}_{t(cs)} \} | \mathbf{N}; \text{MAR})} \\ &= -2 \sum_{s=1}^S \sum_{t=1}^{T_s} \sum_{c=1}^{R_{st}} n_{stc} \left[\ln \left(\mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s \right) - \ln \left(\frac{n_{stc}}{n_{st+}} \right) \right] \\ &= -2 \sum_{s=1}^S \mathbf{N}'_s \left[\ln \left(\mathbf{Z}'_s \hat{\boldsymbol{\theta}}_s \right) - \ln \left(\mathbf{p}_s \right) \right], \end{aligned} \quad (2.22)$$

com a estatística de Pearson, que é uma estatística do tipo escore,

$$\begin{aligned} Q_P(\text{MCAR}|\text{MAR}) &= \sum_{s=1}^S \sum_{t=1}^{T_s} \sum_{c=1}^{R_{st}} \frac{\left(n_{stc} - n_{st+} \mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s\right)^2}{n_{st+} \mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s} \\ &= \sum_{s=1}^S \left(\mathbf{p}_s - \mathbf{Z}'_s \hat{\boldsymbol{\theta}}_s\right)' \left[\mathbf{D}_{\mathbf{N}_{s+}} \mathbf{D}_{\mathbf{Z}'_s \hat{\boldsymbol{\theta}}_s}^{-1}\right] \left(\mathbf{p}_s - \mathbf{Z}'_s \hat{\boldsymbol{\theta}}_s\right) \end{aligned} \quad (2.23)$$

ou com a estatística de Neyman

$$\begin{aligned} Q_N(\text{MCAR}|\text{MAR}) &= \sum_{s=1}^S \sum_{t=1}^{T_s} \sum_{c=1}^{R_{st}} \frac{\left(n_{stc} - n_{st+} \mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s\right)^2}{n_{stc}} \\ &= \sum_{s=1}^S \left(\mathbf{p}_s - \mathbf{Z}'_s \hat{\boldsymbol{\theta}}_s\right)' \left[\mathbf{D}_{\mathbf{N}_{s+}} \mathbf{D}_{\mathbf{p}_s}^{-1}\right] \left(\mathbf{p}_s - \mathbf{Z}'_s \hat{\boldsymbol{\theta}}_s\right), \end{aligned} \quad (2.24)$$

em que $\ln(\mathbf{p}_s)$ denota o operador logaritmo (natural) vetorial que consiste em se tomar o logaritmo natural de cada um dos elementos de \mathbf{p}_s e $\mathbf{N}_{s+} = (n_{st+} \otimes \mathbf{1}'_{R_{st}}, t = 1, \dots, T_s)'$ é o vetor com mesma dimensão de \mathbf{N}_s que contém os totais das frequências observadas em cada cenário de omissão da s -ésima subpopulação repetidos seqüencialmente conforme o número de classes de cada cenário (note que $\mathbf{p}_s = \mathbf{D}_{\mathbf{N}_{s+}}^{-1} \mathbf{N}_s$). As três estatísticas são assintoticamente equivalentes com distribuição nula $\chi_{(g)}^2$, em que $g = S + \sum_{s=1}^S (l_s - T_s)$ e $\chi_{(g)}^2$ denota a distribuição qui-quadrado com g graus de liberdade. Apesar da forma geral apresentada em (2.22), como uma frequência observada n_{stc} nula não traz contribuição à função de probabilidade (2.1), nestes casos deve-se utilizar a definição $0 \times \left[\ln\left(\mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s\right) - \ln(0/n_{st+})\right] \equiv 0$ e evitar o cálculo de $\ln(0)$, que poderia trazer complicações computacionais. A estatística de Neyman (2.24) pressupõe $\{n_{stc} > 0\}$ ou, equivalentemente, $\{p_{c(st)} > 0\}$, o que nem sempre ocorre na prática. Portanto, sugere-se substituir frequências eventualmente nulas por algum valor pequeno antes de se obter \mathbf{p}_s e de calcular a inversa de $\mathbf{D}_{\mathbf{p}_s}$. No contexto de MQG, Koch, Imrey e Reinfurt (1972) sugerem que o valor pequeno seja inversamente proporcional ao número de classes e ao total de frequências observadas no t -ésimo cenário de omissão da s -ésima subpopulação, ou seja, que frequências $n_{stc} = 0$ sejam substituídas por $(R_{st} n_{st+})^{-1}$.

As frequências ampliadas esperadas podem ser estimadas por

$$\hat{y}_{str}^{\text{MAR}} = \hat{E}(y_{str} | n_{s++}, \hat{\theta}_{r(s)}, \hat{\alpha}_{t(cs)}) = n_{s++} \hat{\theta}_{r(s)} \hat{\alpha}_{t(cs)}, \quad (2.25)$$

$s = 1, \dots, S, t = 1, \dots, T_s, r = 1, \dots, R, \{c : r \in \mathcal{C}_{stc}\}$, sob o mecanismo MAR, e por

$$\hat{y}_{str}^{\text{MCAR}} = \hat{E}(y_{str} | n_{s++}, \hat{\theta}_{r(s)}, \hat{\alpha}_{t(s)}) = n_{s++} \hat{\theta}_{r(s)} \hat{\alpha}_{t(s)}, \quad (2.26)$$

$s = 1, \dots, S, t = 1, \dots, T_s, r = 1, \dots, R$, sob o mecanismo MCAR.

2.3.2 Estimação por mínimos quadrados generalizados para modelos estruturais saturados sob o mecanismo MCAR

A ignorabilidade do processo de omissão sob o mecanismo MCAR (Seção 1.2.2) para efeitos de inferências freqüentistas sobre $\boldsymbol{\theta}$ permite a utilização da distribuição de \mathbf{N}_s condicional a $\{n_{st+}\}$, que é um produto de T_s distribuições multinomiais para cada uma das $s = 1, \dots, S$ subpopulações

$$\mathbf{N}_{st}|n_{st+}, \bar{\boldsymbol{\theta}}_{st} \stackrel{\text{indep.}}{\sim} M_{R_{st}}(n_{st+}, \bar{\boldsymbol{\theta}}_{st}), \quad t = 1, \dots, T_s. \quad (2.27)$$

A suposição MCAR implica a adoção de uma estrutura linear nos vetores de parâmetros $\bar{\boldsymbol{\theta}}_{s*} = (\bar{\boldsymbol{\theta}}_{st}, t = 1, \dots, T_s)'$, ou seja,

$$\text{MCAR} : \bar{\boldsymbol{\theta}}_{s*} = \bar{\mathbf{Z}}_s' \bar{\boldsymbol{\theta}}_s, \quad s = 1, \dots, S, \quad (2.28)$$

possibilitando a aplicação da metodologia de MQG, também conhecida por GSK (Grizzle, Starmer e Koch, 1969), com a diferença de que o número de categorias de respostas varia de uma distribuição multinomial para a outra.

A aplicação da metodologia de MQG consiste em minimizar a forma quadrática

$$\begin{aligned} Q_N(\bar{\boldsymbol{\theta}}) &= \sum_{s=1}^S \sum_{t=1}^{T_s} (\bar{\mathbf{p}}_{st} - \bar{\boldsymbol{\theta}}_{st})' [\boldsymbol{\Sigma}(\bar{\mathbf{p}}_{st})]^{-1} (\bar{\mathbf{p}}_{st} - \bar{\boldsymbol{\theta}}_{st}) \\ &= \sum_{s=1}^S (\bar{\mathbf{p}}_s - \bar{\boldsymbol{\theta}}_{s*})' [\boldsymbol{\Sigma}_*(\bar{\mathbf{p}}_s)]^{-1} (\bar{\mathbf{p}}_s - \bar{\boldsymbol{\theta}}_{s*}), \end{aligned} \quad (2.29)$$

em que $\boldsymbol{\Sigma}_*(\bar{\mathbf{p}}_s)$ é uma matriz diagonal em blocos com blocos iguais a $\boldsymbol{\Sigma}(\bar{\mathbf{p}}_{st})$, $t = 1, \dots, T_s$, que resultam de $\boldsymbol{\Sigma}(\bar{\boldsymbol{\theta}}_{st})$ substituindo-se $\bar{\boldsymbol{\theta}}_{st}$ por $\bar{\mathbf{p}}_{st}$. Sob o mecanismo MCAR, o estimador de MQG de $\bar{\boldsymbol{\theta}}_s$ é

$$\tilde{\boldsymbol{\theta}}_s = (\bar{\mathbf{Z}}_s [\boldsymbol{\Sigma}_*(\bar{\mathbf{p}}_s)]^{-1} \bar{\mathbf{Z}}_s')^{-1} \bar{\mathbf{Z}}_s [\boldsymbol{\Sigma}_*(\bar{\mathbf{p}}_s)]^{-1} \bar{\mathbf{p}}_s, \quad (2.30)$$

e uma estimativa de sua matriz de covariâncias assintótica é

$$\tilde{\mathbf{V}}_{\tilde{\boldsymbol{\theta}}_s} = (\bar{\mathbf{Z}}_s [\boldsymbol{\Sigma}_*(\bar{\mathbf{p}}_s)]^{-1} \bar{\mathbf{Z}}_s')^{-1}. \quad (2.31)$$

De acordo com (2.7), pode-se obter o estimador de MQG de $\boldsymbol{\theta}_s$ por meio de

$$\tilde{\boldsymbol{\theta}}_s = \mathbf{b}_s + \mathbf{B}_s \tilde{\boldsymbol{\theta}}_s; \quad (2.32)$$

analogamente, uma estimativa da correspondente matriz de covariâncias assintótica é

$$\tilde{\mathbf{V}}_{\tilde{\boldsymbol{\theta}}_s} = \mathbf{B}_s \tilde{\mathbf{V}}_{\tilde{\boldsymbol{\theta}}_s} \mathbf{B}'_s. \quad (2.33)$$

Uma estimativa da matriz de covariâncias assintótica de $\tilde{\boldsymbol{\theta}}$ é

$$\tilde{\mathbf{V}}_{\tilde{\boldsymbol{\theta}}} = \begin{bmatrix} \tilde{\mathbf{V}}_{\tilde{\boldsymbol{\theta}}_1} & \mathbf{0}_{R-1,R-1} & \cdots & \mathbf{0}_{R-1,R-1} \\ \mathbf{0}_{R-1,R-1} & \tilde{\mathbf{V}}_{\tilde{\boldsymbol{\theta}}_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_{R-1,R-1} \\ \mathbf{0}_{R-1,R-1} & \cdots & \mathbf{0}_{R-1,R-1} & \tilde{\mathbf{V}}_{\tilde{\boldsymbol{\theta}}_S} \end{bmatrix},$$

e a de $\tilde{\boldsymbol{\theta}}$ é $\tilde{\mathbf{V}}_{\tilde{\boldsymbol{\theta}}} = \mathbf{B} \tilde{\mathbf{V}}_{\tilde{\boldsymbol{\theta}}} \mathbf{B}'$.

O teste de ajuste do mecanismo MCAR pode ser realizado com a estatística de Neyman

$$Q_N(\text{MCAR}) = \sum_{s=1}^S \left(\bar{\mathbf{p}}_s - \bar{\mathbf{Z}}'_s \tilde{\boldsymbol{\theta}}_s \right)' [\boldsymbol{\Sigma}_* (\bar{\mathbf{p}}_s)]^{-1} \left(\bar{\mathbf{p}}_s - \bar{\mathbf{Z}}'_s \tilde{\boldsymbol{\theta}}_s \right), \quad (2.34)$$

que tem distribuição nula aproximada $\chi^2_{(g)}$ para grandes valores de $\{n_{stc}\}$, em que $g = S + \sum_{s=1}^S (l_s - T_s)$.

Nas expressões (2.29), (2.30), (2.31) e (2.34) assume-se que $\boldsymbol{\Sigma}_* (\bar{\mathbf{p}}_s)$ seja não singular para que se possa calcular sua inversa unicamente. Sendo assim, necessita-se $\{p_{c(st)} > 0\}$ ou, equivalentemente, $\{n_{stc} > 0\}$. Como isso nem sempre acontece na prática, Koch, Imrey e Reinfurt (1972) sugerem que se substituam frequências eventualmente nulas por $(R_{st} n_{st+})^{-1}$ ou outro valor pequeno qualquer.

As frequências ampliadas esperadas podem ser estimadas por

$$\tilde{y}_{str} = \tilde{E}(y_{str} | n_{st+}, \tilde{\theta}_{r(s)}) = n_{st+} \tilde{\theta}_{r(s)}, \quad (2.35)$$

$s = 1, \dots, S$, $t = 1, \dots, T_s$, $r = 1, \dots, R$.

2.4 Estimação das probabilidades de categorização para modelos estruturais não saturados

Como no caso de dados completos, em geral há necessidade de se modelar as probabilidades de categorização por meio de estruturas não saturadas com a finalidade de dar respostas a questões de interesse. Na Seção 2.4.1, mostra-se algumas dessas estruturas para atender aos objetivos dos Exemplos 1, 2 e 3. Na Seção 2.4.2 expõe-se o processo de estimação de $\boldsymbol{\theta}$ pela metodologia de MV sob os mecanismos MAR e MCAR impondo estruturas lineares e log-lineares em $\boldsymbol{\theta}$ e na Seção 2.4.3, pela metodologia de MQG sob os mecanismos MAR, MCAR e MNAR impondo estruturas funcionais lineares. Na estimação por MV (MQG) utilizam-se as estatísticas de razão de verossimilhanças, Pearson, Neyman e Wald (apenas Wald) para testar o ajuste dessas estruturas e utiliza-se a estatística de Wald para realizar testes de hipóteses lineares redutoras.

Descreve-se as estruturas funcionais lineares, ou particularmente, lineares e log-lineares, de maneira sucinta. Para maiores detalhes, veja Paulino e Singer (2006) ou, unicamente no caso de dados completos, Forthofer e Lehnen (1981) e Koch, Imrey, Singer, Atkinson e Stokes (1985).

2.4.1 Exemplos de modelos estruturais

Exemplo 1 (Continuação, pp.1 e 55)

Pode-se utilizar a razão de chances, ou o seu logaritmo, para avaliar a associação entre o hábito de fumo da mãe e o peso do recém-nascido. O logaritmo da razão de chances tem a vantagem de poder ser expresso na forma do seguinte modelo log-linear

$$\ln \left(\frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} \right) = \ln(\pi_{11}) - \ln(\pi_{12}) - \ln(\pi_{21}) + \ln(\pi_{22}) = \mathbf{A} \ln(\boldsymbol{\theta}) = \mathbf{X}_L \boldsymbol{\beta}, \quad (2.36)$$

em que $(\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})' = (\theta_r, r = 1, \dots, 4)' = \boldsymbol{\theta}$, $\mathbf{A} = (1, -1, -1, 1)$, $\mathbf{X}_L = \mathbf{1}$ e $\boldsymbol{\beta} = \beta$. O ajuste de (2.36) não adiciona restrições, uma vez que $\mathbf{A} \ln(\boldsymbol{\theta})$ e $\boldsymbol{\beta}$ são escalares. No entanto, este modelo permite obter uma estimativa da razão de chances por meio de $\exp(\hat{\beta})$. Querendo verificar se o logaritmo da razão de chances é nulo, basta testar a hipótese

$$\ln \left(\frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} \right) = 0 \iff \ln(\pi_{11}) - \ln(\pi_{12}) - \ln(\pi_{21}) + \ln(\pi_{22}) = 0 \iff \mathbf{U} \ln(\boldsymbol{\theta}) = 0, \quad (2.37)$$

em que $\mathbf{U} = (1, -1, -1, 1)$, ou, equivalentemente avaliar o ajuste do modelo

$$\ln(\boldsymbol{\theta}) = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \nu \\ \beta_1 \\ \beta_2 \end{pmatrix} = \mathbf{1}_4\nu + \mathbf{X}\boldsymbol{\beta}, \quad (2.38)$$

em que $\mathbf{X} = ([0, 1]' \otimes \mathbf{1}_2, \mathbf{1}_2 \otimes [0, 1]')$, $\boldsymbol{\beta} = (\beta_1, \beta_2)'$, ν é uma componente associada à restrição natural $\mathbf{1}'_4\boldsymbol{\theta} = 1$, $\exp(\beta_1) = \pi_{21}/\pi_{11} = \pi_{22}/\pi_{12}$ e $\exp(\beta_2) = \pi_{12}/\pi_{11} = \pi_{22}/\pi_{21}$. Note que o vetor coluna \mathbf{U} de (2.37) é ortogonal às colunas da matriz $(\mathbf{1}_4, \mathbf{X})$ de (2.38), ou seja, $\mathbf{U}(\mathbf{1}_4, \mathbf{X}) = \mathbf{0}_{1,3}$. ■

Exemplo 2 (Continuação, p.56)

Primeiramente, pode-se avaliar se a associação entre o hábito do fumo da mãe e o estado respiratório da criança tem a mesma magnitude para as duas cidades, verificando que os logaritmos das razões de chances adjacentes, por exemplo, são iguais para as duas cidades, ou seja, que

$$\ln \left(\frac{\pi_{ij(1)}\pi_{i+1,j+1(1)}}{\pi_{i,j+1(1)}\pi_{i+1,j(1)}} \right) = \ln \left(\frac{\pi_{ij(2)}\pi_{i+1,j+1(2)}}{\pi_{i,j+1(2)}\pi_{i+1,j(2)}} \right), \quad i, j = 1, 2, \quad (2.39)$$

em que $(\pi_{11(s)}, \pi_{12(s)}, \pi_{13(s)}, \pi_{21(s)}, \pi_{22(s)}, \pi_{23(s)}, \pi_{31(s)}, \pi_{32(s)}, \pi_{33(s)})' = (\theta_{r(s)}, r = 1, \dots, 9)' = \boldsymbol{\theta}_s$, $s = 1, 2$. Uma maneira de se escrever (2.39) na forma de um modelo log-linear

$$\mathbf{A} \ln(\boldsymbol{\theta}) = \mathbf{X}_L \boldsymbol{\beta} \quad (2.40)$$

é considerar

$$\mathbf{A} = \left(\mathbf{I}_2 \otimes \begin{bmatrix} 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 \end{bmatrix} \right) = \mathbf{I}_2 \otimes \mathbf{E} \otimes \mathbf{E},$$

em que

$$\mathbf{E} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix},$$

$\boldsymbol{\theta} = (\boldsymbol{\theta}'_1, \boldsymbol{\theta}'_2)'$, $\mathbf{X}_L = \mathbf{1}_2 \otimes \mathbf{I}_4$ e $\boldsymbol{\beta} = (\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22})'$. Com este modelo, $\exp(\beta_{ij})$ pode ser interpretado como o quociente (comum para as duas cidades) entre a chance de uma criança

de uma mãe com o $(i + 1)$ -ésimo hábito de fumo ter o $(j + 1)$ -ésimo estado respiratório em relação a ter o j -ésimo estado respiratório e a chance correspondente para uma mãe com o i -ésimo hábito de fumo. Não existindo interesse em se estimar os parâmetros β_{ij} , mas apenas avaliar o ajuste do modelo, pode-se utilizar a formulação equivalente em termos de restrições

$$\mathbf{U}_L \mathbf{A} \ln(\boldsymbol{\theta}) = \mathbf{0}_4, \tag{2.41}$$

em que $\mathbf{U}_L = ([1, -1] \otimes \mathbf{I}_4)$ e $\mathbf{A} = \mathbf{I}_2 \otimes \mathbf{E} \otimes \mathbf{E}$. Note que as linhas da matriz \mathbf{U}_L de (2.41) são ortogonais às colunas da matriz \mathbf{X}_L de (2.40), ou seja, $\mathbf{U}_L \mathbf{X}_L = \mathbf{0}_{4,4}$.

Outra hipótese de interesse pode ser em avaliar se as razões de chances adjacentes são iguais, ou seja, se $\beta_{11} = \beta_{12} = \beta_{21} = \beta_{22} = \beta$. Esta hipótese pode ser testada ajustando o modelo log-linear $\mathbf{A} \ln(\boldsymbol{\theta}) = \mathbf{X}_L \boldsymbol{\beta}$, em que $\mathbf{A} = \mathbf{I}_2 \otimes \mathbf{E} \otimes \mathbf{E}$, $\mathbf{X}_L = \mathbf{1}_8$ e $\boldsymbol{\beta} = \beta$ é o logaritmo da razão de chances adjacentes comum. Paulino e Singer (2006) mostram que esta formulação matricial coincide com a modelo log-linear de associação linear por linear. Por fim, existindo interesse em se testar a independência entre o hábito do fumo da mãe e o estado respiratório da criança, pode-se avaliar o ajuste do modelo log-linear sob a formulação $\mathbf{U}_L \mathbf{A} \ln(\boldsymbol{\theta}) = \mathbf{0}_8$, com $\mathbf{U}_L = \mathbf{I}_8$ e $\mathbf{A} = \mathbf{I}_2 \otimes \mathbf{E} \otimes \mathbf{E}$.

■

Exemplo 3 (Continuação, p.58)

Para avaliar a hipótese de homogeneidade de distribuições marginais dos graus de suscetibilidade à cárie dentária obtidas pelos dois métodos pode-se verificar o ajuste do modelo (estritamente) linear

$$\mathbf{A}\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta} \tag{2.42}$$

com

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} [\mathbf{I}_2, \mathbf{0}_2] \otimes \mathbf{1}'_3 \\ \mathbf{1}'_3 \otimes [\mathbf{I}_2, \mathbf{0}_2] \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{1}_2 \otimes \mathbf{I}_2,$$

e $\boldsymbol{\beta} = (\beta_1, \beta_2)'$. Neste modelo, β_1 [β_2] é a probabilidade marginal (comum aos dois métodos) de se classificar uma criança na categoria de alto [médio] grau de suscetibilidade à cárie dentária, ou seja,

$$\beta_i = \pi_{i+} = \pi_{+i}, \quad i = 1, 2, 3,$$

em que $(\pi_{11}, \pi_{12}, \pi_{13}, \pi_{21}, \pi_{22}, \pi_{23}, \pi_{31}, \pi_{32}, \pi_{33})' = (\theta_r, r = 1, \dots, 9)' = \boldsymbol{\theta}$, $\pi_{i+} = \sum_{j=1}^3 \pi_{ij}$, $i = 1, 2, 3$ e $\pi_{+j} = \sum_{i=1}^3 \pi_{ij}$, $j = 1, 2, 3$. Caso não exista interesse em se estimar $\boldsymbol{\beta}$, mas apenas avaliar o ajuste do modelo, pode-se utilizar a formulação equivalente em termos de restrições

$$\pi_{i+} - \pi_{+i} = 0, \quad i = 1, 2, 3, \iff \mathbf{UA}\boldsymbol{\theta} = \mathbf{0}_2, \quad (2.43)$$

com $\mathbf{U} = ([1, -1] \otimes \mathbf{I}_2)$ e $\mathbf{A} = ([\mathbf{I}_2, \mathbf{0}_2]' \otimes \mathbf{1}_3, \mathbf{1}_3 \otimes [\mathbf{I}_2, \mathbf{0}_2]')'$. Note que as linhas da matriz \mathbf{U} de (2.43) são ortogonais às colunas da matriz \mathbf{X} de (2.42), ou seja, $\mathbf{UX} = \mathbf{0}_{2,2}$.

Para avaliar a concordância entre os dois métodos pode-se utilizar a medida *kappa* de Cohen (1960)

$$\kappa = \frac{\sum_{i=1}^3 \pi_{ii} - \sum_{i=1}^3 \pi_{i+} \pi_{+i}}{1 - \sum_{i=1}^3 \pi_{i+} \pi_{+i}}. \quad (2.44)$$

Esta medida pode ser escrita na forma do seguinte modelo funcional linear

$$\mathbf{F} = \boldsymbol{\pi}_1 + \exp(\mathbf{A}_4 \ln\{\mathbf{A}_3 \exp[\mathbf{A}_2 \ln(\mathbf{A}_1 \boldsymbol{\theta})]\}) \quad (2.45)$$

com

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} (\mathbf{1}'_2 \otimes [1, \mathbf{0}'_3], 1) \\ \mathbf{1}'_9 \\ \mathbf{I}_3 \otimes \mathbf{1}'_3 \\ \mathbf{1}'_3 \otimes \mathbf{I}_3 \end{bmatrix},$$

$$\mathbf{A}_2 = \left[\begin{array}{c|cccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right] = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0}_{2,6} \\ \mathbf{0}_{3,2} & \mathbf{1}'_2 \otimes \mathbf{I}_3 \end{bmatrix},$$

$$\mathbf{A}_3 = \left[\begin{array}{c|cc} 1 & 1 & -2 & -2 & -2 \\ 0 & 1 & -1 & -1 & -1 \end{array} \right] = [(1, 0)', \mathbf{1}_2, -(2, 1)' \mathbf{1}'_3], \quad \mathbf{A}_4 = [1, -1] \quad \text{e} \quad \boldsymbol{\pi}_1 = -1.$$

Para entender a formulação do modelo, note que

$$\mathbf{A}_1 \boldsymbol{\theta} = \begin{bmatrix} \pi_{11} + \pi_{22} + \pi_{33} \\ 1 \\ \pi_{1+} \\ \pi_{2+} \\ \pi_{3+} \\ \pi_{+1} \\ \pi_{+2} \\ \pi_{+3} \end{bmatrix}, \quad \exp[\mathbf{A}_2 \ln(\mathbf{A}_1 \boldsymbol{\theta})] = \begin{bmatrix} \pi_{11} + \pi_{22} + \pi_{33} \\ 1 \\ \pi_{1+}\pi_{+1} \\ \pi_{2+}\pi_{+2} \\ \pi_{3+}\pi_{+3} \end{bmatrix},$$

$$\mathbf{A}_3 \exp[\mathbf{A}_2 \ln(\mathbf{A}_1 \boldsymbol{\theta})] = \begin{bmatrix} \sum_{i=1}^3 \pi_{ii} + 1 - 2 \sum_{i=1}^3 \pi_{i+}\pi_{+i} \\ 1 - \sum_{i=1}^3 \pi_{i+}\pi_{+i} \end{bmatrix},$$

$$\exp(\mathbf{A}_4 \ln\{\mathbf{A}_3 \exp[\mathbf{A}_2 \ln(\mathbf{A}_1 \boldsymbol{\theta})]\}) = \frac{\sum_{i=1}^3 \pi_{ii} + 1 - 2 \sum_{i=1}^3 \pi_{i+}\pi_{+i}}{1 - \sum_{i=1}^3 \pi_{i+}\pi_{+i}} = \kappa + 1$$

e $\boldsymbol{\pi}_1 + \exp(\mathbf{A}_4 \ln\{\mathbf{A}_3 \exp[\mathbf{A}_2 \ln(\mathbf{A}_1 \boldsymbol{\theta})]\}) = \kappa$. Apesar da aparente complicação adicional em somar o termo do denominador da medida *kappa* em seu numerador e no final subtrair o valor 1, se isto não tivesse sido feito, não seria possível aplicar a função logarítmica na antepenúltima expressão para os casos em que a medida assume valores negativos, ou seja, quando a concordância entre os dois métodos é menor do que a esperada sob independência. Ainda que esses casos sejam raros, não são impossíveis e devem ser contemplados, uma vez que a medida *kappa* pode assumir valores negativos. Segundo Fleiss *et al.* (2003), se há uma concordância completa, $\kappa = 1$; se a concordância é maior (igual) [menor] que a esperada sob independência, $\kappa > 0$ ($=0$) [< 0] e o valor mínimo de κ depende das probabilidades marginais, assumindo usualmente valores entre -1 e 0 , mas podendo chegar a ser igual a -1 em uma tabela 2×2 quando $\sum_{i=1}^2 \pi_{i+}\pi_{+i} = 0.5$.

Devido ao caráter ordinal do grau de suscetibilidade à cárie dentária, pode ser mais ade-

quando utilizar a medida *kappa* ponderada (Spitzer, Cohen, Fleiss e Endicott, 1967)

$$\kappa_w = \frac{\sum_{i=1}^3 \sum_{j=1}^3 w_{ij} \pi_{ij} - \sum_{i=1}^3 \sum_{j=1}^3 w_{ij} \pi_{i+\pi+j}}{1 - \sum_{i=1}^3 \sum_{j=1}^3 w_{ij} \pi_{i+\pi+j}}, \quad (2.46)$$

em que $\{0 \leq w_{ij} \leq 1\}$ são pesos que quantificam o grau de concordância entre as categorias i e j , com o valor $w_{ii} = 1$ indicando uma concordância total e $w_{ij} = w_{ji} < 1$ assume simetria e alguma discordância entre as categorias $i \neq j$. O modelo funcional linear para esta medida é o mesmo utilizado para a medida *kappa* de Cohen em (2.45), redefinindo as matrizes

$$\mathbf{A}_1 = \left[\begin{array}{ccccccccc} w_{11} & w_{12} & w_{13} & w_{21} & w_{22} & w_{23} & w_{31} & w_{32} & w_{33} \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ \hline 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right] = \left[\begin{array}{c} \mathbf{W}' \\ \mathbf{1}'_9 \\ \mathbf{I}_3 \otimes \mathbf{1}'_3 \\ \mathbf{1}'_3 \otimes \mathbf{I}_3 \end{array} \right],$$

$$\mathbf{A}_2 = \left[\begin{array}{cc|cccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right] = \left[\begin{array}{cc} \mathbf{I}_2 & \mathbf{0}_{2,6} \\ \mathbf{0}_{9,2} & (\mathbf{I}_3 \otimes \mathbf{1}_3, \mathbf{1}_3 \otimes \mathbf{I}_3) \end{array} \right] \quad \text{e}$$

$$\begin{aligned} \mathbf{A}_3 &= \left[\begin{array}{c|c|cccccccc} 1 & 1 & -2w_{11} & -2w_{12} & -2w_{13} & -2w_{21} & -2w_{22} & -2w_{23} & -2w_{31} & -2w_{32} & -2w_{33} \\ 0 & 1 & -w_{11} & -w_{12} & -w_{13} & -w_{21} & -w_{22} & -w_{23} & -w_{31} & -w_{32} & -w_{33} \end{array} \right] \\ &= [(1, 0)', \mathbf{1}_2, -(2, 1)' \mathbf{W}'], \end{aligned}$$

em que $\mathbf{W} = (w_{11}, w_{12}, w_{13}, w_{21}, w_{22}, w_{23}, w_{31}, w_{32}, w_{33})'$ é um vetor contendo os pesos.

Novamente, para facilitar a compreensão da formulação matricial do modelo, observe que

$$\mathbf{A}_1 \boldsymbol{\theta} = \begin{bmatrix} \sum_{i=1}^3 \sum_{j=1}^3 w_{ij} \pi_{ij} \\ 1 \\ \pi_{1+} \\ \pi_{2+} \\ \pi_{3+} \\ \pi_{+1} \\ \pi_{+2} \\ \pi_{+3} \end{bmatrix}, \quad \exp[\mathbf{A}_2 \ln(\mathbf{A}_1 \boldsymbol{\theta})] = \begin{bmatrix} \sum_{i=1}^3 \sum_{j=1}^3 w_{ij} \pi_{ij} \\ 1 \\ \pi_{1+} \pi_{+1} \\ \pi_{1+} \pi_{+2} \\ \pi_{1+} \pi_{+3} \\ \pi_{2+} \pi_{+1} \\ \pi_{2+} \pi_{+2} \\ \pi_{2+} \pi_{+3} \\ \pi_{3+} \pi_{+1} \\ \pi_{3+} \pi_{+2} \\ \pi_{3+} \pi_{+3} \end{bmatrix},$$

$$\mathbf{A}_3 \exp[\mathbf{A}_2 \ln(\mathbf{A}_1 \boldsymbol{\theta})] = \begin{bmatrix} \sum_{i=1}^3 \sum_{j=1}^3 w_{ij} \pi_{ij} + 1 - 2 \sum_{i=1}^3 \sum_{j=1}^3 w_{ij} \pi_{i+} \pi_{+j} \\ 1 - \sum_{i=1}^3 \sum_{j=1}^3 w_{ij} \pi_{i+} \pi_{+j} \end{bmatrix}$$

e $\boldsymbol{\pi}_1 + \exp(\mathbf{A}_4 \ln\{\mathbf{A}_3 \exp[\mathbf{A}_2 \ln(\mathbf{A}_1 \boldsymbol{\theta})]\}) = \kappa_w$.

Fleiss e Cohen (1973) sugerem pesos

$$w_{ij} = 1 - \frac{(i-j)^2}{(I-1)^2}, \quad i, j = 1, \dots, I, \quad (2.47)$$

em que $I = 3$ indica o número de classificações obtidas de cada método, e mostram que κ_w fica com expressão idêntica ao coeficiente de correlação intraclassa, com exceção de um termo inversamente proporcional ao tamanho da amostra. Sob essa sugestão, obtém-se $\mathbf{W}_1 = (1, 0.75, 0, 0.75, 1, 0.75, 0, 0.75, 1)'$. Outra possibilidade sugerida por Cicchetti e Allison (1971) é

$$w_{ij} = 1 - \frac{|i-j|}{I-1}, \quad i, j = 1, \dots, I, \quad (2.48)$$

obtendo $\mathbf{W}_2 = (1, 0.5, 0, 0.5, 1, 0.5, 0, 0.5, 1)'$. \mathbf{W}_1 e \mathbf{W}_2 são, geralmente, conhecidos como pesos quadráticos e absolutos, respectivamente.

■

2.4.2 Estimação por máxima verossimilhança para modelos estruturais não saturados sob os mecanismos MAR e MCAR

Considera-se aqui modelos (estritamente) lineares da forma

$$M_L : \mathbf{A}\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta}, \quad (2.49)$$

em que a matriz \mathbf{A} , de dimensão $u \times SR$, define as u funções lineares de interesse e possui posto $r(\mathbf{A}) = u \leq S(R - 1)$; a matriz \mathbf{X} , de dimensão $u \times p$, com posto $r(\mathbf{X}) = p \leq u$ especifica o modelo e o vetor $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$, de dimensão $p \times 1$, contém como elementos os p parâmetros desconhecidos. À expressão (2.49), que corresponde à formulação do modelo linear geral em equações livres, contrapõe-se a formulação equivalente em termos de restrições (veja Paulino e Singer, 2006, Apêndice A, por exemplo)

$$M_L : \mathbf{U}\mathbf{A}\boldsymbol{\theta} = \mathbf{0}_{u-p}, \quad (2.50)$$

em que a matriz \mathbf{U} , de dimensão $(u - p) \times u$, contém as $u - p$ restrições, possui posto máximo e suas linhas são ortogonais às colunas de \mathbf{X} , ou seja, $\mathbf{U}\mathbf{X} = \mathbf{0}_{(u-p),p}$. Para a especificação do modelo, deve-se incluir também as S restrições naturais $\sum_{r=1}^R \theta_{r(s)} = 1$, $s = 1, \dots, S$, representadas matricialmente por

$$[\mathbf{I}_S \otimes \mathbf{1}'_R] \boldsymbol{\theta} = \mathbf{1}_S. \quad (2.51)$$

Sendo assim, supõe-se que as linhas de \mathbf{A} são linearmente independentes das colunas da matriz $\mathbf{I}_S \otimes \mathbf{1}_R$ definidora das restrições naturais, *i.e.*, $r(\mathbf{A}', \mathbf{I}_S \otimes \mathbf{1}_R) = u + S$.

Para aproveitar o desenvolvimento utilizado na Seção 2.3.1, em função de $\bar{\boldsymbol{\theta}}$, é conveniente reescrever (2.49) conforme a proposta apresentada a seguir. A junção de (2.49) e (2.51) conduz a

$$\begin{pmatrix} \mathbf{A} \\ \mathbf{I}_S \otimes \mathbf{1}'_R \end{pmatrix} \boldsymbol{\theta} = \begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{1}_S \end{pmatrix}. \quad (2.52)$$

Portanto, se $r(\mathbf{A}) = u = S(R - 1)$, pode-se obter $\boldsymbol{\theta}$ unicamente a partir de (\mathbf{A}, \mathbf{X}) e $\boldsymbol{\beta}$ de acordo com

$$\boldsymbol{\theta}(\boldsymbol{\beta}) = \begin{pmatrix} \mathbf{A} \\ \mathbf{I}_S \otimes \mathbf{1}'_R \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{1}_S \end{pmatrix} \quad (2.53)$$

e, $\bar{\boldsymbol{\theta}}$, com

$$\bar{\boldsymbol{\theta}}(\boldsymbol{\beta}) = (\mathbf{I}_S \otimes [\mathbf{I}_{R-1}, \mathbf{0}_{R-1}]) \begin{pmatrix} \mathbf{A} \\ \mathbf{I}_S \otimes \mathbf{1}'_R \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{1}_S \end{pmatrix}. \quad (2.54)$$

Quando $r(\mathbf{A}) = u < S(R-1)$, necessita-se uma matriz \mathbf{A}_0 , com dimensão $(S[R-1]-u) \times SR$, base do complemento ortogonal de $(\mathbf{A}', \mathbf{I}_S \otimes \mathbf{1}'_R)'$, para ampliar o modelo (2.52) de modo a obter-se o novo modelo

$$\begin{pmatrix} \mathbf{A} \\ \mathbf{I}_S \otimes \mathbf{1}'_R \\ \mathbf{A}_0 \end{pmatrix} \boldsymbol{\theta} = \begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{1}_S \\ \boldsymbol{\beta}_0 \end{pmatrix}, \quad (2.55)$$

que engloba o anterior, mas possui também $S[R-1]-u$ parâmetros adicionais, de incômodo (*nuisance*), incluídos em $\boldsymbol{\beta}_0$. Neste caso, obtém-se $\boldsymbol{\theta}$ em função de $(\boldsymbol{\beta}, \boldsymbol{\beta}_0)$ por meio de

$$\boldsymbol{\theta}(\boldsymbol{\beta}, \boldsymbol{\beta}_0) = \begin{pmatrix} \mathbf{A} \\ \mathbf{I}_S \otimes \mathbf{1}'_R \\ \mathbf{A}_0 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{1}_S \\ \boldsymbol{\beta}_0 \end{pmatrix} \quad (2.56)$$

e, $\bar{\boldsymbol{\theta}}$, de

$$\bar{\boldsymbol{\theta}}(\boldsymbol{\beta}, \boldsymbol{\beta}_0) = (\mathbf{I}_S \otimes [\mathbf{I}_{R-1}, \mathbf{0}_{R-1}]) \begin{pmatrix} \mathbf{A} \\ \mathbf{I}_S \otimes \mathbf{1}'_R \\ \mathbf{A}_0 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{1}_S \\ \boldsymbol{\beta}_0 \end{pmatrix}. \quad (2.57)$$

Por simplicidade, no desenvolvimento a seguir, suprime-se o parâmetro $\boldsymbol{\beta}_0$, mencionando-se apenas $\boldsymbol{\beta}$. No entanto, sempre que $u < S(R-1)$, $\boldsymbol{\beta}_0$ também é utilizado. Para exemplos e mais detalhes sobre a ampliação de modelos lineares, veja Paulino e Singer (2006) e Koch *et al.* (1985).

A incorporação desta estrutura linear na log-verossimilhança $\ln L_1(\boldsymbol{\theta}(\boldsymbol{\beta}) | \mathbf{N})$ e a sua diferenciação em relação a $\boldsymbol{\beta}^3$ conduzem ao vetor *score*

$$\mathbf{S}_{1L}(\boldsymbol{\beta}) = \mathbf{W}'\mathbf{S}_1(\bar{\boldsymbol{\theta}}(\boldsymbol{\beta})), \quad (2.58)$$

e às seguintes matrizes de informação de Fisher obtidas sob os mecanismos MAR e MCAR, respectivamente,

$$\mathcal{I}_{1L}(\boldsymbol{\beta}, \{\boldsymbol{\alpha}_{st}^{\text{MAR}}\}) = \mathbf{W}'\mathcal{I}_1(\bar{\boldsymbol{\theta}}(\boldsymbol{\beta}), \{\boldsymbol{\alpha}_{st}^{\text{MAR}}\})\mathbf{W}, \quad (2.59)$$

$$\mathcal{I}_{1L}(\boldsymbol{\beta}, \{\boldsymbol{\alpha}_{st}^{\text{MCAR}}\}) = \mathbf{W}'\mathcal{I}_1(\bar{\boldsymbol{\theta}}(\boldsymbol{\beta}), \{\boldsymbol{\alpha}_{st}^{\text{MCAR}}\})\mathbf{W}, \quad (2.60)$$

³Por meio da diferenciação matricial em cadeia, tem-se

$$\frac{\partial \ln L_1(\bar{\boldsymbol{\theta}}(\boldsymbol{\beta}) | \mathbf{N})}{\partial \boldsymbol{\beta}} = \left[\frac{\partial \bar{\boldsymbol{\theta}}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \right]' \frac{\partial \ln L_1(\bar{\boldsymbol{\theta}} | \mathbf{N})}{\partial \bar{\boldsymbol{\theta}}} \quad \text{e} \quad \frac{\partial^2 \ln L_1(\bar{\boldsymbol{\theta}}(\boldsymbol{\beta}) | \mathbf{N})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \left[\frac{\partial \bar{\boldsymbol{\theta}}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \right]' \frac{\partial^2 \ln L_1(\bar{\boldsymbol{\theta}} | \mathbf{N})}{\partial \bar{\boldsymbol{\theta}} \partial \bar{\boldsymbol{\theta}'}} \frac{\partial \bar{\boldsymbol{\theta}}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}$$

em que

$$\mathbf{W} = (\mathbf{I}_S \otimes [\mathbf{I}_{R-1}, \mathbf{0}_{R-1}]) \begin{pmatrix} \mathbf{A} \\ \mathbf{I}_S \otimes \mathbf{1}'_R \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X} \\ \mathbf{0}_{S,p} \end{pmatrix}, \quad (2.61)$$

se $u = S(R-1)$, ou

$$\mathbf{W} = (\mathbf{I}_S \otimes [\mathbf{I}_{R-1}, \mathbf{0}_{R-1}]) \begin{pmatrix} \mathbf{A} \\ \mathbf{I}_S \otimes \mathbf{1}'_R \\ \mathbf{A}_0 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X} & \mathbf{0}_{u,S(R-1)-u} \\ \mathbf{0}_{S,p} & \mathbf{0}_{S,S(R-1)-u} \\ \mathbf{0}_{S(R-1)-u,p} & \mathbf{I}_{S(R-1)-u} \end{pmatrix}, \quad (2.62)$$

se $u < S(R-1)$; $\mathbf{S}_1(\bar{\boldsymbol{\theta}}(\boldsymbol{\beta}))$ é apresentado em (2.11); $\mathcal{I}_1(\bar{\boldsymbol{\theta}}(\boldsymbol{\beta}), \{\boldsymbol{\alpha}_{st}^{\text{MAR}}\})$ e $\mathcal{I}_1(\bar{\boldsymbol{\theta}}(\boldsymbol{\beta}), \{\boldsymbol{\alpha}_{st}^{\text{MCAR}}\})$ são descritos em (2.19) e (2.20) e $\bar{\boldsymbol{\theta}}(\boldsymbol{\beta})$, em (2.54), se $u = S(R-1)$, ou em (2.57), se $u < S(R-1)$.

O uso do vetor escore e da matriz de informação de Fisher possibilita a obtenção das EMV $\hat{\boldsymbol{\beta}}$ de $\boldsymbol{\beta}$ por meio do método *scoring* de Fisher. O processo iterativo pode ser inicializado com a estimativa de MQG (caso particular do que será exposto na Seção 2.4.3)

$$\hat{\boldsymbol{\beta}}^{(0)} = \left\{ \mathbf{X}' [\mathbf{A} \hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}} \mathbf{A}']^{-1} \mathbf{X} \right\}^{-1} \mathbf{X}' [\mathbf{A} \hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}} \mathbf{A}']^{-1} \mathbf{A} \hat{\boldsymbol{\theta}}, \quad (2.63)$$

se $u = S(R-1)$, ou por meio de expressão análoga proveniente da substituição de \mathbf{A} por $(\mathbf{A}', \mathbf{A}'_0)'$ e \mathbf{X} por

$$\begin{pmatrix} \mathbf{X} & \mathbf{0}_{u,S(R-1)-u} \\ \mathbf{0}_{S(R-1)-u,p} & \mathbf{I}_{S(R-1)-u} \end{pmatrix}, \quad (2.64)$$

se $u < S(R-1)$, em que $\hat{\boldsymbol{\theta}}$ é a EMV irrestrita de $\boldsymbol{\theta}$ e $\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}}$ é uma estimativa proveniente de um estimador consistente da matriz de covariâncias assintótica sob o mecanismo MAR ($\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}}^{\text{MAR}}$) ou MCAR ($\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}}^{\text{MCAR}}$), obtidas conforme sugestão apresentada na Seção 2.3.1.

As estimativas das matrizes de covariâncias assintóticas de $\hat{\boldsymbol{\beta}}$ sob os mecanismos MAR e MCAR são $\hat{\mathbf{V}}_{\hat{\boldsymbol{\beta}L}^{\text{MAR}}}^{\text{MAR}} = [\mathcal{I}_{1L}(\hat{\boldsymbol{\beta}}, \{\hat{\boldsymbol{\alpha}}_{st}^{\text{MAR}}\})]^{-1}$ e $\hat{\mathbf{V}}_{\hat{\boldsymbol{\beta}L}^{\text{MCAR}}}^{\text{MCAR}} = [\mathcal{I}_{1L}(\hat{\boldsymbol{\beta}}, \{\hat{\boldsymbol{\alpha}}_{st}^{\text{MCAR}}\})]^{-1}$. Por meio de (2.54), se $u = S(R-1)$, ou (2.57), se $u < S(R-1)$, obtém-se a EMV $\hat{\boldsymbol{\theta}}(M_L)$ de $\bar{\boldsymbol{\theta}}$ sob M_L e, com o método *delta*, suas respectivas estimativas das matrizes de covariâncias assintóticas sob os mecanismos MAR e MCAR, a saber $\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}(M_L)}^{\text{MAR}} = \mathbf{W} \hat{\mathbf{V}}_{\hat{\boldsymbol{\beta}L}^{\text{MAR}}}^{\text{MAR}} \mathbf{W}'$ e $\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}(M_L)}^{\text{MCAR}} = \mathbf{W} \hat{\mathbf{V}}_{\hat{\boldsymbol{\beta}L}^{\text{MCAR}}}^{\text{MCAR}} \mathbf{W}'$. Usando (2.8) e o método *delta*, obtém-se as estimativas das matrizes de covariâncias assintóticas de $\hat{\boldsymbol{\theta}}(M_L)$ sob os mecanismos MAR, $\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}(M_L)}^{\text{MAR}} = \mathbf{B} \hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}(M_L)}^{\text{MAR}} \mathbf{B}'$, e MCAR, $\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}(M_L)}^{\text{MCAR}} = \mathbf{B} \hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}(M_L)}^{\text{MCAR}} \mathbf{B}'$. As EMV das funções lineares $\mathbf{A}\boldsymbol{\theta}$ sob M_L são obtidas a partir de $\mathbf{X}\hat{\boldsymbol{\beta}}$ e, por intermédio do

método *delta*, as estimativas das suas respectivas matrizes de covariâncias assintóticas sob os mecanismos MAR e MCAR são obtidas por $\hat{\mathbf{V}}_{\mathbf{A}\hat{\boldsymbol{\theta}}(M_L)}^{\text{MAR}} = \mathbf{X}\hat{\mathbf{V}}_{\hat{\boldsymbol{\beta}}_L}^{\text{MAR}}\mathbf{X}'$ e $\hat{\mathbf{V}}_{\mathbf{A}\hat{\boldsymbol{\theta}}(M_L)}^{\text{MCAR}} = \mathbf{X}\hat{\mathbf{V}}_{\hat{\boldsymbol{\beta}}_L}^{\text{MCAR}}\mathbf{X}'$.

Outra estrutura muito utilizada é a log-linear, que pode ser expressa na forma

$$M_{LL} : \ln(\boldsymbol{\theta}_s) = \mathbf{1}_R \nu_s + \mathbf{X}_s \boldsymbol{\beta}, \quad s = 1, \dots, S$$

ou, de uma forma condensada, por

$$M_{LL} : \ln(\boldsymbol{\theta}) = [\mathbf{I}_S \otimes \mathbf{1}_R] \boldsymbol{\nu} + \mathbf{X} \boldsymbol{\beta}, \quad (2.65)$$

em que $\boldsymbol{\nu} = (\nu_1, \dots, \nu_S)'$ é um vetor de S componentes associadas às restrições naturais, tal que $\boldsymbol{\nu} = -\ln[(\mathbf{I}_S \otimes \mathbf{1}'_R) \exp(\mathbf{X} \boldsymbol{\beta})]$, $\exp(\mathbf{X} \boldsymbol{\beta})$ denota o operador exponencial vetorial que consiste em se aplicar o exponencial a cada um dos elementos de $\mathbf{X} \boldsymbol{\beta}$ e $\nu_s = -\ln[\mathbf{1}'_R \exp(\mathbf{X}_s \boldsymbol{\beta})]$, $s = 1, \dots, S$; o vetor $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$, de dimensão $p \times 1$, engloba os $p \leq S(R-1)$ parâmetros desconhecidos e a matriz $\mathbf{X} = (\mathbf{X}'_1, \dots, \mathbf{X}'_S)'$, de dimensão $SR \times p$, é tal que cada submatriz \mathbf{X}_s , de dimensão $R \times p$, tem suas colunas linearmente independentes do vetor $\mathbf{1}_R$ definidor da s -ésima restrição natural, $\mathbf{1}'_R \boldsymbol{\theta}_s = 1$, *i.e.*, $r(\mathbf{1}_R, \mathbf{X}_s) = 1 + r(\mathbf{X}_s)$, $s = 1, \dots, S$ e $r(\mathbf{I}_S \otimes \mathbf{1}_R, \mathbf{X}) = S + p$. Reescrevendo (2.65), pode-se obter $\boldsymbol{\theta}$ a partir de $\boldsymbol{\beta}$ por

$$\boldsymbol{\theta}(\boldsymbol{\beta}) = \mathbf{D}_{\boldsymbol{\psi}}^{-1} \exp(\mathbf{X} \boldsymbol{\beta}), \quad (2.66)$$

em que $\boldsymbol{\psi} = [\mathbf{I}_S \otimes (\mathbf{1}_R \mathbf{1}'_R)] \exp(\mathbf{X} \boldsymbol{\beta}) = (\psi'_s, s = 1, \dots, S)'$, $\boldsymbol{\theta}(\boldsymbol{\beta}) = (\theta'_s(\boldsymbol{\beta}), s = 1, \dots, S)'$, $\theta_s(\boldsymbol{\beta}) = \mathbf{D}_{\psi_s}^{-1} \exp(\mathbf{X}_s \boldsymbol{\beta})$ e $\psi_s = (\mathbf{1}_R \mathbf{1}'_R) \exp(\mathbf{X}_s \boldsymbol{\beta})$.

Pode-se também considerar uma classe mais ampla de modelos log-lineares, expressável por

$$M_{LL} : \mathbf{A} \ln(\boldsymbol{\theta}) = \mathbf{X}_L \boldsymbol{\beta}, \quad (2.67)$$

em que a matriz \mathbf{A} , de dimensão $u \times SR$, tem posto $r(\mathbf{A}) = u \leq S(R-1)$ e suas linhas são ortogonais às colunas da matriz $\mathbf{I}_S \otimes \mathbf{1}_R$ definidora das restrições naturais, *i.e.*, $\mathbf{A}(\mathbf{I}_S \otimes \mathbf{1}_R) = \mathbf{0}_{u,S}$. Neste contexto, por exemplo, a matriz $\mathbf{A} = \mathbf{I}_S \otimes [\mathbf{I}_{R-1}, -\mathbf{1}_{R-1}]$ indica a utilização dos logitos de referência (com relação à categoria R). Se $u = S(R-1)$, a matriz \mathbf{X}_L , de dimensão $S(R-1) \times p$, tem as seguintes relações com \mathbf{X}

$$\mathbf{X}_L = \mathbf{A} \mathbf{X} \quad \text{e} \quad \mathbf{X} = \mathbf{A}' (\mathbf{A} \mathbf{A}')^{-1} \mathbf{X}_L. \quad (2.68)$$

Se $u < S(R-1)$, necessita-se uma matriz \mathbf{A}_0 , com dimensão $(S[R-1] - u) \times SR$, base do complemento ortogonal de $(\mathbf{A}', \mathbf{I}_S \otimes \mathbf{1}_R)'$, tal que o modelo a ser ajustado

$$M_{LL} : \begin{pmatrix} \mathbf{A} \\ \mathbf{A}_0 \end{pmatrix} \ln(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{X}_L \boldsymbol{\beta} \\ \boldsymbol{\beta}_0 \end{pmatrix} \quad (2.69)$$

pode ser escrito na forma (2.65) por

$$M_{LL} : \ln(\boldsymbol{\theta}) = [\mathbf{I}_S \otimes \mathbf{1}_R] \boldsymbol{\nu} + \begin{pmatrix} \mathbf{A}' (\mathbf{A}\mathbf{A}')^{-1} \mathbf{X}_L & , & \mathbf{A}'_0 (\mathbf{A}_0\mathbf{A}'_0)^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\beta}_0 \end{pmatrix}. \quad (2.70)$$

Por simplicidade, no desenvolvimento a seguir, suprime-se o parâmetro $\boldsymbol{\beta}_0$, mencionando-se apenas $\boldsymbol{\beta}$. No entanto, sempre que $u < S(R-1)$, $\boldsymbol{\beta}_0$ também é utilizado e considera-se $\mathbf{X} = (\mathbf{A}' (\mathbf{A}\mathbf{A}')^{-1} \mathbf{X}_L, \mathbf{A}'_0 (\mathbf{A}_0\mathbf{A}'_0)^{-1})$.

As formulações (2.65) e (2.67) em equações livres são equivalentes, respectivamente, às formulações em termos de restrições

$$\mathbf{U} \ln(\boldsymbol{\theta}) = \mathbf{0}_{S(R-1)-p}, \quad (2.71)$$

$$\mathbf{U}_L \mathbf{A} \ln(\boldsymbol{\theta}) = \mathbf{0}_{u-p}, \quad (2.72)$$

em que a matriz \mathbf{U} (\mathbf{U}_L), de dimensão $[S\{R-1\}-p] \times SR$ ($[u-p] \times u$), define as $S[R-1]-p$ ($u-p$) restrições, possui posto máximo e suas linhas são ortogonais às colunas de $[\mathbf{I}_S \otimes \mathbf{1}_R, \mathbf{X}]$ (\mathbf{X}_L), ou seja, $\mathbf{U}[\mathbf{I}_S \otimes \mathbf{1}_R, \mathbf{X}] = \mathbf{0}_{(SR-p),p}$ ($\mathbf{U}_L\mathbf{X}_L = \mathbf{0}_{(u-p),p}$).

Diferenciando $L_1(\boldsymbol{\theta}(\boldsymbol{\beta}) | \mathbf{N})$ em relação a $\boldsymbol{\beta}$ obtém-se o vetor *score*

$$\mathbf{S}_{1LL}(\boldsymbol{\beta}) = \sum_{s=1}^S \mathbf{X}'_s \left\{ \mathbf{N}_{s1} + \sum_{t=2}^{T_s} \left[\mathbf{D}_{\boldsymbol{\theta}_s(\boldsymbol{\beta})} \mathbf{Z}_{st} \mathbf{D}_{\mathbf{Z}'_{st} \boldsymbol{\theta}_s(\boldsymbol{\beta})}^{-1} \mathbf{N}_{st} \right] - n_{s++} \boldsymbol{\theta}_s(\boldsymbol{\beta}) \right\}. \quad (2.73)$$

A diferenciação adicional do gradiente de $L_1(\boldsymbol{\theta}(\boldsymbol{\beta}) | \mathbf{N})$ em relação a $\boldsymbol{\beta}'$ permite a obtenção da matriz hessiana

$$\mathbf{H}_{1LL}(\boldsymbol{\beta}) = \sum_{s=1}^S \mathbf{X}'_s \left\{ -n_{s++} \mathbf{I}_R + \sum_{t=2}^{T_s} \left[\mathbf{D}_{\mathbf{u}_{st}^I} - \mathbf{D}_{\mathbf{u}_{st}^{II}} \mathbf{Z}_{st} \mathbf{Z}'_{st} \right] \right\} \left\{ \mathbf{D}_{\boldsymbol{\theta}_s(\boldsymbol{\beta})} - \boldsymbol{\theta}_s(\boldsymbol{\beta}) [\boldsymbol{\theta}_s(\boldsymbol{\beta})]' \right\} \mathbf{X}_s, \quad (2.74)$$

em que $\mathbf{u}_{st}^I = \mathbf{Z}_{st} \mathbf{D}_{\mathbf{Z}'_{st} \boldsymbol{\theta}_s(\boldsymbol{\beta})}^{-1} \mathbf{N}_{st}$ e $\mathbf{u}_{st}^{II} = \mathbf{D}_{\boldsymbol{\theta}_s(\boldsymbol{\beta})} \mathbf{Z}_{st} \mathbf{D}_{\mathbf{Z}'_{st} \boldsymbol{\theta}_s(\boldsymbol{\beta})}^{-2} \mathbf{N}_{st}$, $s = 1, \dots, S$, $t = 1, \dots, T_s$. Apresentam-se alguns detalhes da obtenção de (2.73) e (2.74) no Apêndice A.5. A matriz de informação de Fisher sob o mecanismo MAR é expressa por

$$\mathcal{I}_{1LL}(\boldsymbol{\beta}, \{\boldsymbol{\alpha}_{st}^{\text{MAR}}\}) = \sum_{s=1}^S \mathbf{X}'_s \left\{ n_{s++} \mathbf{I}_R - \sum_{t=2}^{T_s} \left[\mathbf{D}_{\mathbf{v}_{st}^I} - \mathbf{D}_{\mathbf{v}_{st}^{II}} \mathbf{Z}_{st} \mathbf{Z}'_{st} \right] \right\} \left\{ \mathbf{D}_{\boldsymbol{\theta}_s(\boldsymbol{\beta})} - \boldsymbol{\theta}_s(\boldsymbol{\beta}) [\boldsymbol{\theta}_s(\boldsymbol{\beta})]' \right\} \mathbf{X}_s, \quad (2.75)$$

em que $\mathbf{v}_{st}^I = n_{s++} \mathbf{Z}_{st} \boldsymbol{\alpha}_{st}^{\text{MAR}}$ e $\mathbf{v}_{st}^{II} = n_{s++} \mathbf{D}_{\boldsymbol{\theta}_s(\boldsymbol{\beta})} \mathbf{Z}_{st} \mathbf{D}_{\mathbf{Z}'_{st} \boldsymbol{\theta}_s(\boldsymbol{\beta})}^{-1} \boldsymbol{\alpha}_{st}^{\text{MAR}}$, $s = 1, \dots, S$, $t = 1, \dots, T_s$.

A matriz de informação de Fisher sob o mecanismo MCAR é representada por

$$\mathcal{I}_{1LL}(\boldsymbol{\beta}, \{\boldsymbol{\alpha}_{st}^{\text{MCAR}}\}) = \sum_{s=1}^S \mathbf{X}'_s \left\{ n_{s++} \mathbf{I}_R - \sum_{t=2}^{T_s} \left[\mathbf{D}_{\mathbf{w}_{st}^I} - \mathbf{D}_{\mathbf{w}_{st}^{II}} \mathbf{Z}_{st} \mathbf{Z}'_{st} \right] \right\} \left\{ \mathbf{D}_{\boldsymbol{\theta}_s(\boldsymbol{\beta})} - \boldsymbol{\theta}_s(\boldsymbol{\beta}) [\boldsymbol{\theta}_s(\boldsymbol{\beta})]' \right\} \mathbf{X}_s, \quad (2.76)$$

em que $\mathbf{w}_{st}^I = n_{s++}\boldsymbol{\alpha}_{t(s)}\mathbf{1}_R$ e $\mathbf{w}_{st}^{II} = n_{s++}\boldsymbol{\alpha}_{t(s)}\mathbf{D}_{\boldsymbol{\theta}_s(\boldsymbol{\beta})}\mathbf{Z}_{st}\mathbf{D}_{\mathbf{Z}'_{st}\boldsymbol{\theta}_s(\boldsymbol{\beta})}^{-1}\mathbf{1}_{Rst}$, $s = 1, \dots, S$, $t = 1, \dots, T_s$.

O processo iterativo *scoring* de Fisher pode ser inicializado com a estimativa de MQG (caso particular do que será exposto na Seção 2.4.3)

$$\hat{\boldsymbol{\beta}}^{(0)} = \left[\mathbf{X}'_L \left(\mathbf{A}\mathbf{D}_{\hat{\boldsymbol{\theta}}^{-1}\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}}\mathbf{D}_{\hat{\boldsymbol{\theta}}^{-1}\mathbf{A}'} \right)^{-1} \mathbf{X}_L \right]^{-1} \mathbf{X}'_L \left(\mathbf{A}\mathbf{D}_{\hat{\boldsymbol{\theta}}^{-1}\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}}\mathbf{D}_{\hat{\boldsymbol{\theta}}^{-1}\mathbf{A}'} \right)^{-1} \mathbf{A} \ln(\hat{\boldsymbol{\theta}}), \quad (2.77)$$

se $u = S(R-1)$, ou por meio de expressão análoga proveniente da substituição de \mathbf{A} por $(\mathbf{A}', \mathbf{A}'_0)'$ e \mathbf{X}_L por

$$\begin{pmatrix} \mathbf{X}_L & \mathbf{0}_{u, S(R-1)-u} \\ \mathbf{0}_{S(R-1)-u, p} & \mathbf{I}_{S(R-1)-u} \end{pmatrix},$$

se $u < S(R-1)$, em que $\hat{\boldsymbol{\theta}}$ é a EMV irrestrita de $\boldsymbol{\theta}$ e $\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}}$ é uma estimativa proveniente de um estimador consistente da matriz de covariâncias assintótica sob o mecanismo MAR ($\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}}^{\text{MAR}}$) ou MCAR ($\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}}^{\text{MCAR}}$), ambas obtidas conforme sugestão apresentada na Seção 2.3.1.

Uma vez obtida a EMV $\hat{\boldsymbol{\beta}}$ de $\boldsymbol{\beta}$, as estimativas de suas matrizes de covariâncias assintóticas sob os mecanismos MAR e MCAR são, respectivamente, $\hat{\mathbf{V}}_{\hat{\boldsymbol{\beta}}LL}^{\text{MAR}} = \left[\mathcal{I}_{1LL} \left(\hat{\boldsymbol{\beta}}, \{ \hat{\boldsymbol{\alpha}}_{st}^{\text{MAR}} \} \right) \right]^{-1}$ e $\hat{\mathbf{V}}_{\hat{\boldsymbol{\beta}}LL}^{\text{MCAR}} = \left[\mathcal{I}_{1LL} \left(\hat{\boldsymbol{\beta}}, \{ \hat{\boldsymbol{\alpha}}_{st}^{\text{MCAR}} \} \right) \right]^{-1}$. Por meio de (2.66), obtém-se a EMV $\hat{\boldsymbol{\theta}}(M_{LL})$ de $\boldsymbol{\theta}$ sob M_{LL} e, utilizando o método *delta*, uma estimativa da matriz de covariâncias assintótica sob o mecanismo MAR

$$\begin{aligned} \hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}(M_{LL})}^{\text{MAR}} &= \frac{\widehat{\partial\boldsymbol{\theta}}}{\partial\boldsymbol{\beta}'} \hat{\mathbf{V}}_{\hat{\boldsymbol{\beta}}LL}^{\text{MAR}} \left(\frac{\widehat{\partial\boldsymbol{\theta}}}{\partial\boldsymbol{\beta}'} \right)' \\ &= \hat{\mathbf{V}}_{LL} \mathbf{X} \hat{\mathbf{V}}_{\hat{\boldsymbol{\beta}}LL}^{\text{MAR}} \mathbf{X}' \hat{\mathbf{V}}_{LL}, \end{aligned} \quad (2.78)$$

em que $\hat{\mathbf{V}}_{LL}$ é uma matriz diagonal em blocos com blocos iguais⁴ a $\mathbf{D}_{\boldsymbol{\theta}_s(\hat{\boldsymbol{\beta}})} - \boldsymbol{\theta}_s(\hat{\boldsymbol{\beta}}) \left[\boldsymbol{\theta}_s(\hat{\boldsymbol{\beta}}) \right]'$, $s = 1, \dots, S$. Da mesma forma, uma estimativa da matriz de covariâncias assintótica de $\hat{\boldsymbol{\theta}}(M_{LL})$ sob o mecanismo MCAR é $\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}(M_{LL})}^{\text{MCAR}} = \hat{\mathbf{V}}_{LL} \mathbf{X} \hat{\mathbf{V}}_{\hat{\boldsymbol{\beta}}LL}^{\text{MCAR}} \mathbf{X}' \hat{\mathbf{V}}_{LL}$. As EMV das funções log-lineares $\mathbf{A} \ln(\boldsymbol{\theta})$ sob M_{LL} são obtidas a partir de $\mathbf{X}_L \hat{\boldsymbol{\beta}}$ e, por intermédio do método *delta*, as estimativas das suas respectivas matrizes de covariâncias assintóticas sob os mecanismos MAR e MCAR são obtidas por $\hat{\mathbf{V}}_{\mathbf{A} \ln(\hat{\boldsymbol{\theta}}(M_{LL}))}^{\text{MAR}} = \mathbf{X}_L \hat{\mathbf{V}}_{\hat{\boldsymbol{\beta}}LL}^{\text{MAR}} \mathbf{X}'_L$ e $\hat{\mathbf{V}}_{\mathbf{A} \ln(\hat{\boldsymbol{\theta}}(M_{LL}))}^{\text{MCAR}} = \mathbf{X}_L \hat{\mathbf{V}}_{\hat{\boldsymbol{\beta}}LL}^{\text{MCAR}} \mathbf{X}'_L$.

Seja \mathcal{M} um mecanismo de omissão mais restritivo que o MAR para as probabilidades condicionais de omissão e M (e.g., M_L ou M_{LL}) um modelo reduzido para o vetor $\boldsymbol{\theta}$. A

⁴Veja o cálculo de $\partial\boldsymbol{\theta}_s/\partial\boldsymbol{\beta}'$ em (A.8), Apêndice A.5.

estatística de teste do modelo conjunto (M, \mathcal{M}) condicional ao mecanismo MAR pelo critério da razão de verossimilhanças de Wilks é desdobrável na soma das correspondentes estatísticas de teste, separadamente, de M e de \mathcal{M} , isto é,

$$\begin{aligned} Q_V(M, \mathcal{M}|\text{MAR}) &= -2 \ln \frac{L_1(\hat{\boldsymbol{\theta}}(M) | \mathbf{N}; M) L_2(\{\hat{\alpha}_{t(cs)}(\mathcal{M})\} | \mathbf{N}; \mathcal{M})}{L_1(\hat{\boldsymbol{\theta}} | \mathbf{N}) L_2(\{\hat{\alpha}_{t(cs)}\} | \mathbf{N}; \text{MAR})} \\ &= Q_V(M) + Q_V(\mathcal{M}|\text{MAR}) \end{aligned} \quad (2.79)$$

em que $\hat{\boldsymbol{\theta}} [\hat{\boldsymbol{\theta}}(M)]$ é a EMV de $\boldsymbol{\theta}$ sob o modelo saturado [sob M] e $\{\hat{\alpha}_{t(cs)}\} [\{\hat{\alpha}_{t(cs)}(\mathcal{M})\}]$ é a EMV de $\{\alpha_{t(cs)}\}$ sob o mecanismo MAR (\mathcal{M}). Williamson e Haber (1994) notaram que este particionamento de Q_V mostra que, por este critério, a comparação de qualquer par de modelos, para as probabilidades de categorização e as probabilidades condicionais de omissão, não depende da estrutura tão ou mais reduzida que se imponha, respectivamente, ao mecanismo de omissão e às probabilidades de categorização. Se o parâmetro de interesse for apenas $\boldsymbol{\theta}$, a estatística de razão de verossimilhanças do teste de ajuste do modelo M é expressa por

$$Q_V(M|\mathcal{M}) = Q_V(M) = -2 \ln \frac{L_1(\hat{\boldsymbol{\theta}}(M) | \mathbf{N})}{L_1(\hat{\boldsymbol{\theta}} | \mathbf{N})} = -2 \sum_{s=1}^S \mathbf{N}'_s \left\{ \ln [\mathbf{Z}'_s \hat{\boldsymbol{\theta}}_s(M)] - \ln [\mathbf{Z}'_s \hat{\boldsymbol{\theta}}_s] \right\}, \quad (2.80)$$

sendo independente do mecanismo \mathcal{M} mais restritivo que o MAR (*e.g.*, MCAR) que se esteja assumindo.

As estatísticas de Pearson e Neyman de teste de (M, MCAR) condicionalmente ao mecanismo MAR

$$\begin{aligned} Q_P(M, \text{MCAR}|\text{MAR}) &= \sum_{s=1}^S \sum_{t=1}^{T_s} \sum_{c=1}^{R_{st}} \frac{\left(n_{stc} - n_{st+} \mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s(M) \right)^2}{n_{st+} \mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s(M)} \\ &= \sum_{s=1}^S \left(\mathbf{p}_s - \mathbf{Z}'_s \hat{\boldsymbol{\theta}}_s(M) \right)' \left[\mathbf{D}_{\mathbf{N}_{s+}} \mathbf{D}_{\mathbf{Z}'_s \hat{\boldsymbol{\theta}}_s(M)}^{-1} \right] \left(\mathbf{p}_s - \mathbf{Z}'_s \hat{\boldsymbol{\theta}}_s(M) \right), \end{aligned} \quad (2.81)$$

$$\begin{aligned} Q_N(M, \text{MCAR}|\text{MAR}) &= \sum_{s=1}^S \sum_{t=1}^{T_s} \sum_{c=1}^{R_{st}} \frac{\left(n_{stc} - n_{st+} \mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s(M) \right)^2}{n_{stc}} \\ &= \sum_{s=1}^S \left(\mathbf{p}_s - \mathbf{Z}'_s \hat{\boldsymbol{\theta}}_s(M) \right)' \left[\mathbf{D}_{\mathbf{N}_{s+}} \mathbf{D}_{\mathbf{p}_s}^{-1} \right] \left(\mathbf{p}_s - \mathbf{Z}'_s \hat{\boldsymbol{\theta}}_s(M) \right), \end{aligned} \quad (2.82)$$

de teste do modelo M condicionalmente ao mecanismo MAR

$$\begin{aligned}
 Q_P(M|\text{MAR}) &= \sum_{s=1}^S \sum_{t=1}^{T_s} \sum_{c=1}^{R_{st}} \frac{\left(n_{stc} - n_{s++} \mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s(M) \hat{\alpha}_{t(cs)} \right)^2}{n_{s++} \mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s(M) \hat{\alpha}_{t(cs)}} \\
 &= \sum_{s=1}^S \sum_{t=1}^{T_s} \sum_{c=1}^{R_{st}} \frac{n_{stc}}{\mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s} \frac{\left(\mathbf{z}'_{stc} [\hat{\boldsymbol{\theta}}_s - \hat{\boldsymbol{\theta}}_s(M)] \right)^2}{\mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s(M)} \\
 &= \sum_{s=1}^S \left(\mathbf{z}'_s [\hat{\boldsymbol{\theta}}_s - \hat{\boldsymbol{\theta}}_s(M)] \right)' \left[\mathbf{D}_{\mathbf{N}_s} \mathbf{D}_{\mathbf{z}'_s \hat{\boldsymbol{\theta}}_s}^{-1} \mathbf{D}_{\mathbf{z}'_s \hat{\boldsymbol{\theta}}_s(M)}^{-1} \right] \left(\mathbf{z}'_s [\hat{\boldsymbol{\theta}}_s - \hat{\boldsymbol{\theta}}_s(M)] \right), \quad (2.83)
 \end{aligned}$$

$$\begin{aligned}
 Q_N(M|\text{MAR}) &= \sum_{s=1}^S \sum_{t=1}^{T_s} \sum_{c=1}^{R_{st}} \frac{\left(n_{stc} - n_{s++} \mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s(M) \hat{\alpha}_{t(cs)} \right)^2}{n_{stc}} \\
 &= \sum_{s=1}^S \sum_{t=1}^{T_s} \sum_{c=1}^{R_{st}} n_{stc} \left(1 - \frac{\mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s(M)}{\mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s} \right)^2 \\
 &= \sum_{s=1}^S \left(\mathbf{1}_{R+l_s} - \mathbf{D}_{\mathbf{z}'_s \hat{\boldsymbol{\theta}}_s}^{-1} \mathbf{z}'_s \hat{\boldsymbol{\theta}}_s(M) \right)' \mathbf{D}_{\mathbf{N}_s} \left(\mathbf{1}_{R+l_s} - \mathbf{D}_{\mathbf{z}'_s \hat{\boldsymbol{\theta}}_s}^{-1} \mathbf{z}'_s \hat{\boldsymbol{\theta}}_s(M) \right), \quad (2.84)
 \end{aligned}$$

e de teste do modelo M condicionalmente ao mecanismo MCAR

$$\begin{aligned}
 Q_P(M|\text{MCAR}) &= \sum_{s=1}^S \sum_{t=1}^{T_s} \sum_{c=1}^{R_{st}} \frac{\left(n_{st+\mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s} - n_{st+\mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s(M)} \right)^2}{n_{st+\mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s(M)}} \\
 &= \sum_{s=1}^S \sum_{t=1}^{T_s} \sum_{c=1}^{R_{st}} \frac{n_{st+} \left(\mathbf{z}'_{stc} [\hat{\boldsymbol{\theta}}_s - \hat{\boldsymbol{\theta}}_s(M)] \right)^2}{\mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s(M)} \\
 &= \sum_{s=1}^S \left(\mathbf{z}'_s [\hat{\boldsymbol{\theta}}_s - \hat{\boldsymbol{\theta}}_s(M)] \right)' \left[\mathbf{D}_{\mathbf{N}_{s+}} \mathbf{D}_{\mathbf{z}'_s \hat{\boldsymbol{\theta}}_s(M)}^{-1} \right] \left(\mathbf{z}'_s [\hat{\boldsymbol{\theta}}_s - \hat{\boldsymbol{\theta}}_s(M)] \right), \quad (2.85)
 \end{aligned}$$

$$\begin{aligned}
 Q_N(M|\text{MCAR}) &= \sum_{s=1}^S \sum_{t=1}^{T_s} \sum_{c=1}^{R_{st}} \frac{\left(n_{st+\mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s} - n_{st+\mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s(M)} \right)^2}{n_{st+\mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s}} \\
 &= \sum_{s=1}^S \sum_{t=1}^{T_s} \sum_{c=1}^{R_{st}} \frac{n_{st+} \left(\mathbf{z}'_{stc} [\hat{\boldsymbol{\theta}}_s - \hat{\boldsymbol{\theta}}_s(M)] \right)^2}{\mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s} \\
 &= \sum_{s=1}^S \left(\mathbf{z}'_s [\hat{\boldsymbol{\theta}}_s - \hat{\boldsymbol{\theta}}_s(M)] \right)' \left[\mathbf{D}_{\mathbf{N}_{s+}} \mathbf{D}_{\mathbf{z}'_s \hat{\boldsymbol{\theta}}_s}^{-1} \right] \left(\mathbf{z}'_s [\hat{\boldsymbol{\theta}}_s - \hat{\boldsymbol{\theta}}_s(M)] \right), \quad (2.86)
 \end{aligned}$$

em que $\mathbf{N}_{s+} = (n_{st+} \otimes \mathbf{1}'_{R_{st}}, t = 1, \dots, T_s)'$ e $\hat{\alpha}_{t(cs)} = n_{stc} / (n_{s++} \mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s)$, evidenciam que já não se tem os vantajosos resultados obtidos com a estatística de razão de verossimilhanças.

As estatísticas de Wald de teste de, respectivamente, M_L e M_{LL} condicionalmente ao mecanismo de omissão \mathcal{M} (MAR ou MCAR) são

$$Q_W(M_L|\mathcal{M}) = (\mathbf{U}\mathbf{A}\hat{\boldsymbol{\theta}})' [\mathbf{U}\mathbf{A}\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}}^{\mathcal{M}}\mathbf{A}'\mathbf{U}']^{-1} \mathbf{U}\mathbf{A}\hat{\boldsymbol{\theta}}, \quad (2.87)$$

$$Q_W(M_{LL}|\mathcal{M}) = (\mathbf{U}_L \mathbf{A} \ln(\hat{\boldsymbol{\theta}}))' [\mathbf{U}\mathbf{A}\mathbf{D}_{\hat{\boldsymbol{\theta}}}^{-1}\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}}^{\mathcal{M}}\mathbf{D}_{\hat{\boldsymbol{\theta}}}^{-1}\mathbf{A}'\mathbf{U}']^{-1} \mathbf{U}_L \mathbf{A} \ln(\hat{\boldsymbol{\theta}}). \quad (2.88)$$

Assintoticamente, sob o modelo M e o mecanismo MAR

$$Q_V(M) \stackrel{a}{\approx} Q_P(M|\text{MAR}) \stackrel{a}{\approx} Q_N(M|\text{MAR}) \stackrel{a}{\approx} Q_W(M|\text{MAR}) \xrightarrow{a} \chi^2_{(u-p)}$$

e, adicionalmente sob o mecanismo MCAR,

$$Q_P(M|\text{MCAR}) \stackrel{a}{\approx} Q_N(M|\text{MCAR}) \stackrel{a}{\approx} Q_W(M|\text{MCAR}) \xrightarrow{a} \chi^2_{(u-p)},$$

$$Q_V(M, \text{MCAR}|\text{MAR}) \stackrel{a}{\approx} Q_P(M, \text{MCAR}|\text{MAR}) \stackrel{a}{\approx} Q_N(M, \text{MCAR}|\text{MAR}) \xrightarrow{a} \chi^2_{(u-p+g)},$$

em que $g = S + \sum_{s=1}^S (l_s - T_s)$.

Existindo interesse em se efetuar uma redução da dimensão de $\boldsymbol{\beta}$ através de uma hipótese do tipo

$$H : \mathbf{C}\boldsymbol{\beta} = \mathbf{C}_0, \quad (2.89)$$

em que \mathbf{C}_0 é um vetor, de dimensão $c \times 1$, com constantes conhecidas (geralmente, $\mathbf{C}_0 = \mathbf{0}_c$) e a matriz \mathbf{C} , de dimensão $c \times p$, tem posto máximo c ($\leq p$) e suas linhas definem os contrastes de interesse, pode-se recorrer à estatística de Wald

$$Q_W(H|M, \mathcal{M}) = (\mathbf{C}\hat{\boldsymbol{\beta}}(M) - \mathbf{C}_0)' [\mathbf{C}\hat{\mathbf{V}}_{\hat{\boldsymbol{\beta}}(M)}^{\mathcal{M}}\mathbf{C}']^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}}(M) - \mathbf{C}_0), \quad (2.90)$$

que tem distribuição nula assintótica $\chi^2_{(c)}$.

2.4.3 Estimação por mínimos quadrados generalizados para modelos estruturais não saturados sob os mecanismos MAR, MCAR e MNAR

Para efeitos de análise de modelos funcionais lineares para θ sob a validade do mecanismo MCAR, Koch, Imrey e Reinfurt (1972) propõem a aplicação de MQG numa segunda fase à estimativa de MQG, $\tilde{\theta}$, obtida na Seção 2.3.2, com o auxílio de uma estimativa da matriz de covariâncias aproximada $\tilde{V}_{\tilde{\theta}}$. No espírito da regressão funcional assintótica para dados completos conforme Imrey, Koch, Stokes *et al.* (1981, 1982) e utilizada em contextos diferentes em Koch, Singer e Amara (1985) e Ho e Singer (2001), Paulino (1991) sugeriu a aplicação de MQG num segundo estágio à EMV $\hat{\theta}$ sob qualquer mecanismo de omissão, com estimativa da matriz de covariâncias aproximada $\hat{V}_{\hat{\theta}}$. Com esta metodologia híbrida se espera facilitar o traçado de certas inferências sobre θ , principalmente no contexto de modelos não-ignoráveis para o mecanismo de omissão (MAR e MNAR), através de procedimentos que continuem a usufruir de boas propriedades em grandes amostras. Sendo assim, pode-se utilizar a EMV $\hat{\theta}$ de θ discutida na Seção 2.3.1, com o auxílio de uma estimativa proveniente de um estimador consistente da matriz de covariâncias aproximada $\hat{V}_{\hat{\theta}}^{\text{MAR}}$, sob o mecanismo MAR, ou $\hat{V}_{\hat{\theta}}^{\text{MCAR}}$, sob o mecanismo MCAR. Devido à imensa variedade de estruturas MNAR, não se apresentou procedimentos de estimação de θ sob estes mecanismos. Entretanto, obtendo-se a EMV irrestrita $\hat{\theta}$ de θ sob alguma estrutura MNAR e respectiva estimativa de sua matriz de covariâncias assintótica, também pode-se aplicar esta abordagem híbrida.

Por exemplo, considerando a estrutura MNAR1 da Seção 1.4, $\hat{\theta}$ pode ser obtido maximizando

$$\begin{aligned}
 L(\theta, \alpha | \mathbf{N}) &\propto \prod_{i=1}^2 \prod_{j=1}^2 (\theta_{ij} \lambda_{1(ij)})^{n_{1ij}} \times \prod_{i=1}^2 (\theta_{i1} \lambda_{2(i1)} + \theta_{i2} \lambda_{2(i2)})^{n_{2i+}} \times \\
 &\quad \prod_{j=1}^2 (\theta_{1j} \lambda_{3(1j)} + \theta_{2j} \lambda_{3(2j)})^{n_{3+j}} \times \left(\sum_{i=1}^2 \sum_{j=1}^2 \theta_{ij} \lambda_{4(ij)} \right)^{n_{4++}} \quad (2.91) \\
 &= \prod_{i=1}^2 \prod_{j=1}^2 (\theta_{ij} \psi_{1(ij)} \psi_{21(ij)})^{n_{1ij}} \times \prod_{i=1}^2 \left(\sum_{j=1}^2 \theta_{ij} \psi_{1(ij)} (1 - \psi_{21(ij)}) \right)^{n_{2i+}} \times \\
 &\quad \prod_{j=1}^2 \left(\sum_{i=1}^2 \theta_{ij} (1 - \psi_{1(ij)}) \psi_{20(ij)} \right)^{n_{3+j}} \times \left(\sum_{i=1}^2 \sum_{j=1}^2 \theta_{ij} (1 - \psi_{1(ij)}) (1 - \psi_{20(ij)}) \right)^{n_{4++}}
 \end{aligned}$$

$$\begin{aligned}
&= \prod_{i=1}^2 \prod_{j=1}^2 \left(\theta_{ij} \frac{e^{\alpha_{10} + \alpha_1(i-1) + \alpha_2(j-1)}}{1 + e^{\alpha_{10} + \alpha_1(i-1) + \alpha_2(j-1)}} \frac{e^{\alpha_{20} + \alpha_1(i-1) + \alpha_2(j-1)}}{1 + e^{\alpha_{20} + \alpha_1(i-1) + \alpha_2(j-1)}} \right)^{n_{1ij}} \times \\
&\prod_{i=1}^2 \left(\sum_{j=1}^2 \theta_{ij} \frac{e^{\alpha_{10} + \alpha_1(i-1) + \alpha_2(j-1)}}{1 + e^{\alpha_{10} + \alpha_1(i-1) + \alpha_2(j-1)}} \frac{1}{1 + e^{\alpha_{20} + \alpha_1(i-1) + \alpha_2(j-1)}} \right)^{n_{2i+}} \times \\
&\prod_{j=1}^2 \left(\sum_{i=1}^2 \theta_{ij} \frac{1}{1 + e^{\alpha_{10} + \alpha_1(i-1) + \alpha_2(j-1)}} \frac{e^{\alpha_{30} + \alpha_1(i-1) + \alpha_2(j-1)}}{1 + e^{\alpha_{30} + \alpha_1(i-1) + \alpha_2(j-1)}} \right)^{n_{3+j}} \times \\
&\left(\sum_{i=1}^2 \sum_{j=1}^2 \theta_{ij} \frac{1}{1 + e^{\alpha_{10} + \alpha_1(i-1) + \alpha_2(j-1)}} \frac{1}{1 + e^{\alpha_{30} + \alpha_1(i-1) + \alpha_2(j-1)}} \right)^{n_{4++}}
\end{aligned}$$

em relação a $(\boldsymbol{\theta}, \boldsymbol{\alpha})$, em que $\boldsymbol{\alpha} = (\alpha_{10}, \alpha_{20}, \alpha_{30}, \alpha_1, \alpha_2)'$. Uma estimativa da matriz de covariâncias aproximada de $\hat{\boldsymbol{\theta}}$ pode ser extraída da respectiva componente da matriz de covariâncias aproximada de $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}})$, obtida por meio da inversão seja da matriz de informação observada ou da matriz de informação de Fisher.

Por simplicidade, nesta seção $\tilde{\boldsymbol{\theta}}$ representa qualquer estimador consistente de $\boldsymbol{\theta}$ que reflita todos os dados disponíveis, como por exemplo, o estimador de MQG sob o mecanismo MCAR ou o EMV sob algum mecanismo de omissão \mathcal{M} . Da mesma forma, $\tilde{\mathbf{V}}_{\tilde{\boldsymbol{\theta}}}$ denota uma estimativa proveniente de um estimador consistente da matriz de covariâncias desse estimador sob o mecanismo de omissão \mathcal{M} . Adicionalmente, supõe-se que

$$\tilde{\boldsymbol{\theta}} \stackrel{a}{\sim} N_{SR}(\boldsymbol{\theta}, \tilde{\mathbf{V}}_{\tilde{\boldsymbol{\theta}}}). \quad (2.92)$$

Considera-se o modelo funcional linear

$$M_F : \mathbf{F} \equiv \mathbf{F}(\boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\beta}, \quad (2.93)$$

em que o vetor $\mathbf{F}(\boldsymbol{\theta}) = (F_i(\boldsymbol{\theta}), i = 1, \dots, u)'$, de dimensão $u \times 1$, define as $u \leq S(R - 1)$ funções, e é tal que

$$\mathbf{G} \equiv \mathbf{G}(\boldsymbol{\theta}) = \frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}'} \quad \text{e} \quad \frac{\partial^2 \mathbf{F}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$$

existem e são contínuas num conjunto aberto contendo $\boldsymbol{\theta}$; a matriz \mathbf{X} , de dimensão $u \times p$, com posto $r(\mathbf{X}) = p \leq u$, especifica o modelo e o vetor $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$, de dimensão $p \times 1$, contém os p parâmetros desconhecidos. A expressão (2.93) é a formulação em equações livres, que equivale à formulação em termos de restrições

$$M_F : \mathbf{U} \mathbf{F}(\boldsymbol{\theta}) = \mathbf{0}_{u-p}, \quad (2.94)$$

em que a matriz \mathbf{U} , de dimensão $(u-p) \times u$, contém as $u-p$ restrições, possui posto máximo e suas linhas são ortogonais às colunas de \mathbf{X} , ou seja, $\mathbf{UX} = \mathbf{0}_{(u-p),p}$.

Sob (2.92) e condições que assegurem a aplicabilidade do método *delta*, tem-se que

$$\tilde{\mathbf{F}} \equiv \mathbf{F}(\tilde{\boldsymbol{\theta}}) \stackrel{a}{\sim} N_u(\mathbf{F}, \tilde{\mathbf{V}}_{\tilde{\mathbf{F}}}), \quad (2.95)$$

em que

$$\tilde{\mathbf{V}}_{\tilde{\mathbf{F}}} = \tilde{\mathbf{G}}\tilde{\mathbf{V}}_{\tilde{\boldsymbol{\theta}}}\tilde{\mathbf{G}}', \quad \text{com } \tilde{\mathbf{G}} \equiv \mathbf{G}(\tilde{\boldsymbol{\theta}}), \quad (2.96)$$

se supõe não singular.

O estimador de MQG de $\boldsymbol{\beta}$ de (2.93) é expresso por

$$\tilde{\boldsymbol{\beta}} = \left(\mathbf{X}'\tilde{\mathbf{V}}_{\tilde{\mathbf{F}}}^{-1}\mathbf{X} \right)^{-1} \mathbf{X}'\tilde{\mathbf{V}}_{\tilde{\mathbf{F}}}^{-1}\tilde{\mathbf{F}} \quad (2.97)$$

e uma estimativa de sua matriz de covariâncias assintótica pode ser obtida por

$$\tilde{\mathbf{V}}_{\tilde{\boldsymbol{\beta}}} = \left(\mathbf{X}'\tilde{\mathbf{V}}_{\tilde{\mathbf{F}}}^{-1}\mathbf{X} \right)^{-1}. \quad (2.98)$$

O estimador de MQG das funções \mathbf{F} sob M_F é obtido a partir de $\mathbf{X}\tilde{\boldsymbol{\beta}}$ e, por intermédio do método *delta*, uma estimativa de sua matriz de covariâncias assintótica é obtida por $\tilde{\mathbf{V}}_{\tilde{\mathbf{F}}(M_F)} = \mathbf{X}\tilde{\mathbf{V}}_{\tilde{\boldsymbol{\beta}}}\mathbf{X}'$.

O teste de ajuste do modelo M_F condicional ao mecanismo de omissão \mathcal{M} que se está considerando pode ser realizado com a estatística de Wald

$$Q_W(M_F|\mathcal{M}) = \left(\mathbf{U}\tilde{\mathbf{F}} \right)' \left[\mathbf{U}\tilde{\mathbf{V}}_{\tilde{\mathbf{F}}}\mathbf{U}' \right]^{-1} \mathbf{U}\tilde{\mathbf{F}}, \quad (2.99)$$

que tem distribuição nula assintótica $\chi_{(u-p)}^2$.

Existindo interesse em se efetuar uma redução da dimensão de $\boldsymbol{\beta}$ através de uma hipótese do tipo

$$H : \mathbf{C}\boldsymbol{\beta} = \mathbf{C}_0, \quad (2.100)$$

em que \mathbf{C}_0 é um vetor, de dimensão $c \times 1$, com constantes conhecidas (geralmente, $\mathbf{C}_0 = \mathbf{0}_c$) e a matriz \mathbf{C} , de dimensão $c \times p$, tem posto máximo c ($\leq p$) e suas linhas definem os contrastes de interesse, pode-se recorrer à estatística de Wald

$$Q_W(H|M_F, \mathcal{M}) = \left(\mathbf{C}\tilde{\boldsymbol{\beta}} - \mathbf{C}_0 \right)' \left[\mathbf{C}\tilde{\mathbf{V}}_{\tilde{\mathbf{F}}}\mathbf{C}' \right]^{-1} \left(\mathbf{C}\tilde{\boldsymbol{\beta}} - \mathbf{C}_0 \right), \quad (2.101)$$

que tem distribuição nula assintótica $\chi_{(c)}^2$.

Em muitos casos o vetor de funções $\mathbf{F}(\boldsymbol{\theta})$ pode ser expresso como uma composição de funções lineares, logarítmicas, exponenciais e adição de constantes. Alguns exemplos de funções e matrizes de primeiras derivadas associadas são

$$\begin{aligned}\mathbf{F}(\boldsymbol{\theta}) &= \mathbf{A}_1 \boldsymbol{\theta} \implies \mathbf{G}(\boldsymbol{\theta}) = \mathbf{A}_1, \\ \mathbf{F}(\boldsymbol{\theta}) &= \ln(\boldsymbol{\theta}) \implies \mathbf{G}(\boldsymbol{\theta}) = \mathbf{D}_{\boldsymbol{\theta}}^{-1}, \\ \mathbf{F}(\boldsymbol{\theta}) &= \exp(\boldsymbol{\theta}) \implies \mathbf{G}(\boldsymbol{\theta}) = \mathbf{D}_{\exp(\boldsymbol{\theta})}, \\ \mathbf{F}(\boldsymbol{\theta}) &= \boldsymbol{\pi}_1 + \boldsymbol{\theta} \implies \mathbf{G}(\boldsymbol{\theta}) = \mathbf{I}_{SR}, \\ \mathbf{F}(\boldsymbol{\theta}) &= \mathbf{A}_1 \ln(\boldsymbol{\theta}) \implies \mathbf{G}(\boldsymbol{\theta}) = \mathbf{A}_1 \mathbf{D}_{\boldsymbol{\theta}}^{-1}, \\ \mathbf{F}(\boldsymbol{\theta}) &= \exp[\mathbf{A}_1 \ln(\boldsymbol{\pi}_1 + \boldsymbol{\theta})] \implies \mathbf{G}(\boldsymbol{\theta}) = \mathbf{D}_{\exp[\mathbf{A}_1 \ln(\boldsymbol{\pi}_1 + \boldsymbol{\theta})]} \mathbf{A}_1 \mathbf{D}_{\boldsymbol{\pi}_1 + \boldsymbol{\theta}}^{-1},\end{aligned}$$

em que \mathbf{A}_1 é uma matriz de dimensão $u \times SR$, com $u \leq S(R-1)$, e $\boldsymbol{\pi}_1$ é um vetor com constantes conhecidas, de dimensão $SR \times 1$. Note que as matrizes $\mathbf{G}(\boldsymbol{\theta})$ correspondentes às últimas duas funções são obtidas através da regra da cadeia, utilizando como base as primeiras derivadas das outras quatro funções. Por exemplo, para as medidas *kappa* de Cohen e *kappa* ponderada, em (2.44) e (2.46), a função $\mathbf{F}(\boldsymbol{\theta}) = \boldsymbol{\pi}_1 + \exp(\mathbf{A}_4 \ln\{\mathbf{A}_3 \exp[\mathbf{A}_2 \ln(\mathbf{A}_1 \boldsymbol{\theta})]\})$ em (2.45) tem matriz de primeiras derivadas dada por

$$\mathbf{G}(\boldsymbol{\theta}) = \mathbf{D}_{\exp(\mathbf{A}_4 \ln\{\mathbf{A}_3 \exp[\mathbf{A}_2 \ln(\mathbf{A}_1 \boldsymbol{\theta})]\})} \mathbf{A}_4 \mathbf{D}_{\mathbf{A}_3 \exp[\mathbf{A}_2 \ln(\mathbf{A}_1 \boldsymbol{\theta})]}^{-1} \mathbf{A}_3 \mathbf{D}_{\exp[\mathbf{A}_2 \ln(\mathbf{A}_1 \boldsymbol{\theta})]} \mathbf{A}_2 \mathbf{D}_{\mathbf{A}_1 \boldsymbol{\theta}}^{-1} \mathbf{A}_1.$$

Se forem utilizadas as funções lineares $\mathbf{F}(\boldsymbol{\theta}) = \mathbf{A}_1 \boldsymbol{\theta}$ ou log-lineares $\mathbf{F}(\boldsymbol{\theta}) = \ln(\boldsymbol{\theta})$ ou $\mathbf{F}(\boldsymbol{\theta}) = \mathbf{A}_1 \ln(\boldsymbol{\theta})$, neste último caso com \mathbf{A}_1 ortogonal às restrições naturais, pode-se obter uma estimativa de $\boldsymbol{\theta}$ sob M_F e de sua respectiva matriz de covariâncias assintótica por meio de desenvolvimentos análogos àqueles da Seção 2.4.2.

CAPÍTULO 3

Aspectos práticos

Na Seção 3.1 descreve-se a implementação da teoria exposta no Capítulo 2 em uma biblioteca para o ambiente estatístico R (R Development Core Team, 2006), apresentando cada uma das rotinas desenvolvidas. Na Seção 3.2 ilustra-se a flexibilidade do uso das rotinas com análises dos exemplos introduzidos no capítulo anterior e de outros descritos neste.

3.1 Implementação da biblioteca para o ambiente estatístico R

A notação do Capítulo 2 é mantida no decorrer desta seção para facilitar a inter-relação entre o que foi lá apresentado e o que está implementado na biblioteca.

Sucintamente, descreve-se as principais atribuições de cada uma das rotinas¹ criadas:

- `readCatdata()`, abreviação de *read categorical data*, é responsável pela leitura dos dados, sejam completos ou com omissão;
- `satMarML()`, minguamento de *saturated structural models under MAR (or MCAR) mechanism by maximum likelihood*, realiza a estimação por MV de modelos saturados sob os mecanismos MAR e MCAR com base num objeto² `readCatdata()`. Essa rotina só pode ser utilizada no contexto de dados com omissão;

¹Os nomes das rotinas foram adotados à luz das recomendações de convenções de nomeação utilizadas no R (veja, *e.g.*, o sítio <http://www.maths.lth.se/help/R/RCC/#3.%20Naming%20Conventions>).

²No R, um objeto pode ser um escalar, um vetor, uma matriz, um arranjo (*array*), um conjunto de dados ou até uma lista de objetos. Rotinas que realizam análises estatísticas no R, *e.g.*, modelos lineares generalizados, mistos ou de análise de sobrevivência, em geral, armazenam uma lista de objetos como resultado. As rotinas criadas também foram construídas dessa forma.

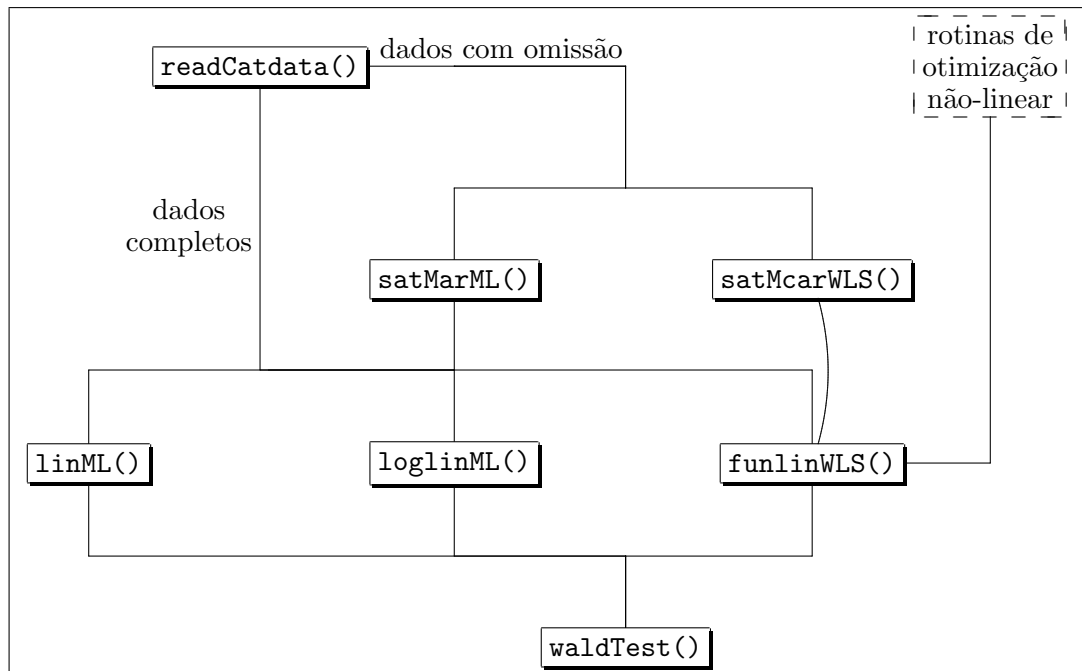


Figura 3.1: Hierarquia de uso das rotinas

- `satMcarWLS()`, encurtamento de *saturated structural models under MCAR mechanism by weighted least squares*, faz a estimação por MQG de modelos saturados sob o mecanismo MCAR com base num objeto `readCatdata()`; por isso, só deve ser utilizada em conjuntos de dados com omissão;
- `linML()`, supressão de *linear models by maximum likelihood*, ajusta modelos lineares pelo procedimento de MV com base num objeto `readCatdata()`, para dados completos, ou num objeto `satMarML()`, no caso de dados com omissão;
- `loglinML()`, redução de *log-linear models by maximum likelihood*, ajusta modelos log-lineares pelo procedimento de MV com base num objeto `readCatdata()`, para dados completos, ou num objeto `satMarML()`, no caso de dados com omissão;
- `funlinWLS()`, desfalcamento de *functional linear models by weighted least squares*, ajusta modelos funcionais lineares pela abordagem de MQG com base num objeto `readCatdata()`, para dados completos, ou num objeto `satMarML()` ou `satMcarWLS()`, no caso de dados com omissão, ou a partir da EMV de θ e correspondente matriz de covariâncias consistentemente estimada obtida, por exemplo, por meio de outras rotinas de otimização não-linear do R, seja no caso de dados completos

ou com omissão, sob qualquer mecanismo de omissão considerado (*e.g.*, alguma estrutura MNAR);

- `waldTest()` realiza testes de hipóteses redutoras por meio da estatística de Wald com base em objetos `linML()`, `loglinML()` ou `funlinWLS()`, desde que se tenha utilizado a formulação em equações livres (veja as Seções 2.4.2 e 2.4.3).

As Figuras 3.1 e 3.2 contêm esquemas que representam as informações apresentadas.

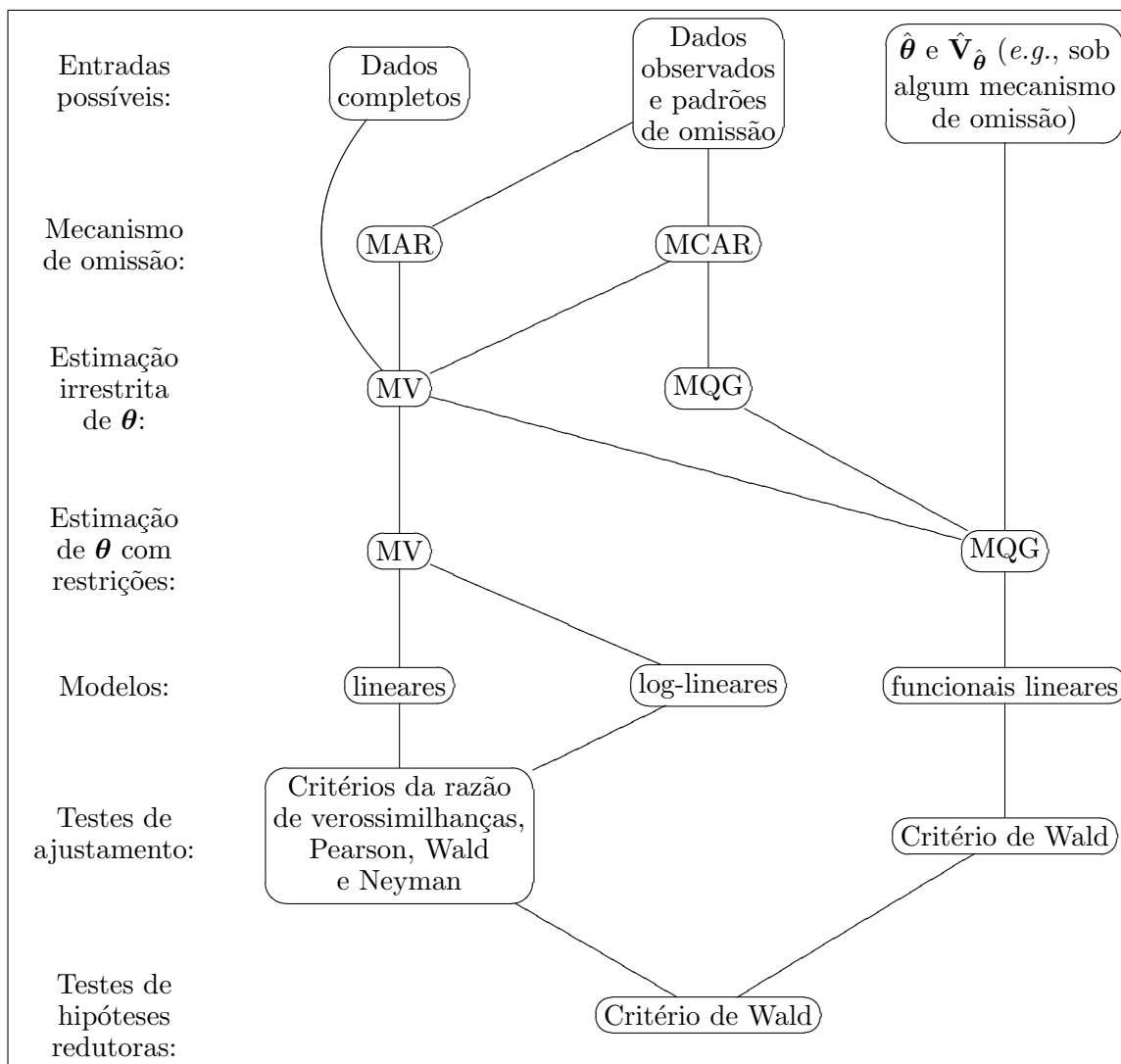


Figura 3.2: Análises possíveis de serem realizadas com a biblioteca

Rodrigues (1996), também tomando como base a teoria apresentada por Paulino (1988, 1991), implementou os processos de estimação por MV e MQG no sistema NTIA³, mas, em ambos, considerou apenas o ajuste de modelos lineares e log-lineares sob o mecanismo MCAR. As funções desenvolvidas neste trabalho generalizam aquelas criadas por Rodrigues (1996) em diversas direções: (i) unificam análises de dados categorizados completos e com omissão; (ii) fornecem estimativas para as matrizes de covariâncias de $\hat{\theta}$ sob o mecanismo MAR; (iii) permitem estimar modelos funcionais lineares por MQG e adotar uma abordagem híbrida (MV/MQG); (iv) possibilitam usar estimativas de θ e de sua matriz de covariâncias obtidas externamente, com estruturas para as probabilidades de categorização ou condicionais de omissão não contempladas (*e.g.*, alguma estrutura MNAR), numa posterior análise por MQG; (v) admitem a inclusão de variáveis explicativas na análise, permitindo que cada estrato referente às subpopulações tenha um padrão de omissão diferenciado.

As rotinas, correspondente documentação e comandos utilizados para reproduzir as análises desta dissertação estão disponíveis no sítio <http://www.poleto.com/missing.html>.

Outras rotinas implementadas para a análise de dados com omissão (categorizados ou não) costumam se basear na metodologia de imputação múltipla, descrita na Seção 1.5 (p.51). *E.g.*, no SAS, veja os procedimentos “MI” e “MIANALYZE”, no R, as bibliotecas “cat”, “mitools”, “mix”, “mvnmle”, “norm” e “pan” (disponíveis em <http://cran.r-project.org>) e, no S-Plus, as bibliotecas “cat”, “mix”, “norm” e “pan” (disponíveis em <http://www.stat.psu.edu/~jls>).

3.2 Análises de dados

Exemplo 2 (Continuação, pp.56 e 72)

Recorde a Tabela 2.1 com as frequências observadas. Como o objetivo do estudo é a avaliação da associação entre o hábito de fumo da mãe (Y_1) e o estado respiratório da criança (Y_2), volta-se a atenção aos logaritmos das razões de chances adjacentes, definidos por

$$\omega_{ij(s)} = \ln \left(\frac{\pi_{ij(s)}\pi_{i+1,j+1(s)}}{\pi_{i,j+1(s)}\pi_{i+1,j(s)}} \right), \quad i, j, s = 1, 2,$$

³As rotinas para a análise de dados categorizados com omissão de Rodrigues (1996) e as rotinas de análise de dados categorizados completos, apresentada em Savalli, Paulino, Silva, Singer, Chicarino, Castro e Tavares (1999), foram escritas na linguagem matricial CM do NTIA. A versão 4.2.2 do NTIA (EMBRAPA, 1997), última que se tem notícia, foi desenvolvida sob o sistema operacional MS-DOS. Devido à falta de continuidade de atualizações, atualmente o NTIA está praticamente em desuso.

ao invés de se trabalhar diretamente com as probabilidades de categorização $\{\pi_{ij(s)}\}$.

Os intervalos para o melhor-pior caso para $\{\omega_{ij(s)}\}$ são apresentados na Tabela 3.1. As suas grandes amplitudes indicam, *a priori*, que a ausência de informações sobre o mecanismo de omissão provavelmente será traduzida em grande ignorância e incerteza sobre as inferências de interesse.

Tabela 3.1: Intervalos para o melhor-pior caso para $\{\omega_{ij(s)}\}$

Limite	$\omega_{11(1)}$	$\omega_{12(1)}$	$\omega_{21(1)}$	$\omega_{22(1)}$	$\omega_{11(2)}$	$\omega_{12(2)}$	$\omega_{21(2)}$	$\omega_{22(2)}$
inferior	-4.32	-4.96	-4.25	-6.21	-2.23	-4.59	-3.22	-3.40
superior	4.34	6.03	5.12	5.12	3.14	3.23	3.07	4.95

Obs.: o Apêndice B.1 contém as alocações que geraram estes intervalos.

Com relação ao mecanismo de omissão, há 36 probabilidades condicionais de omissão $\{\lambda_{t(ijs)}\}$ e apenas 12 freqüências parcialmente categorizadas. Sendo assim, necessita-se da imposição de, no mínimo, 24 restrições, ou equivalentemente, a adoção de modelos em formulação de equações livres que possuam, no máximo, 12 parâmetros. Com este objetivo, considera-se, arbitrariamente, as seguintes estruturas para o mecanismo de omissão: MCAR, MAR,

$$\text{MNAR1} : \begin{cases} \lambda_{2(ijs)} = \alpha_{2(js)}, \\ \lambda_{3(ijs)} = \alpha_{3(is)}, \end{cases}$$

$$\text{MNAR2} : \begin{cases} \lambda_{t(ijs)} = \alpha_{t(1s)}, & \text{se } i = j, \\ \lambda_{t(ijs)} = \alpha_{t(2s)}, & \text{se } |i - j| = 1, \\ \lambda_{t(ijs)} = \alpha_{t(3s)}, & \text{se } |i - j| = 2, \end{cases}$$

$$\text{MNAR3} : \ln \left(\frac{\lambda_{t(ijs)}}{\lambda_{1(ijs)}} \right) = \alpha_{t00(s)} + \alpha_{t02}\mathbf{I}_{[Y_2=2]} + \alpha_{t03}\mathbf{I}_{[Y_2=3]},$$

$$\text{MNAR4} : \ln \left(\frac{\lambda_{t(ijs)}}{\lambda_{1(ijs)}} \right) = \alpha_{t00(s)} + \alpha_{t20}\mathbf{I}_{[Y_1=2]} + \alpha_{t30}\mathbf{I}_{[Y_1=3]} + \alpha_{t02}\mathbf{I}_{[Y_2=2]} + \alpha_{t03}\mathbf{I}_{[Y_2=3]},$$

$$\begin{aligned} \text{MNAR5} : \ln \left(\frac{\lambda_{t(ijs)}}{\lambda_{1(ijs)}} \right) &= \alpha_{t00(s)} + \alpha_{020}\mathbf{I}_{[Y_1=2]} + \alpha_{030}\mathbf{I}_{[Y_1=3]} + \alpha_{002}\mathbf{I}_{[Y_2=2]} + \alpha_{003}\mathbf{I}_{[Y_2=3]} + \\ &\quad \alpha_{022}\mathbf{I}_{[Y_1=2, Y_2=2]} + \alpha_{023}\mathbf{I}_{[Y_1=2, Y_2=3]} + \\ &\quad \alpha_{032}\mathbf{I}_{[Y_2=3, Y_2=2]} + \alpha_{033}\mathbf{I}_{[Y_2=3, Y_2=3]}, \end{aligned}$$

$t = 2, 3$, $s = 1, 2$, $i, j = 1, 2, 3$, em que $I_{[\]}$ denota a função indicadora, assumindo o valor 1 quando a condição explicitada entre chaves for satisfeita e resultando o valor 0 em caso contrário.

Sob a estrutura MNAR1, as probabilidades condicionais de omissão dependem apenas das respostas não-observadas e da cidade de residência, utilizando a idéia dos estimadores protetores, como em Michiels e Molenberghs (1997).

O mecanismo MNAR2 assume que as probabilidades condicionais de omissão dependem da distância entre as categorias de resposta do hábito de fumo da mãe e do estado respiratório da criança, mas, condicionalmente a essas distâncias, essas probabilidades são independentes dos níveis de Y_1 e Y_2 , embora variem para as cidades de residência e os padrões de omissão.

Sob a estrutura MNAR3, $\alpha_{t00(s)}$ é a log-chance de se observar apenas Y_{t-1} em relação a (Y_1, Y_2) para a s -ésima cidade de residência quando $Y_2 = 1$, independentemente dos níveis de Y_1 . α_{t0j} é o acréscimo nessas log-chances, comum às duas cidades, de $Y_2 = j$ em relação a $Y_2 = 1$.

Sob o mecanismo MNAR4, $\alpha_{t00(s)}$ é a log-chance de se observar apenas Y_{t-1} em relação a (Y_1, Y_2) para a s -ésima cidade de residência quando $(Y_1 = 1, Y_2 = 1)$. α_{ti0} é o acréscimo nessas log-chances, comum às duas cidades, de $Y_1 = i$ em relação a $Y_1 = 1$. α_{t0j} tem a mesma interpretação do mecanismo MNAR3.

Sob a estrutura MNAR5, $\alpha_{t00(s)}$ possui a mesma explicação obtida sob a estrutura MNAR4. α_{0i0} e α_{00j} têm interpretações semelhantes às de α_{ti0} e α_{t0j} , mas independentes dos padrões de omissão. O termo α_{0ij} corresponde ao incremento nas log-chances quando $Y_1 = i$ e $Y_2 = j$.

Os modelos estruturais para os mecanismos de omissão, juntamente com modelos saturados para as probabilidades de categorização, foram ajustados pela abordagem de MV. A Tabela 3.2 contém os valores das estatísticas de razão de verossimilhanças de ajuste dos mecanismos de omissão, correspondentes graus de liberdade e número de frequências ampliadas estimadas inferiores a 0.1, indicando o número de probabilidades condicionais de omissão com estimativas na fronteira do espaço paramétrico.

O mecanismo MCAR não se ajusta aos dados ($p < 0.001$), enquanto que o mecanismo MNAR3, única estrutura MNAR não saturada, teve um ajuste satisfatório ($p = 0.482$). Dentre todas as estruturas saturadas, a MAR foi a única que se ajustou perfeitamente. Todas as estruturas MNAR tiveram estimativas de algumas das probabilidades condicionais de omissão na fronteira do espaço paramétrico, ilustrando que, apesar de essa característica estar associada

Tabela 3.2: Estatísticas de razão de verossimilhanças de ajuste dos mecanismos de omissão (Q_V), graus de liberdade (g.l.) e número de freqüências ampliadas < 0.1 ($\#\{\hat{y}_{stij} < 0.1\}$)

Quantidade	MCAR	MAR	MNAR1	MNAR2	MNAR3	MNAR4	MNAR5
Q_V	45.54	0.00	2.78	0.19	3.48	1.44	1.71
g.l.	8	0	0	0	4	0	0
$\#\{\hat{y}_{stij} < 0.1\}$	0	0	9	10	12	24	12

a mecanismos de omissão saturados sem ajuste perfeito (Baker e Laird, 1988), pode também ocorrer em estruturas reduzidas (Smith, Skinner e Clarke, 1999).

As Tabelas 3.3 e 3.4 ilustram os dados observados e as freqüências ampliadas esperadas estimadas sob cada uma das estruturas dos mecanismos de omissão. Como esperado, as somas das freqüências estimadas dos padrões de omissão coincidem sob as estruturas MCAR e MAR. Comparando com a estrutura MAR,

- MNAR1 aloca as unidades com omissão com menor freqüência no estado respiratório normal ($Y_2 = 1$) e em maior número nos estados ofegantes com e sem resfriado ($Y_2 = 2$ e $Y_2 = 3$) para a cidade de Kingston-Harriman. Na cidade de Portage, a estrutura aloca menos freqüentemente em ofegantes com resfriado ($Y_2 = 2$), mais em ofegantes sem resfriado ($Y_2 = 3$) e praticamente o mesmo número no estado respiratório normal ($Y_2 = 1$). Em ambas cidades, as unidades com omissão são alocadas mais freqüentemente nas categorias de hábitos de fumo moderado ($Y_1 = 2$) e forte ($Y_1 = 3$).
- MNAR2 distribui as unidades com omissão menos freqüentemente nas categorias em que $Y_1 = Y_2$ e em um número muito maior quando $|Y_1 - Y_2| = 2$. Com exceção de quando apenas Y_2 está omissa na cidade de Portage, em que nenhuma unidade omissa é alocada em $|Y_1 - Y_2| = 1$, em geral, as unidades são alocadas nessas categorias praticamente da mesma forma que o mecanismo MAR.
- MNAR3 aloca todas as unidades com o estado respiratório da criança omissa na categoria ofegante sem resfriado ($Y_2 = 3$). Quando apenas o hábito de fumo da mãe (Y_1) está omissa, a distribuição é feita similarmente ao mecanismo MAR, já que ambas estruturas dependem apenas do valor observado do estado respiratório da criança (Y_2), embora o

Tabela 3.3: Dados observados e EMV das frequências ampliadas esperadas*

Modelo	Cidade	$Y_1 \setminus Y_2$	$t = 1$			$t = 2$				$t = 3$			Total			
			1	2	3	1	2	3	total	1	2	3	1	2	3	total
Dados observados	Kingston-H.	1	167	17	19	?	?	?	176	?	?	?	?	?	?	
		2	10	1	3	?	?	?	24	?	?	?	?	?	?	
		3	52	10	11	?	?	?	121	?	?	?	?	?	?	
		total								28	10	12				661
	Portage	1	120	22	19	?	?	?	103	?	?	?	?	?	?	
		2	8	5	1	?	?	?	3	?	?	?	?	?	?	
3		39	12	12	?	?	?	80	?	?	?	?	?	?		
	total								31	8	14				477	
MCAR	Kingston-H.	1	143	17	19	158	19	21	198	25	3	3	326	39	43	408
		2	12	2	5	13	2	5	20	2	0	1	28	3	10	41
		3	62	15	16	68	16	18	103	11	3	3	140	34	37	211
		total								37	6	7	494	76	91	661
	Portage	1	108	20	20	84	16	16	115	24	4	4	216	40	40	296
		2	5	3	1	4	3	1	7	1	1	0	11	7	2	19
3		48	15	18	37	12	14	63	11	3	4	95	30	37	162	
	total								36	9	9	322	77	78	477	
MAR	Kingston-H.	1	167	17	19	141	17	19	176	18	5	6	326	39	43	408
		2	10	1	3	16	2	6	24	2	0	1	28	3	10	41
		3	52	10	11	80	19	21	121	8	4	5	140	34	37	211
		total								28	10	12	494	76	91	661
	Portage	1	120	22	19	75	14	14	103	21	4	7	216	40	40	296
		2	8	5	1	2	1	0	3	1	1	0	11	7	2	19
3		39	12	12	47	15	18	80	9	3	7	95	30	37	162	
	total								31	8	14	322	77	78	477	
MNAR1	Kingston-H.	1	166	17	19	119	29	30	177	0	0	0	285	45	49	379
		2	10	1	3	11	4	10	24	5	1	3	26	6	15	47
		3	53	10	11	54	33	33	120	23	9	9	130	52	53	235
		total								28	10	12	440	103	118	661
	Portage	1	121	22	19	76	6	20	102	0	0	0	197	28	39	264
		2	6	4	1	4	1	1	6	0	0	0	10	6	1	17
3		39	13	12	45	5	28	78	32	7	15	116	25	55	196	
	total								32	7	15	323	59	95	477	

*Alguns totais não coincidem com as somas das frequências respectivas devido aos arredondamentos.

Tabela 3.4: EMV das frequências ampliadas esperadas*

Modelo	Cidade	$Y_1 \setminus Y_2$	$t = 1$			$t = 2$				$t = 3$			Total			
			1	2	3	1	2	3	total	1	2	3	1	2	3	total
MNAR2	Kingston-H.	1	167	17	20	108	31	37	176	0	6	10	275	54	67	396
		2	10	1	3	18	1	6	24	4	0	1	31	2	10	43
		3	51	10	11	96	18	7	121	25	4	0	172	32	18	222
		total									29	10	11	479	87	95
	Portage	1	120	22	19	68	0	35	103	0	5	14	188	27	68	283
		2	8	5	1	0	3	0	3	2	0	0	10	8	1	19
		3	39	12	12	73	0	7	80	29	3	0	141	15	19	175
		total									31	8	14	339	50	88
MNAR3	Kingston-H.	1	165	19	18	0	0	177	177	22	4	7	187	23	202	412
		2	10	1	3	0	0	24	24	1	0	1	11	1	28	40
		3	51	11	12	0	0	120	120	7	3	4	58	14	136	208
		total									31	7	12	257	38	366
	Portage	1	122	20	18	0	0	104	104	20	6	8	142	27	130	299
		2	8	5	1	0	0	3	3	1	1	0	9	6	4	20
		3	40	11	13	0	0	78	78	7	3	6	46	14	98	159
		total									28	11	14	198	47	232
MNAR4	Kingston-H.	1	167	17	19	0	0	176	176	0	0	0	167	17	195	379
		2	10	1	3	0	0	24	24	0	0	0	10	1	27	38
		3	50	11	11	0	0	122	122	30	9	11	80	20	144	244
		total									30	9	11	257	38	366
	Portage	1	120	22	19	0	0	103	103	0	0	0	120	22	122	264
		2	8	5	1	0	0	3	3	0	0	0	8	5	4	17
		3	41	11	12	0	0	79	79	29	9	15	70	20	106	196
		total									29	9	15	198	47	232
MNAR5	Kingston-H.	1	166	17	19	145	0	31	176	23	0	5	333	17	55	405
		2	10	1	3	0	0	24	24	0	0	4	10	1	31	42
		3	52	12	11	44	53	25	121	7	8	4	102	73	40	214
		total									29	8	12	445	91	125
	Portage	1	121	22	19	79	0	24	103	23	0	7	223	22	50	295
		2	8	5	1	0	0	3	3	0	0	1	8	5	5	18
		3	39	10	12	25	34	21	80	7	10	6	71	54	39	164
		total									30	10	14	302	81	94

*Alguns totais não coincidem com as somas das frequências respectivas devido aos arredondamentos.

mecanismo MAR permita que as probabilidades condicionais desse padrão de omissão difiram entre as duas cidades e a estrutura MNAR3, não.

- MNAR4 distribui todas as unidades com omissão nas categorias ofegante sem resfriado ($Y_2 = 3$) e no hábito de fumo forte ($Y_1 = 3$).
- MNAR5 aloca as unidades mais em $(Y_1 = 1, Y_2 = 1)$, $(Y_1 = 3, Y_2 = 2)$ e $Y_2 = 3$.

Utilizando a abordagem híbrida (MV/MQG), descrita na Seção 2.4.3, e a matriz de informação observada, como medida de precisão na metodologia de MV, estimativas de $\{\omega_{ij(s)}\}$ e respectivos erros padrões são resumidos na Tabela 3.5. Também são apresentados valores-p de testes de Wald de ajuste do modelo de razão de chances adjacentes comum, $H : \omega_{ij(s)} = \omega$, com correspondentes estimativas do parâmetro e de seu erro padrão.

Todos os resultados obtidos sob a ACC e os mecanismos MCAR e MAR são praticamente os mesmos, embora os erros padrões dos últimos sejam menores a partir da 3^a casa decimal. As estimativas obtidas sob as estruturas MNAR variam consideravelmente, mas mantêm-se sempre dentro dos intervalos para o melhor-pior caso. O modelo de razão de chances adjacentes se ajusta bem (razoavelmente) aos dados na ACC e com as estruturas MCAR/MAR, MNAR1 e MNAR2 (MNAR3). Entretanto, a estimativa negativa do logaritmo da razão de chances adjacentes comum sob o mecanismo MNAR2 não é esperada, bem como algumas das outras estimativas obtidas para os $\{\omega_{ij(s)}\}$ sob os demais mecanismos.

Para avaliar se essas associações inversas podem ser uma conseqüência de instabilidades ocasionadas ou agravadas pelas baixas freqüências observadas de mães com hábito de fumo moderado, reanalisa-se os dados, a seguir, após agrupar as categorias de hábito de fumo moderado e forte, resultando numa nova variável (Y'_1), indicando simplesmente se a mãe é ($i = 2$) ou não ($i = 1$) fumante. Considera-se também a união das categorias ofegante com e sem resfriado ($j = 2$), resguardando a categoria normal ($j = 1$), numa variável com o mesmo rótulo anterior (Y'_2).

Conforme Bishop *et al.* (1975, pp.27-29) demonstram, se houver independência na tabela original, esta será mantida na tabela com as categorias combinadas, enquanto que o inverso pode não ocorrer devido a uma estrutura mais complexa de dependência existente na tabela expandida. Contudo, Kateri e Iliopoulos (2004) indicam que quando há homogeneidade das distribuições condicionais às linhas ou colunas que se quer combinar, ou, equivalentemente, a falta de associação entre as celas correspondentes, a estrutura de associação original não

Tabela 3.5: Estimativas e respectivos erros padrões de $\{\omega_{ij(s)}\}$ e ω , e valor-p do teste de Wald de $H : \omega_{ij(s)} = \omega$

	Quant.	ACC	MCAR/MAR	MNAR1	MNAR2	MNAR3	MNAR4	MNAR5
estimativas	$\omega_{11(1)}$	-0.02	0.05	0.31	-1.31	-0.02	-0.02	0.67
	$\omega_{12(1)}$	0.99	0.99	0.95	1.57	0.86	0.86	2.25
	$\omega_{21(1)}$	0.65	0.65	0.62	1.25	0.65	0.92	1.96
	$\omega_{22(1)}$	-1.00	-1.00	-1.00	-2.35	-0.72	-1.32	-4.02
	$\omega_{11(2)}$	1.23	1.22	1.38	1.72	1.23	1.23	1.85
	$\omega_{12(2)}$	-1.46	-1.50	-1.76	-2.79	-1.94	-1.94	-0.82
	$\omega_{21(2)}$	-0.71	-0.70	-0.94	-2.04	-0.71	-0.78	0.20
	$\omega_{22(2)}$	1.61	1.71	2.18	2.11	2.26	1.89	-0.33
	ω	0.19	0.20	0.24	-0.23	0.20	0.10	2.15
erros padrões	$\omega_{11(1)}$	1.08	1.08	1.02	1.13	1.08	1.08	1.09
	$\omega_{12(1)}$	1.20	1.20	1.16	1.23	1.05	1.05	1.54
	$\omega_{21(1)}$	1.10	1.10	1.06	1.14	1.10	1.08	1.31
	$\omega_{22(1)}$	1.23	1.23	1.23	1.26	1.07	1.05	1.77
	$\omega_{11(2)}$	0.62	0.62	0.66	0.99	0.62	0.62	0.65
	$\omega_{12(2)}$	1.14	1.14	1.17	1.34	0.71	0.71	1.24
	$\omega_{21(2)}$	0.66	0.66	0.71	0.99	0.66	0.62	0.87
	$\omega_{22(2)}$	1.17	1.17	1.20	1.41	0.74	0.71	1.57
	ω	0.07	0.07	0.06	0.08	0.04	0.03	0.28
	valor-p	0.623	0.597	0.379	0.303	0.121	0.026	<0.001

é afetada. Desta forma, a aceitação da hipótese $H : \omega_{12(s)} = \omega_{21(s)} = \omega_{22(s)} = 0$, $s = 1, 2$, implica que as categorias podem ser combinadas como sugerido. Aplicando-se o teste de Wald, isso é verificado sob a ACC e os mecanismos MCAR/MAR e MNAR1 ($p > 0.500$), mas não sob as outras estruturas MNAR ($p < 0.100$). Como não se conhece o verdadeiro mecanismo de omissão, os resultados com as categorias combinadas devem ser analisados com cautela.

Os intervalos para o melhor-pior caso para os logaritmos das razões de chances das duas cidades com as categorias mescladas, $\{\omega'_{(s)}\}$, são expostos na Tabela 3.6. As amplitudes dos intervalos são apenas um pouco menores do que as dos anteriores (compare com a Tabela 3.1), apontando que a tabela condensada não deve reduzir a ignorância sobre o mecanismo de omissão e nem a incerteza sobre as inferências de interesse.

Tabela 3.6: Intervalos para o melhor-pior caso para $\{\omega'_{(s)}\}$

Limite	$\omega'_{(1)}$	$\omega'_{(2)}$
inferior	-2.58	-2.00
superior	3.46	2.88

A nova tabela de contingência está associada a 16 probabilidades condicionais de omissão $\{\lambda_{t(ijs)}\}$ e 8 frequências parcialmente categorizadas. Por livre-arbítrio, considera-se as estruturas para o mecanismo de omissão: MCAR, MAR,

$$\text{MNAR1} : \begin{cases} \lambda_{2(ijs)} = \alpha_{2(js)}, \\ \lambda_{3(ijs)} = \alpha_{3(is)}, \end{cases}$$

$$\text{MNAR2} : \begin{cases} \lambda_{t(ijs)} = \alpha_{t(1s)}, & \text{se } i = j, \\ \lambda_{t(ijs)} = \alpha_{t(2s)}, & \text{se } |i - j| = 1, \end{cases}$$

$$\text{MNAR3} : \ln \left(\frac{\lambda_{t(ijs)}}{\lambda_{1(ijs)}} \right) = \alpha_{t00(s)} + \alpha_{t02} \mathbf{I}_{[Y'_2=2]},$$

$$\text{MNAR4} : \ln \left(\frac{\lambda_{t(ijs)}}{\lambda_{1(ijs)}} \right) = \alpha_{t00(s)} + \alpha_{t20} \mathbf{I}_{[Y'_1=2]} + \alpha_{t02} \mathbf{I}_{[Y'_2=2]},$$

$$\text{MNAR5} : \ln \left(\frac{\lambda_{t(ijs)}}{\lambda_{1(ijs)}} \right) = \alpha_{t00(s)} + \alpha_{020} \mathbf{I}_{[Y'_1=2]} + \alpha_{002} \mathbf{I}_{[Y'_2=2]} + \alpha_{022} \mathbf{I}_{[Y'_1=2, Y'_2=2]},$$

$$\text{MNAR6} : \ln \left(\frac{\lambda_{t(ijs)}}{\lambda_{1(ijs)}} \right) = \alpha_{t00(s)} + \alpha_{020(s)} \mathbf{I}_{[Y'_1=2]} + \alpha_{002(s)} \mathbf{I}_{[Y'_2=2]},$$

$$\text{MNAR7} : \lambda_{t(ijs)} = \alpha_{t(ij)},$$

$t = 2, 3$, $i, j, s = 1, 2$. Note que o mecanismo MNAR5 deixa de ser saturado e, com exceção das estruturas MNAR6 e MNAR7, as demais são análogas às utilizadas anteriormente, isto é, antes de combinar as categorias. O mecanismo MNAR6 pode ser obtido do MNAR4 assumindo que as probabilidades condicionais de omissão dependem do hábito de fumo da mãe e do estado respiratório da criança da mesma maneira para os padrões de omissão, mas de formas diferentes para as duas cidades. Em contrapartida, a estrutura MNAR7 assume que cada padrão de omissão pode depender de forma diferente do hábito de fumo da mãe e do estado

respiratório da criança, mas supondo que este mecanismo seja igual para as duas cidades.

Os resultados correspondentes às Tabelas 3.2, 3.3, 3.4 e 3.5 são apresentados nas Tabelas 3.7, 3.8, 3.9 e 3.10.

Tabela 3.7: Estatísticas de razão de verossimilhanças de ajuste dos mecanismos de omissão (Q_V), graus de liberdade (g.l.) e número de frequências estimadas ampliadas < 0.1 ($\#\{\hat{y}_{stij} < 0.1\}$) da tabela com as categorias combinadas

Quantidade	MCAR	MAR	MNAR1	MNAR2	MNAR3	MNAR4	MNAR5	MNAR6	MNAR7
Q_V	33.92	0.00	0.08	0.73	2.11	1.42	1.33	0.00	6.69
g.l.	4	0	0	0	2	0	1	0	0
$\#\{\hat{y}_{stij} < 0.1\}$	0	0	4	2	4	8	0	0	8

Os mecanismos MCAR e MNAR3 permanecem, respectivamente, mal e bem ajustados ($p < 0.001$ e $= 0.349$). A estrutura MNAR5 também apresentou um ajuste satisfatório ($p = 0.249$) e agora não gera mais estimativas das probabilidades condicionais de omissão na fronteira do espaço paramétrico. Dentre todas as estruturas MNAR, a MNAR6 foi a única que teve um ajuste perfeito e a MNAR7 teve o pior ajuste.

Comparando os mecanismos MNAR com o MAR, as estruturas MNAR1 a MNAR3 alocam as unidades com omissão similarmente às descritas antes de combinar as categorias. A estrutura MNAR4, que antes alocava as unidades com omissão apenas nos ofegantes sem resfriado e no hábito forte de fumar, quando somente respostas das variáveis respectivas estão omissas, agora distribui apenas nas categorias normal e não-fumante. O mecanismo MNAR5 (MNAR7) distribui as unidades com omissão semelhantemente ao MNAR2 (MNAR4), mas de forma mais extrema, quando comparados com o MAR. O mesmo acontece para a estrutura MNAR6 em relação à MNAR1, mas de maneira menos extrema.

A associação entre o hábito de fumo das mães e o estado respiratório das crianças é a mesma para as cidades de Kingston-Harriman e Portage em todas as análises da Tabela 3.10. Embora as estruturas MNAR2, MNAR4, MNAR5 e MNAR7 apresentem estimativas negativas para ω' , apenas sob a primeira a associação contraria o esperado ($p = 0.002$), enquanto que sob as outras três há independência entre as duas variáveis ($p > 0.300$). Sob as demais análises, mães fumantes estão associadas às crianças com estado respiratório ofegante.

Adicionando o termo $\alpha_{022}I_{[Y'_1=2, Y'_2=2]}$ do mecanismo MNAR5 à estrutura MNAR6, obtém-se

Tabela 3.8: Dados observados e EMV das frequências ampliadas esperadas*

Modelo	Cidade	MFum	EstRC	$t = 1$		$t = 2$			$t = 3$		Total		
				norm.	ofeg.	norm.	ofeg.	total	norm.	ofeg.	norm.	ofeg.	total
Dados observados	Kingston	não		167	36	?	?	176	?	?	?	?	
		sim		62	25	?	?	145	?	?	?	?	
		total							28	22		661	
	Portage	não		120	41	?	?	103	?	?	?	?	
		sim		47	30	?	?	83	?	?	?	?	
		total							31	22		477	
MCAR	Kingston	não		143	36	158	40	198	25	6	326	82	408
		sim		74	37	81	41	123	13	6	168	85	253
		total							37	13	494	167	661
	Portage	não		108	40	84	31	115	24	9	216	80	296
		sim		53	37	41	29	71	12	8	106	75	181
		total							36	17	322	155	477
MAR	Kingston	não		167	36	141	35	176	18	11	326	82	408
		sim		62	25	96	49	145	10	11	168	85	253
		total							28	22	494	167	661
	Portage	não		120	41	75	28	103	21	11	216	80	296
		sim		47	30	49	34	83	10	11	106	75	181
		total							31	22	322	155	477
MNAR1	Kingston	não		166	36	119	59	177	0	0	285	94	379
		sim		63	25	65	79	144	28	23	155	127	282
		total							28	23	440	221	661
	Portage	não		120	41	74	29	102	0	0	194	70	264
		sim		47	30	48	36	84	31	22	125	88	213
		total							31	22	319	158	477
MNAR2	Kingston	não		167	38	91	85	176	0	20	258	143	401
		sim		60	25	131	14	145	30	0	221	39	260
		total							30	20	480	181	661
	Portage	não		120	41	39	64	103	8	20	167	125	292
		sim		47	30	73	10	83	23	2	143	42	185
		total							31	22	310	167	477

MFum: Mãe fumante, EstRC: Estado respiratório da criança, norm.: normal, ofeg.: ofegante.

*Alguns totais não coincidem com as somas das frequências respectivas devido aos arredondamentos.

Tabela 3.9: EMV das frequências ampliadas esperadas*

Modelo	Cidade	MFum	EstRC	$t = 1$		$t = 2$			$t = 3$		Total		
				norm.	ofeg.	norm.	ofeg.	total	norm.	ofeg.	norm.	ofeg.	total
MNAR3	Kingston	não		165	35	0	178	178	22	11	187	224	412
		sim		61	28	0	143	143	8	9	70	180	249
		total							31	19	257	404	661
	Portage	não		122	38	0	104	104	20	14	142	156	299
		sim		48	30	0	82	82	8	11	56	123	178
		total							28	25	198	279	477
MNAR4	Kingston	não		164	38	176	0	176	30	20	371	58	429
		sim		62	25	145	0	145	0	0	207	25	232
		total							30	20	578	83	661
	Portage	não		123	39	103	0	103	29	24	254	63	317
		sim		47	30	83	0	83	0	0	130	30	160
		total							29	24	384	93	477
MNAR5	Kingston	não		167	39	57	118	175	9	18	232	175	408
		sim		60	25	137	9	146	21	1	217	36	253
		total							30	20	450	211	661
	Portage	não		120	38	27	77	104	8	22	155	137	292
		sim		49	30	75	7	82	21	2	145	39	185
		total							29	24	300	177	477
MNAR6	Kingston	não		167	36	111	65	176	17	10	295	111	406
		sim		62	25	69	76	145	11	12	142	113	255
		total							28	22	437	224	661
	Portage	não		120	41	67	36	103	19	10	206	87	293
		sim		47	30	42	41	83	12	12	100	83	184
		total							31	22	307	170	477
MNAR7	Kingston	não		170	37	166	0	166	35	21	371	58	429
		sim		67	25	140	0	140	0	0	207	25	232
		total							35	21	578	83	661
	Portage	não		117	40	113	0	113	24	23	254	63	317
		sim		42	30	88	0	88	0	0	130	30	160
		total							24	23	384	93	477

MFum: Mãe fumante, EstRC: Estado respiratório da criança, norm.: normal, ofeg.: ofegante.

*Alguns totais não coincidem com as somas das frequências respectivas devido aos arredondamentos.

Tabela 3.10: Estimativas e respectivos erros padrões de $\{\omega'_{(s)}\}$ e ω' , e valor-p do teste de Wald de $H : \omega'_{(s)} = \omega'$

Quant.	ACC	MCAR/MAR	MNAR1	MNAR2	MNAR3	MNAR4	MNAR5	MNAR6	MNAR7*	
estima- tivas	$\omega'_{(1)}$	0.63	0.69	0.90	-1.15	0.77	-0.26	-1.52	0.75	-0.26
	$\omega'_{(2)}$	0.63	0.65	0.67	-0.94	0.69	-0.07	-1.19	0.68	-0.07
	ω'	0.63	0.67	0.78	-1.07	0.74	-0.16	-0.16	0.71	-0.20
erros padrões	$\omega'_{(1)}$	0.30	0.30	0.28	0.43	0.18	0.25	3.29	0.29	1.03
	$\omega'_{(2)}$	0.30	0.30	0.27	0.55	0.21	0.25	2.69	0.29	1.06
	ω'	0.21	0.21	0.19	0.34	0.14	0.18	1.58	0.20	1.02
valor-p	0.998	0.910	0.541	0.761	0.788	0.600	0.638	0.863	0.664	

*Estimativas das probabilidades condicionais deste mecanismo de omissão na fronteira do espaço paramétrico foram substituídas por 0.005, pois valores menores que este geravam variâncias negativas.

o mecanismo de omissão sobre-parametrizado

$$\text{MNAR8} : \ln \left(\frac{\lambda_{t(ijs)}}{\lambda_{1(ijs)}} \right) = \alpha_{t00(s)} + \alpha_{020(s)} I_{[Y'_1=2]} + \alpha_{002(s)} I_{[Y'_2=2]} + \alpha_{022} I_{[Y'_1=2, Y'_2=2]},$$

com $\tau = \alpha_{022}$ representando o parâmetro de sensibilidade e os demais, o parâmetro estimável. Quanto maior é α_{022} , maior é a probabilidade de Y'_1 ou Y'_2 estar omissos quando $(Y'_1 = 2, Y'_2 = 2)$. Para compensar isto, à medida que α_{022} aumenta, o mecanismo de omissão distribui mais as unidades com omissão em $(Y'_1 = 2, Y'_2 = 2)$ e, conseqüentemente, maior será ω' . As estimativas e intervalos de 95% de confiança para ω' variando α_{022} de -5 a 5 estão ilustradas na Figura 3.3. O intervalo de ignorância para ω' é $(-1.62; 2.78)$, e o de 95% de incerteza, construído por meio da união dos intervalos de 95% de confiança de cada valor de α_{022} , é $(-2.00; 3.16)$.

Outra estrutura sobre-parametrizada também foi considerada, adicionando $\alpha_{022} I_{[Y'_1=2, Y'_2=2]}$ do mecanismo MNAR5 na estrutura MNAR4, resultando em

$$\text{MNAR9} : \ln \left(\frac{\lambda_{t(ijs)}}{\lambda_{1(ijs)}} \right) = \alpha_{t00(s)} + \alpha_{t20} I_{[Y'_1=2]} + \alpha_{t02} I_{[Y'_2=2]} + \alpha_{022} I_{[Y'_1=2, Y'_2=2]}.$$

Manteve-se $\tau = \alpha_{022}$ como o parâmetro de sensibilidade. Ao variar α_{022} de -5 a 5 com incrementos de 0.01, obteve-se estimativas negativas para as variâncias quando $-2.82 \leq \alpha_{022} \leq -2.38$ e $\alpha_{022} = 2.05$. Na Figura 3.4 são ilustradas as estimativas e intervalos de 95% de confiança para ω' variando α_{022} de -2 a 2. Pode-se observar que as estimativas para $\omega'(\alpha_{022})$ e suas variâncias não variam suavemente para valores de α_{022} próximos, como foi observado

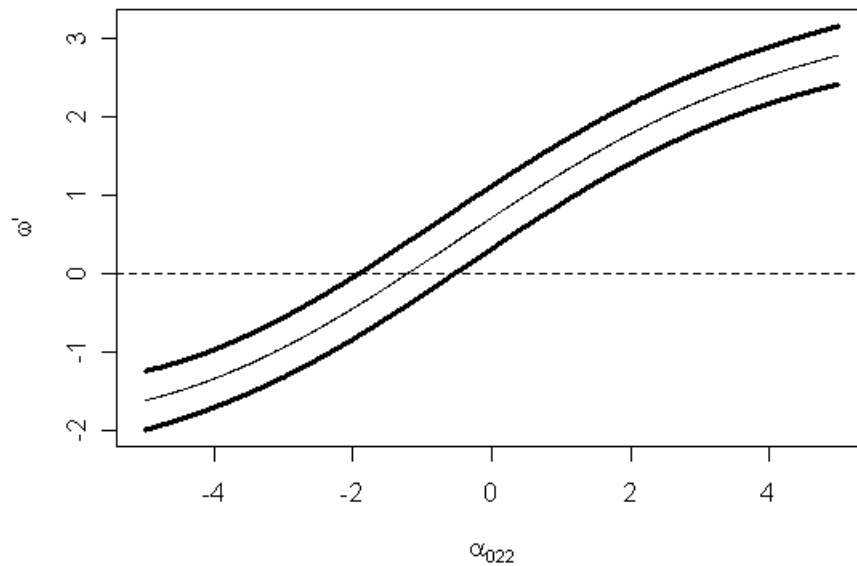


Figura 3.3: Estimativas e intervalos de 95% de confiança para ω' variando α_{022} no MNAR8

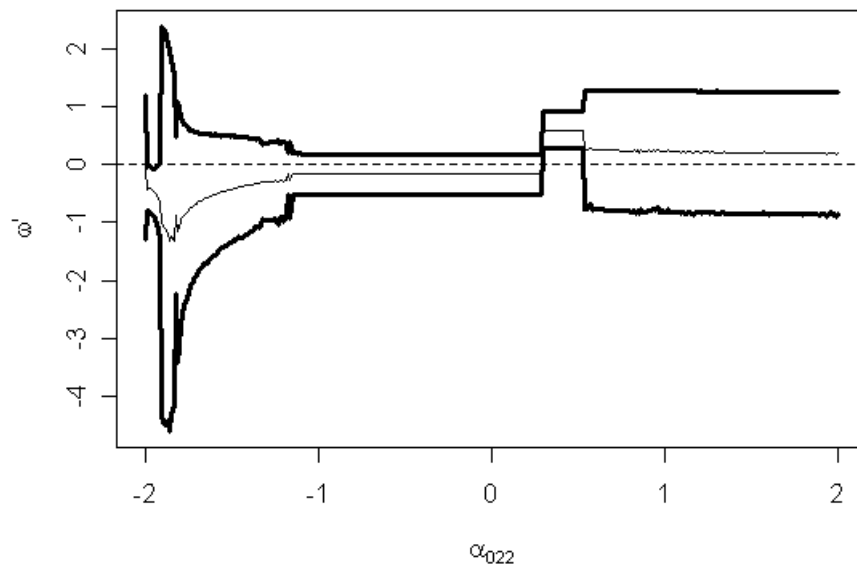


Figura 3.4: Estimativas e intervalos de 95% de confiança para ω' variando α_{022} no MNAR9

sob mecanismo MNAR8 (Figura 3.3), demonstrando a grande instabilidade das inferências sob a estrutura MNAR9.

Tanto com os dados originais, quanto com as categorias agrupadas, sem maiores informações sobre o mecanismo de omissão não há um consenso entre as análises de sensibilidade de que mães fumantes estejam associadas às crianças em estado respiratório ofegante. ■

Exemplo 3 (Continuação, pp.58 e 73)

Reveja a Tabela 2.3 com as freqüências observadas. Como o propósito do estudo é a comparação das distribuições marginais dos graus de suscetibilidade à cárie dentária obtidas pelos dois métodos e a avaliação da concordância entre suas classificações, os parâmetros de interesse são $\pi_{i+} - \pi_{+i}$, $i = 1, 2$, κ (*kappa*⁴ de Cohen), κ_{w1} (*kappa* ponderada⁵ com pesos quadráticos) e κ_{w2} (*kappa* ponderada com pesos⁶ absolutos).

Os intervalos para o melhor-pior caso para os parâmetros de interesse são apresentados na Tabela 3.11. Como todos os intervalos contêm o valor zero, a ausência total de informações sobre o mecanismo de omissão deve criar demasiada ignorância e incerteza sobre as inferências de interesse.

Tabela 3.11: Intervalos para o melhor-pior caso para os parâmetros de interesse

Limite	$\pi_{1+} - \pi_{+1}$	$\pi_{2+} - \pi_{+2}$	κ	κ_{w1}	κ_{w2}
inferior	-0.052	-0.351	-0.187	-0.014	-0.107
superior	0.134	0.124	0.317	0.502	0.398

Obs.: o Apêndice B.2 contém as alocações que geraram estes intervalos.

⁴Há controvérsias de que medidas do tipo *kappa* são adequadas para mensurar concordância, uma vez que dependem fortemente das distribuições marginais (Sprott, 2000, pp.55-59). Uma análise dos resíduos (*e.g.*, padronizados de Pearson), sob o modelo de independência, e o ajuste de modelos de quase-independência e/ou quase-simetria (Agresti, 2002; Paulino e Singer, 2006) possibilitam uma investigação mais apurada dos padrões de concordância ou discordância do que o simples embasamento num único índice. Por simplicidade, tais análises não serão aqui consideradas.

⁵Prado (2004) apresenta uma modelagem da medida *kappa* ponderada com o uso de covariáveis por meio de equações de estimação generalizadas.

⁶Se existisse possibilidade, seria mais adequado contatar o pesquisador para atribuir os pesos que melhor descrevem as discordâncias de acordo com suas gravidades.

Diferentemente dos outros exemplos, em que a omissão se deve às variáveis serem parcialmente observadas, o padrão de omissões neste caso é provocado pelo confundimento de categorias vizinhas. Uma vez que o cenário de omissão $t = 2$ ($t = 3$) é constituído pelo emaranhamento das categorias $i = 1, 2$ ($i = 2, 3$), em todas as estruturas para os mecanismos de omissão assume-se que $\lambda_{2(3j)} = 0$ ($\lambda_{3(1j)} = 0$), já que não há outros motivos para se supor o contrário. Obviamente, o mecanismo MCAR, $\lambda_{t(ij)} = \alpha_t$, não se encaixa em tal conjunto de estruturas consideradas⁷. Note também que, sob essa ótica, as frequências associadas às classes artificialmente criadas anteriormente, n_{24} e n_{31} , constituem zeros estruturais e não amostrais. Na Tabela 3.12 são apresentadas as estruturas para as probabilidades condicionais de omissão examinadas.

Sob a estrutura MNAR1, $\alpha_{t(1)}$ e $\alpha_{t(2)}$ são as probabilidades condicionais do t -ésimo cenário de omissão associadas às categorias que indicam, respectivamente, maior e menor concordância dentro de cada resultado do método convencional.

O mecanismo MNAR2 utiliza o conceito dos estimadores protetores, ou seja, de que a omissão depende apenas das categorias não-observadas.

A estrutura MNAR3 foi idealizada supondo que as probabilidades condicionais de omissão dependem apenas do método convencional (sempre observado) e da ordenação das categorias não-observadas, independentemente do padrão de omissão.

Em outra tentativa de associar o mecanismo de omissão à concordância ou discordância entre os métodos, a estrutura MNAR4 assume que as probabilidades condicionais de omissão estariam dispostas em ordens inversas nos dois padrões de omissão.

As estruturas para os mecanismos de omissão, juntamente com modelos saturados para as probabilidades de categorização, foram ajustadas pelo procedimento de MV. A Tabela 3.13 contém os valores das estatísticas de razão de verossimilhanças de ajuste dos mecanismos de omissão, correspondentes graus de liberdade e número de probabilidades condicionais de omissão com estimativas na fronteira do espaço paramétrico.

As estruturas não saturadas MAR_{red} , MNAR1 e MNAR2 se ajustam bem aos dados ($p=0.585$, 0.780 e 0.495), apesar de a última ter apresentado estimativas das probabilidades condicionais de omissão na fronteira do espaço paramétrico. Dentre as estruturas saturadas, a MAR foi a única que se ajustou perfeitamente, enquanto que as estruturas MNAR3 e MNAR4

⁷Por curiosidade, o mecanismo MCAR não se ajusta aos dados ($p<0.001$), de acordo com o TRV.

Tabela 3.12: Estruturas para as probabilidades condicionais de omissão $\{\lambda_{t(ij)}\}$

Modelo	Métodos simpl.\conv.	$t = 2$			$t = 3$		
		alto	médio	baixo	alto	médio	baixo
MAR _{red}	alto	α_2	α_2	α_2	0	0	0
	médio	α_2	α_2	α_2	α_3	α_3	α_3
	baixo	0	0	0	α_3	α_3	α_3
MAR	alto	$\alpha_{2(1)}$	$\alpha_{2(2)}$	$\alpha_{2(3)}$	0	0	0
	médio	$\alpha_{2(1)}$	$\alpha_{2(2)}$	$\alpha_{2(3)}$	$\alpha_{3(1)}$	$\alpha_{3(2)}$	$\alpha_{3(3)}$
	baixo	0	0	0	$\alpha_{3(1)}$	$\alpha_{3(2)}$	$\alpha_{3(3)}$
MNAR1	alto	$\alpha_{2(1)}$	$\alpha_{2(2)}$	$\alpha_{2(2)}$	0	0	0
	médio	$\alpha_{2(2)}$	$\alpha_{2(1)}$	$\alpha_{2(1)}$	$\alpha_{3(1)}$	$\alpha_{3(1)}$	$\alpha_{3(2)}$
	baixo	0	0	0	$\alpha_{3(2)}$	$\alpha_{3(2)}$	$\alpha_{3(1)}$
MNAR2	alto	$\alpha_{2(1)}$	$\alpha_{2(1)}$	$\alpha_{2(1)}$	0	0	0
	médio	$\alpha_{2(2)}$	$\alpha_{2(2)}$	$\alpha_{2(2)}$	$\alpha_{3(1)}$	$\alpha_{3(1)}$	$\alpha_{3(1)}$
	baixo	0	0	0	$\alpha_{3(2)}$	$\alpha_{3(2)}$	$\alpha_{3(2)}$
MNAR3	alto	α_1	α_3	α_5	0	0	0
	médio	α_2	α_4	α_6	α_1	α_3	α_5
	baixo	0	0	0	α_2	α_4	α_6
MNAR4	alto	α_1	α_3	α_5	0	0	0
	médio	α_2	α_4	α_6	α_6	α_4	α_2
	baixo	0	0	0	α_5	α_3	α_1

$\lambda_{1(ij)} = 1 - \lambda_{2(ij)} - \lambda_{3(ij)}$, simpl.: simplificado, conv.: convencional.

Tabela 3.13: Estatísticas de razão de verossimilhanças de ajuste dos mecanismos de omissão (Q_V), graus de liberdade (g.l.) e número de probabilidades condicionais de omissão na fronteira do espaço paramétrico ($\#\{\hat{\lambda}_{t(ij)} \approx 0\}$)

Quantidade	MAR _{red}	MAR	MNAR1	MNAR2	MNAR3	MNAR4
Q_V	2.84	0.00	0.50	1.41	1.18	3.51
g.l.	4	0	2	2	0	0
$\#\{\hat{\lambda}_{t(ij)} \approx 0\}$	0	0	0	3	2	6

também tiveram estimativas das probabilidades condicionais de omissão na fronteira do espaço paramétrico.

A Tabela 3.14 ilustra os dados observados e as frequências ampliadas esperadas estimadas sob cada uma das estruturas dos mecanismos de omissão.

Para cada resultado do método convencional, a estrutura MNAR1 aloca as unidades com omissão mais freqüentemente nas categorias associadas a uma maior discordância, quando há confundimento entre alto e médio, e nas categorias correspondentes a uma maior concordância, quando não há distinção entre médio e baixo.

O mecanismo MNAR2 distribui as unidades com omissão mais fortemente nos graus de suscetibilidade à cárie mais elevados.

A estrutura MNAR3 aloca as unidades com omissão mais freqüentemente na categoria mais elevada do método simplificado, quando o método convencional é alto ou baixo, e na categoria mais baixa do método simplificado, quando o método convencional é médio.

O mecanismo MNAR4 distribui todas as unidades com omissão nas categorias associadas a uma concordância moderada, $|i - j| = 1$, e nenhuma nas categorias de concordância e discordância totais.

Utilizando a abordagem híbrida (MV/MQG) e a matriz de informação observada, como medida de precisão na metodologia de MV, estimativas, erros padrões e valores-p dos testes de Wald dos parâmetros de interesse são resumidos na Tabela 3.15.

A homogeneidade das distribuições marginais dos níveis de suscetibilidade à cárie é evidenciada em todas as análises que incorporam os dados omissos, o que não acontece na ACC, indicando que os padrões de omissão desse exemplo, sejam aleatórios ou não, afetam gravemente os valores das estimativas, já que o mecanismo MCAR não se afigura adequado, único caso em que a ACC proporcionaria estimadores não enviesados.

Tanto na ACC quanto nas análises sob os mecanismos de omissão considerados, a concordância entre os métodos convencional e simplificado parece estatisticamente igual à esperada ao acaso, segundo a medida *kappa*. Ao incorporar concordâncias/discordâncias moderadas, $|i - j| = 1$, com pesos absolutos ou, principalmente, quadráticos, já não é tão evidente que a concordância entre os dois métodos seja igual à esperada fortuitamente sob os mecanismos MAR e MNAR2.

Tabela 3.14: Dados observados e EMV das frequências ampliadas esperadas*

Modelo	Métodos simpl.\conv.	$t = 1$			$t = 2$			$t = 3$			Total			
		alto	médio	baixo	alto	médio	baixo	alto	médio	baixo	alto	médio	baixo	total
Dados obs.	alto	7	11	2	?	?	?	⊗	⊗	⊗	?	?	?	
	médio	3	9	5	?	?	?	?	?	?	?	?	?	
	baixo	0	10	4	⊗	⊗	⊗	?	?	?	?	?	?	
	total				8	7	3	7	14	7				97
MAR _{red}	alto	7.8	10.4	1.9	2.5	3.4	0.6	⊗	⊗	⊗	10.3	13.8	2.5	26.6
	médio	5.3	7.6	4.3	3.6	5.2	2.9	5.8	8.4	4.8	14.7	21.2	12.0	48.0
	baixo	0.0	9.7	3.9	⊗	⊗	⊗	0.0	6.4	2.6	0.0	16.0	6.4	22.5
	total				6.1	8.5	3.6	5.8	14.8	7.3	25.0	51.0	21.0	97.0
MAR	alto	7.0	11.0	2.0	3.3	2.8	0.5	⊗	⊗	⊗	10.3	13.8	2.5	26.6
	médio	3.0	9.0	5.0	4.7	4.2	2.5	7.0	8.0	4.6	14.7	21.2	12.0	48.0
	baixo	0.0	10.0	4.0	⊗	⊗	⊗	0.0	6.0	2.4	0.0	16.0	6.4	22.5
	total				8.0	7.0	3.0	7.0	14.0	7.0	25.0	51.0	21.0	97.0
MNAR1	alto	7.0	10.1	2.3	0.7	6.5	1.5	⊗	⊗	⊗	7.8	16.6	3.8	28.2
	médio	3.3	8.7	5.2	6.8	1.6	1.0	7.2	7.4	4.2	17.2	17.7	10.3	45.3
	baixo	0.0	9.9	4.0	⊗	⊗	⊗	0.0	6.8	2.9	0.0	16.7	6.8	23.5
	total				7.5	8.2	2.5	7.2	14.2	7.0	25.0	51.0	21.0	97.0
MNAR2	alto	8.0	9.5	2.5	6.6	7.8	2.0	⊗	⊗	⊗	14.6	17.4	4.5	36.4
	médio	3.8	8.6	4.6	0.4	0.8	0.4	6.3	14.2	7.5	10.4	23.6	12.5	46.6
	baixo	0.0	10.0	4.0	⊗	⊗	⊗	0.0	0.0	0.0	0.0	10.0	4.0	14.0
	total				6.9	8.6	2.4	6.3	14.2	7.5	25.0	51.0	21.0	97.0
MNAR3	alto	6.3	11.0	2.0	6.3	0.0	2.5	⊗	⊗	⊗	12.5	11.0	4.5	28.0
	médio	3.4	7.6	5.0	2.8	8.4	0.5	6.2	0.0	6.8	12.5	16.0	12.4	40.9
	baixo	0.0	11.4	4.0	⊗	⊗	⊗	0.0	12.6	0.2	0.0	24.0	4.2	28.2
	total				9.1	8.4	3.0	6.3	12.6	7.0	25.0	51.0	21.0	97.0
MNAR4	alto	7.0	9.0	2.0	0.0	9.0	0.0	⊗	⊗	⊗	7.0	18.0	2.0	27.0
	médio	4.4	9.0	3.6	8.2	0.0	4.5	5.5	0.0	6.8	18.0	9.0	15.0	42.0
	baixo	0.0	12.0	4.0	⊗	⊗	⊗	0.0	12.0	0.0	0.0	24.0	4.0	28.0
	total				8.2	9.0	4.5	5.5	12.0	6.8	25.0	51.0	21.0	97.0

*Alguns totais não coincidem com as somas das frequências respectivas devido aos arredondamentos.

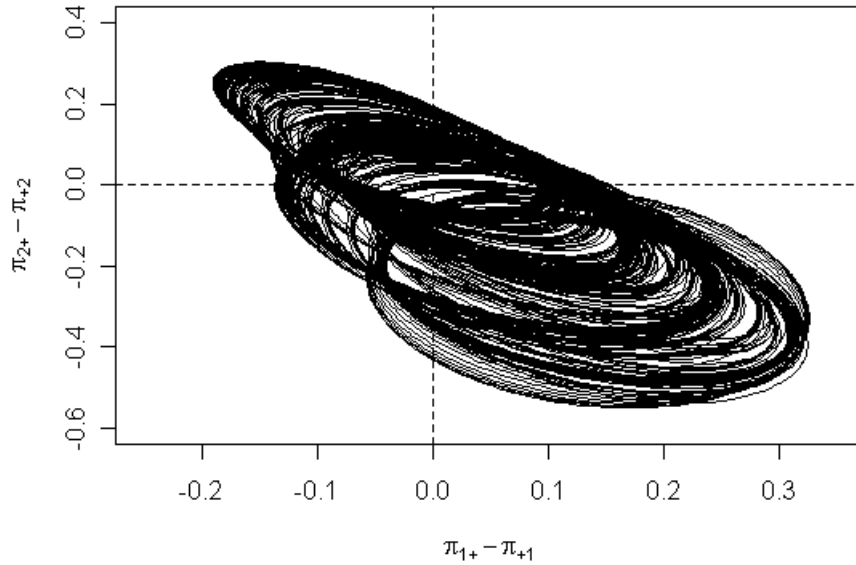
⊗ representa zeros estruturais, simpl.: simplificado, conv.: convencional.

Tabela 3.15: Estimativas, erros padrões e valores-p dos testes de Wald dos parâmetros de interesse

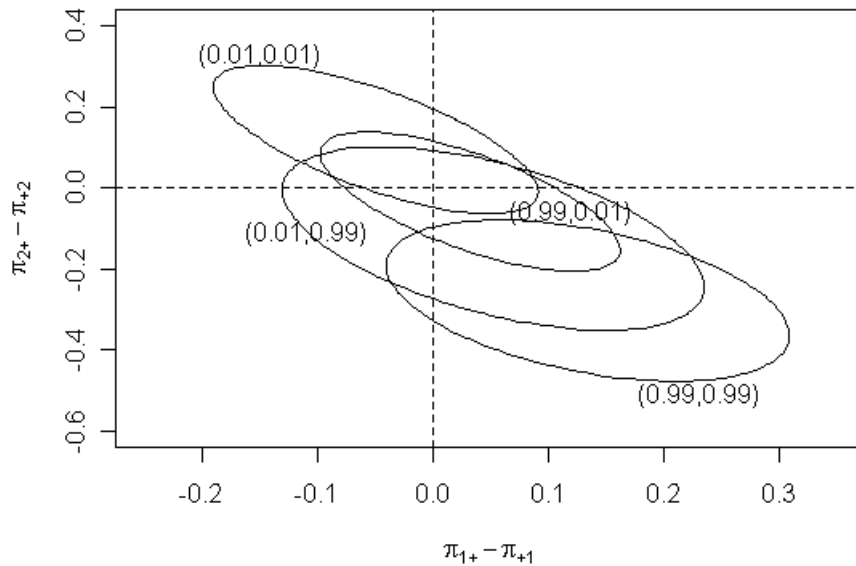
	Quantidade	ACC	MAR _{red} /MAR	MNAR1	MNAR2	MNAR3	MNAR4
estimativas	$\pi_{1+} - \pi_{+1}$	0.196	0.016	0.033	0.118	0.031	0.021
	$\pi_{2+} - \pi_{+2}$	-0.255	-0.031	-0.059	-0.046	-0.105	-0.093
	κ	0.090	0.017	-0.063	0.089	-0.034	-0.244
	κ_{w1}	0.330	0.297	0.218	0.279	0.239	0.179
	κ_{w2}	0.197	0.140	0.062	0.173	0.089	-0.054
erros padrões	$\pi_{1+} - \pi_{+1}$	0.074	0.062	0.073	0.147	0.104	0.158
	$\pi_{2+} - \pi_{+2}$	0.099	0.087	0.108	0.167	0.107	0.265
	κ	0.100	0.102	0.163	0.091	0.130	0.485
	κ_{w1}	0.106	0.090	0.137	0.114	0.217	0.390
	κ_{w2}	0.097	0.092	0.144	0.076	0.150	0.449
valores-p sob	$\pi_{i+} - \pi_{+i} = 0$	0.014	0.938	0.860	0.656	0.582	0.789
	$\kappa = 0$	0.368	0.866	0.699	0.330	0.794	0.615
	$\kappa_{w1} = 0$	0.002	0.001	0.111	0.014	0.269	0.646
	$\kappa_{w2} = 0$	0.042	0.127	0.669	0.023	0.553	0.905

Considere agora a estrutura sobre-saturada para o mecanismo de omissão da Tabela 3.16, em que $\boldsymbol{\tau} = (\alpha_7, \alpha_8)'$ denota o parâmetro de sensibilidade e os demais parâmetros do mecanismo de omissão e categorização são englobados no parâmetro estimável. α_7 é a probabilidade condicional de omissão correspondente às concordâncias entre os métodos para os níveis extremos de suscetibilidade à cárie. α_8 é a probabilidade condicional de omissão associada às discordâncias entre os métodos quando o método simplificado possui graus extremos de suscetibilidade à cárie. Desta forma, quanto maior for α_7 e menor for α_8 , mais acentuada será a concordância entre os métodos.

A Figura 3.5 ilustra as regiões de 95% de confiança para $(\pi_{1+} - \pi_{+1}, \pi_{2+} - \pi_{+2})$ variando (α_7, α_8) . Em (a), α_7 e α_8 assumem valores $e^a/(1 + e^a)$, com a de -5 a 5 em incrementos de 0.5 , enquanto que em (b), apresenta-se casos extremos atribuindo os valores 0.01 e 0.99 para α_7 e α_8 . A maior parte das regiões de confiança contém o valor $(\pi_{1+} - \pi_{+1}, \pi_{2+} - \pi_{+2}) = (0, 0)$, caso em que a homogeneidade marginal não é rejeitada, embora isso não aconteça a medida que (α_7, α_8) tendem a $(1, 1)$. O intervalo de ignorância para $\pi_{1+} - \pi_{+1}$ é $(-0.050; 0.134)$ e o



(a) α_7 e α_8 assumindo valores $\frac{e^a}{1+e^a}$, com $a = -5, -4.5, \dots, 5$



(b) Ilustração dos casos extremos atribuindo os valores 0.01 e 0.99 para α_7 e α_8

Figura 3.5: Regiões de 95% de confiança para $(\pi_{1+} - \pi_{+1}, \pi_{2+} - \pi_{+2})$ variando (α_7, α_8)

Tabela 3.16: Estrutura sobre-saturada para as probabilidades condicionais de omissão $\{\lambda_{t(ij)}\}$

Métodos simpl.\conv.	$t = 2$			$t = 3$		
	alto	médio	baixo	alto	médio	baixo
alto	α_7	α_8	α_8	0	0	0
médio	α_1	α_2	α_3	α_4	α_5	α_6
baixo	0	0	0	α_8	α_8	α_7

$\lambda_{1(ij)} = 1 - \lambda_{2(ij)} - \lambda_{3(ij)}$, simpl.: simplificado, conv.: convencional.

para $\pi_{2+} - \pi_{+2}$ é $(-0.278; 0.121)$. Note que estes intervalos são parecidos com os intervalos para o melhor-pior caso da Tabela 3.11, embora os últimos tenham comprimentos maiores.

Ainda variando α_7 e α_8 como foi descrito na Figura 3.5a, a Figura 3.6 apresenta as EMV e intervalos de 95% de confiança para $\kappa(\alpha_7, \alpha_8)$, $\kappa_{w1}(\alpha_7, \alpha_8)$ e $\kappa_{w2}(\alpha_7, \alpha_8)$. Nota-se algumas mudanças bruscas entre estimativas dos parâmetros para (α_7, α_8) vizinhos, casos em que simultaneamente se observa um aumento desproporcional no comprimento dos intervalos de confiança. Os intervalos de ignorância e de 95% de incerteza obtidos para κ são $(-0.226; 0.315)$ e $(-0.393; 0.571)$, para κ_{w1} , $(0.115; 0.501)$ e $(-0.069; 0.755)$, e para κ_{w2} , $(-0.072; 0.397)$ e $(-0.212; 0.635)$.

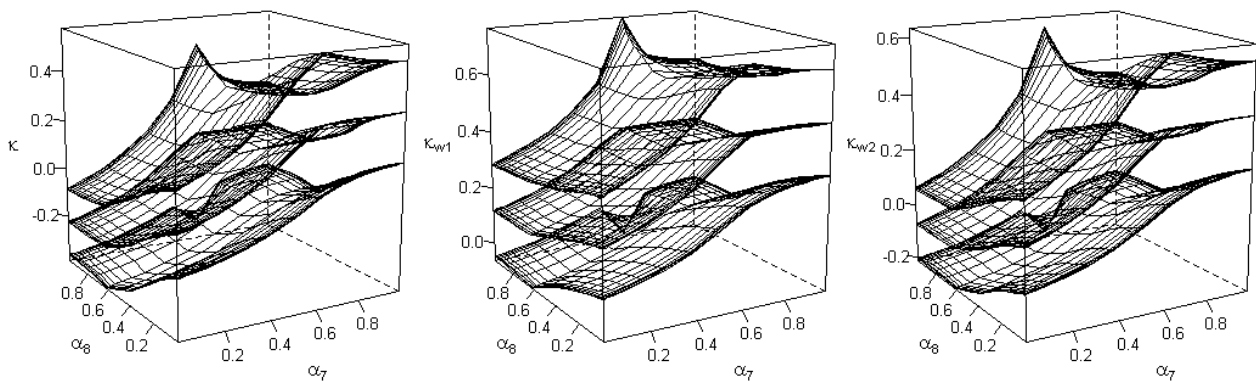


Figura 3.6: EMV e intervalos de 95% de confiança para κ , κ_{w1} e κ_{w2} variando (α_7, α_8)

As análises de sensibilidade demonstram que não se pode concluir a favor da (ou contra a) homogeneidade marginal ou de uma concordância baixa ou alta entre os métodos sem maiores informações sobre o mecanismo de omissão.



Exemplo 4

Em 2004, o resultado do 1º turno das eleições municipais da cidade de São Paulo, realizadas em 3 de outubro, gerou polêmica⁸ ao se constatar que José Serra venceu Marta Suplicy por uma diferença absoluta de 7.7% na proporção de votos válidos, enquanto que o resultado da pesquisa de intenção de voto divulgado no dia anterior às eleições pelo Datafolha Instituto de Pesquisas apontava que a diferença era igual a 3% e, mais distante ainda do resultado almejado, o Instituto Brasileiro de Opinião Pública e Estatística (IBOPE), também no dia 2 de outubro, indicou Marta com 2% de votos a mais do que Serra. Até mesmo na pesquisa de boca-de-urna, o IBOPE continuou bem distante do resultado oficial, estimando que ambos os candidatos estariam empatados com 40% dos votos. A Tabela 3.17 resume os resultados⁹ apresentados pelo Datafolha, IBOPE e Tribunal Regional Eleitoral (TRE).

Tabela 3.17: Comparação entre os resultados do Datafolha, IBOPE e TRE

Entidade	Datafolha	IBOPE	IBOPE	TRE
Dia	02	02	03	03
José Serra	40%	38%	40%	43.6%
Marta Suplicy	37%	40%	40%	35.8%
Paulo Maluf	11%	12%	10%	11.9%
Demais 11 candidatos	13%	10%	10%	8.7%
Margem de erro máxima	2%	3%	1.7%	
Nível de confiança	95%	95%	99%	
Nº de entrevistados	4 963	1 204	6 000	
Indecisos, brancos e nulos	8%	10%	5%	

Nenhum dos intervalos de confiança para a diferença entre as proporções de votos de Serra (π_S) e Marta (π_M) das pesquisas realizadas contém o resultado obtido pelo TRE, uma vez

⁸Vide, por exemplo, as reportagens do jornal Folha de São Paulo — “Datafolha detecta onda de crescimento de Serra” (da reportagem local, 04/10/2004, p.Especial-8), “Vitória de Serra se consolidou no domingo, diz Datafolha” (da redação, 05/10/2004, p.Especial-3) e “Na boca da urna” (de M.Beraba, 10/10/2004, p. A-6) —, jornal O Estado de São Paulo — “Para diretora, IBOPE não errou em boca-de-urna” (de R.Tavares e V.Freire, 04/10/2004, <http://www.estadao.com.br>) — e revista Veja — “Guerra do calendário” (M.Weinberg e C.Menezes, 13/10/2004, pp.46-47).

⁹Pesquisas registradas no Tribunal Regional Eleitoral de São Paulo sob números 005400104-SPPE (Datafolha), 005200104-SPPE e 005300104-SPPE (IBOPE). Os resultados estão disponíveis em <http://www.datafolha.com.br> e <http://www.ibope.com.br>.

que, mesmo construídos de maneira conservadora¹⁰ a partir da Tabela 3.17, são iguais¹¹ a $(-1; 7)\%$, $(-8; 4)\%$ e $(-3.4; 3.4)\%$.

Diversas hipóteses podem ser formuladas para explicar o motivo de as inferências por intervalo das três pesquisas não conterem o resultado final. A mais simples e inócua é a de variação amostral; afinal, de acordo com a teoria freqüentista, espera-se que cerca de 1 em 20 (100) intervalos de 95% (99%) de confiança não contenham o valor do parâmetro de interesse. Suspeita-se também que os eleitores possam ter mudado suas intenções de voto, já que o trabalho de campo das pesquisas divulgadas no dia 2 foi iniciado no dia anterior e no mesmo dia 1, à noite, houve um debate entre os candidatos, organizado e televisionado pela Globo, em que o desempenho de Serra foi apontado como o melhor, de acordo com uma pesquisa realizada pelo Datafolha, e poderia ter repercutido na opinião da população lentamente até o momento da votação. Além disso, desconfia-se de que o resultado dos institutos tenha sido tão distante do obtido pelo TRE, possivelmente, devido ao viés de seleção da amostra, erros de classificação e/ou viés de não-resposta. Os dois últimos, por exemplo, poderiam ter ocorrido com pessoas que queriam se ver livres do entrevistador ou preferiam não expressar sua intenção de voto, atitudes que enviesariam os estimadores se fossem realizadas com uma maior freqüência por entrevistados com intenção de votar em alguns dos candidatos. Por outro lado, o viés de seleção pode ter ocorrido por causa de se ter utilizado a amostragem por quotas¹² (veja a

¹⁰Apesar de não se conhecer o efeito do planejamento amostral das pesquisas, sabe-se que

$$Var(\hat{\pi}_S - \hat{\pi}_M) = Var(\hat{\pi}_S) + Var(\hat{\pi}_M) - 2 Corr(\hat{\pi}_S, \hat{\pi}_M) DP(\hat{\pi}_S) DP(\hat{\pi}_M)$$

atinge seu valor máximo se $Corr(\hat{\pi}_S, \hat{\pi}_M) = -1$. Como $DP(\hat{\pi}_S)$ e $DP(\hat{\pi}_M)$ são menores que $B/z_{1-\alpha/2}$, em que B representa a margem de erro máxima e $z_{1-\alpha/2}$ denota o quantil $1 - \alpha/2$ da distribuição normal padrão, então

$$Var(\hat{\pi}_S - \hat{\pi}_M) \leq Var(\hat{\pi}_S) + Var(\hat{\pi}_M) + 2DP(\hat{\pi}_S) DP(\hat{\pi}_M) \leq \frac{B^2}{z_{1-\alpha/2}^2} + \frac{B^2}{z_{1-\alpha/2}^2} + 2 \frac{B}{z_{1-\alpha/2}} \frac{B}{z_{1-\alpha/2}} = 4 \frac{B^2}{z_{1-\alpha/2}^2}$$

e, conseqüentemente, $\hat{\pi}_S - \hat{\pi}_M \pm 2B$ pode ser utilizado como intervalo de confiança para $\pi_S - \pi_M$ com coeficiente de confiança assintótico de, pelo menos, $1 - \alpha$.

¹¹Esses intervalos foram construídos a partir dos percentuais arredondados, sem casas decimais (com exceção da margem de erro do IBOPE do dia 3), refletidos na Tabela 3.17 da mesma forma como foram divulgados pelos institutos de pesquisa.

¹²Na amostragem por quotas, trabalhadores de campo são designados para entrevistar um número pré-fixado de pessoas de determinadas subpopulações proporcionalmente aos seus tamanhos, tal como na amostragem estratificada com alocação proporcional. A diferença é que enquanto na última a seleção dos entrevistados é feita de acordo com alguma distribuição de probabilidades hipoteticamente conhecida — usualmente, supõe-se probabilidades de seleção iguais para todos os indivíduos —, na primeira o critério de seleção depende do entrevistador e de a pessoa transitar (ou não) pelo local em que se está entrevistando.

descrição completa do esquema amostral a seguir), tanto por causa da seleção, obviamente subjetiva, realizada pelo entrevistador, quanto devido à distribuição das intenções de voto dos eleitores que passam pelos pontos de fluxo, escolhidos pelos institutos como locais das entrevistas, potencialmente ser diferente da correspondente distribuição dos eleitores que não passam (*e.g.*, pessoas que se locomovam apenas por meio de automóvel ou que permaneçam predominantemente em suas residências), casos em que a amostra obtida não constitui uma amostra aleatória da população de interesse. Para uma discussão mais abrangente de erros amostrais e não-amostrais, veja, por exemplo, Kish (1965), Cochran (1977), Särndal, Swensson, e Wretman (1991) ou Bolfarine e Bussab (2005). Algumas dessas e outras questões envolvendo especificamente pesquisas de intenção de voto podem ser encontradas em Pereira e Wechsler (1994) e Almeida (2002).

Dentre todas as especulações, avalia-se apenas o possível viés de não-resposta, em que se propõe uma análise para os dados da pesquisa de intenção de voto divulgada pelo IBOPE no dia 2, assumindo que todos os entrevistados classificados como indecisos, brancos e nulos são não-respondentes¹³. Naturalmente, no dia da eleição algumas pessoas votam em branco, nulo ou mesmo não comparecem para votar. No entanto, como a decisão sobre o candidato vencedor, ou sobre os dois candidatos que irão para o segundo turno, baseia-se apenas nos votos válidos¹⁴, opta-se por considerar todas as intenções de votos não-válidos como dados omissos.

Na Tabela 3.18, apresenta-se uma comparação mais detalhada entre os resultados obtidos pelo IBOPE e TRE.

A pesquisa foi realizada com eleitores de 16 anos ou mais em esquema amostral de dois estágios. No primeiro estágio, os setores censitários¹⁵ foram selecionados com probabilidade proporcional ao número de eleitores. No segundo estágio, a seleção dos entrevistados foi realizada utilizando-se quotas proporcionais às populações associadas às combinações dos níveis das variáveis sexo (masculino, feminino), faixa etária (16 a 17, 18 a 24, 25 a 29, 30 a 39, 40 a 49, 50 a 69, 70 anos ou mais), instrução (até a 4ª série do ensino fundamental, 5ª a 8ª

¹³Dessa forma, os percentuais de votos válidos, geralmente analisados, são uma ACC.

¹⁴Com exceção de quando há mais de 50% de votos nulos, caso em que uma nova eleição é convocada.

¹⁵“Os setores censitários são demarcados pelo Instituto Brasileiro de Geografia e Estatística (IBGE), obedecendo a critérios de operacionalização da coleta de dados, de tal maneira que abranjam uma área que possa ser percorrida por um único entrevistador em um mês e que possua em torno de 250 a 350 domicílios (em áreas urbanas).” (Fonte: Fiocruz, <http://www.sig.cict.fiocruz.br/setorescensitarios.htm>). O município de São Paulo possui 13,193 setores censitários (Censo, 2000).

Tabela 3.18: Comparação entre os resultados do IBOPE e TRE

Candidato – nº (partido)	IBOPE (02/10/2004)				TRE (03/10/2004)			
	Intenções	%Tot.	%Resp.	%Vál.	Votos	%Tot.	%Resp.	%Vál.
José Serra – 45 (PSDB)	414	34.4	35.3	38.2	2 686 396	34.6	40.6	43.6
Marta Suplicy – 13 (PT)	437	36.3	37.3	40.3	2 209 264	28.4	33.4	35.8
Paulo Maluf – 11 (PP)	127	10.5	10.8	11.7	734 580	9.5	11.1	11.9
Luiza Erundina – 40 (PSB)	53	4.4	4.5	4.9	244 090	3.1	3.7	4.0
Paulinho – 12 (PDT)	14	1.2	1.2	1.3	86 549	1.1	1.3	1.4
Francisco Rossi – 31 (PHS)	19	1.6	1.6	1.8	77 957	1.0	1.2	1.3
Dra. Havanir – 56 (PRONA)	7	0.6	0.6	0.6	47 579	0.6	0.7	0.8
Penna – 43 (PV)	7	0.6	0.6	0.6	43 868	0.6	0.7	0.7
Osmar Lins – 26 (PAN)	2	0.2	0.2	0.2	16 339	0.2	0.2	0.3
Dirceu Travesso – 16 (PSTU)	1	0.1	0.1	0.1	8 394	0.1	0.1	0.1
Ciro – 36 (PTC)	2	0.2	0.2	0.2	6 111	0.1	0.1	0.1
Prof. Walter Canoas – 21 (PCB)	2	0.2	0.2	0.2	3 138	0.0	0.0	0.1
João Manuel – 27 (PSDC)	0	0.0	0.0	0.0	1 627	0.0	0.0	0.0
Anaí Caproni – 29 (PCO)	0	0.0	0.0	0.0	1 479	0.0	0.0	0.0
Válidos	1 085	90.1	92.5	100.0	6 167 371	79.4	93.3	100.0
Branco e nulos	88	7.3	7.5		442 012	5.7	6.7	
Respondentes	1 173	97.4	100.0		6 609 383	85.0	100.0	
Não respondeu/compareceu	31	2.6			1 162 120	15.0		
Total	1 204	100.0			7 771 503	100.0		

série do ensino fundamental, ensino médio, ensino superior) e setor de atividade (agricultura, indústria de transformação, indústria de construção, outras indústrias, comércio, prestação de serviços, transporte e comunicação, atividade social, administração pública, outras atividades, estudantes, inativos).

Como a amostragem por quotas é não-probabilística¹⁶, não se pode aplicar fórmulas de erro amostral com segurança. Os institutos de pesquisa comumente assumem que a amostra obtida provém de uma amostragem aleatória simples, embora, conforme se discutiu previamente, não se acredita que esta suposição seja razoável. Note que as margens de erro máximas considerando níveis de 95% e 99% de confiança, calculadas por meio da distribuição binomial (ou

¹⁶Veja Kish (1965, pp.562-566) e Stephan e McCarthy (1958, Capítulos 10 e 13) para mais discussões e críticas sobre as amostragens por quotas.

seja, a distribuição marginal univariada relativa à distribuição multinomial) para os tamanhos de amostra das pesquisas de intenção de voto dos dias 2 e 3 realizadas pelo IBOPE são de, respectivamente,

$$1.96\sqrt{\frac{0.5(1-0.5)}{1\,204}} \cong 2.8\% \quad \text{e} \quad 2.58\sqrt{\frac{0.5(1-0.5)}{6\,000}} \cong 1.7\%,$$

que são aproximadamente iguais aos 3% e 2% divulgados. O mesmo não ocorre para a margem de erro máxima de 2%, divulgada pelo Datafolha, que difere do arredondamento de $1.96\sqrt{\frac{0.5(1-0.5)}{4\,963}} \cong 1.4\%$. Talvez o Datafolha utilize alguma das seguintes abordagens: (i) arredonde o valor para o número inteiro subsequente independentemente das casas decimais; (ii) inflacione a variância em 50%, fator obtido por Stephan e McCarthy (1958, p.233) em seus estudos empíricos como aproximação das diferenças relativas entre os valores obtidos nas tentativas de se estimar a variância em procedimentos amostrais por quotas em relação às correspondentes estimativas obtidas de amostragens aleatórias simples; (iii) calcule os erros máximos considerando os setores censitários como conglomerados e as quotas como estratos, prática que é mais adequada do que se supor uma amostragem aleatória simples por não ignorar a possível dependência entre as opiniões de eleitores de um mesmo setor censitário.

O IBOPE não disponibilizou os pesos ou probabilidades de inclusão no banco de dados¹⁷ da pesquisa do dia 2 e, aparentemente, nem chega a calculá-los, divulgando a margem de erro máxima das proporções calculada previamente às entrevistas e deixando que as proporções se autoponderem, já que as unidades amostrais dos dois estágios são selecionadas proporcionalmente às suas respectivas populações. A exclusão dos pesos é claramente inapropriada, uma vez que pode levar a flutuações desnecessárias nas estimativas das proporções populacionais devido aos arredondamentos dos tamanhos de amostra calculados para as quotas. Porém, ainda que se utilize as mesmas fontes de dados que o IBOPE¹⁸, não seria possível reconstruir os pesos, pois no banco de dados também não estão disponíveis os setores censitários e nem os setores de atividade dos entrevistados. Portanto, a análise será realizada sem a utilização dos pesos, assumindo uma amostragem aleatória simples, embora se julgue mais apropriado que o planejamento amostral fosse incorporado conforme descrito na abordagem (iii) do parágrafo

¹⁷Fonte: Banco de Dados do CESOP – Centro de Estudos de Opinião Pública, UNICAMP (IBO/SPcap04. OUT-02302).

¹⁸Dados do Censo (2000), Pesquisa Nacional por Amostra de Domicílios (PNAD, 2002) e Tribunal Superior Eleitoral (2002).

anterior¹⁹. Conseqüentemente, os erros padrões podem estar incorretos e os resultados dos testes estatísticos devem ser analisados com cautela.

Um intervalo para o melhor-pior caso para $\pi_S - \pi_M$ é²⁰ $(-11.8; 8.0)\%$, demonstrando que se a maior parte dos entrevistados com omissão votaram em Serra, isso já seria suficiente para, possivelmente, explicar a diferença entre os resultados do IBOPE e do TRE.

Além das variáveis sexo, faixa etária e instrução, também constavam do banco de dados: região (norte, sul, leste, centro-oeste), grau de satisfação com a vida²¹ (muito satisfeito, satisfeito, insatisfeito, muito insatisfeito), renda familiar do mês anterior (até 1 salário mínimo (s.m.), 1 a 2 s.m., 2 a 5 s.m., 5 a 10 s.m., 10 a 20 s.m., mais de 20 s.m.) e candidatos rejeitados — resultado de pergunta de múltipla escolha em que o entrevistado mencionava todos os candidatos em que não votaria “de jeito nenhum”. Tabelas de contingência foram construídas cruzando essas variáveis, individualmente e conjuntamente, com a intenção de voto e $\pi_S - \pi_M$ foi estimado sob o mecanismo MAR em cada caso. Esperava-se que a omissão da intenção de voto poderia depender apenas do que foi observado para as outras variáveis e isso bastaria para corrigir as estimativas pontuais, o que não ocorreu, a julgar pelas EMV de $\pi_S - \pi_M$ que variaram de -2.4% a -1.9% , dependendo da análise.

Mecanismos MNAR, construídos utilizando a idéia de estimadores protetores (Seção 1.2.3 e Exemplos 2 e 3), foram ajustados para tabelas de contingência resultantes do cruzamento da intenção de voto agrupada (Serra, Marta, Maluf, outros 11 candidatos) separadamente com cada uma das variáveis instrução e grau de satisfação com a vida. As EMV para $\pi_S - \pi_M$ foram de, respectivamente, -1.9% e -2.4% , novamente não estando próximas do resultado do TRE. Ambas as estruturas não parecem adequadas, uma vez que alocam quase todos os entrevistados com omissão para a intenção de voto em Maluf, na análise da variável instrução,

¹⁹A teoria apresentada no Capítulo 2 pode ser adaptada para incorporar o planejamento amostral analogamente ao que Chambers e Skinner (2003) e Lehtonen e Pahkinen (2004) ilustram, de maneira geral, para a distribuição multinomial considerando modelos log-lineares e de regressão logística, em que, por exemplo, ao aplicar os pesos na verossimilhança, obtém-se uma quase-verossimilhança, e as distribuições das estatísticas de teste tomam a forma de uma soma de distribuições qui-quadrados independentes, com 1 grau de liberdade cada, ponderadas com pesos obtidos de autovalores de uma matriz geral de efeito do esquema amostral.

²⁰Veja a coluna %Tot. (percentual em relação ao total) do IBOPE da Tabela 3.18 e note que este intervalo é construído primeiro supondo que todos os resultados com omissão ($7.3\% + 2.6\% = 9.9\%$) são de eleitores com intenção de votar em Marta ($34.4\% - [36.3\% + 9.9\%]$) e, a seguir, em Serra ($[34.4\% + 9.9\%] - 36.3\%$).

²¹Como Marta era a atual prefeita da cidade de São Paulo na época e estava buscando a reeleição, eleitores satisfeitos tendem a preferi-la e *vice-versa*.

e para a intenção de voto nos outros candidatos, na análise da variável grau de satisfação com a vida. Não se estudou outras estruturas MNAR utilizando mais variáveis conjuntamente, pois à medida que a dimensão das tabelas de contingência cresce, a grande quantidade de zeros amostrais gera instabilidades no procedimento de estimação e a proposição das estruturas também é dificultada.

Mesmo se todas as omissões fossem de eleitores com intenção de votar em Maluf, ainda assim ele atingiria apenas 20.4% das intenções de voto, insuficientes para fazer frente a Serra ou Marta. Portanto, para simplificar a análise de sensibilidade com estruturas sobre-saturadas MNAR, estima-se apenas as proporções de votos de Serra, Marta e dos outros 12 candidatos agrupados. Ao mesmo tempo, desconsidera-se outras variáveis, em primeiro lugar, porque não há questões de interesse relacionadas, e em segundo, por parcimônia. Sejam α_S , α_M e α_O as probabilidades de eleitores com intenção de votar em Serra, Marta e outros candidatos, respectivamente, de não expressarem suas vontades na pesquisa do IBOPE. Ao assumir que $\mathbf{N} = (n_S, n_M, n_O, n_N)'$, o vetor com as frequências de entrevistados com intenção de votar em Serra, Marta, outros candidatos e os não-respondentes, segue uma distribuição multinomial, tem-se a função de verossimilhança sob o mecanismo MNAR sobre-parametrizado supracitado

$$L(\boldsymbol{\mu}, \boldsymbol{\tau} | \mathbf{N}) \propto \left(\pi_S[1 - \alpha_S]\right)^{n_S} \left(\pi_M[1 - \alpha_M]\right)^{n_M} \left([1 - \pi_S - \pi_M][1 - \alpha_O]\right)^{n_O} \times \\ \left(\pi_S \alpha_S + \pi_M \alpha_M + [1 - \pi_S - \pi_M]\alpha_O\right)^{n_N},$$

em que $\boldsymbol{\mu} = (\pi_S, \pi_M, \alpha_O)'$ denota o parâmetro estimável e $\boldsymbol{\tau} = (\alpha_S, \alpha_M)'$, o parâmetro de sensibilidade. Uma vez que 9.9% das intenções de voto dos entrevistados estão omissas, varia-se α_S e α_M de 5% a 20% em incrementos de 1%. Note que o mecanismo MCAR, indicando a ausência do viés de não-resposta, é um caso particular da estrutura sobre-saturada quando $\alpha_S = \alpha_M = \alpha_O$. As EMV e intervalos de 95% de confiança para $\pi_S - \pi_M$ em função de α_S e α_M são apresentados na Figura 3.7.

Como esperado, à medida que α_S cresce e α_M diminui, mais entrevistados são alocados para Serra e menos, para Marta, fazendo com que a estimativa de $\pi_S - \pi_M$ aumente. O intervalo de ignorância para $\pi_S - \pi_M$ é $(-8.5; 4.4)\%$, e o correspondente intervalo de 95% de incerteza, $(-13.9; 9.8)\%$, que já contém o resultado do TRE, 7.7%. Por exemplo, se $\alpha_S = 16.0\%$ e $\alpha_M = 5.0\%$, o mecanismo MNAR em consideração tem um ajuste perfeito, com $\hat{\pi}_S(\boldsymbol{\tau}) = 40.9\%$, $\hat{\pi}_M(\boldsymbol{\tau}) = 38.2\%$ e $\hat{\alpha}_O(\boldsymbol{\tau}) = 6.8\%$, caso em que o intervalo de 95% de confiança para $\pi_S - \pi_M$ é de $(-2.6; 8.1)\%$. Se houver suspeitas que o viés de não-resposta seja tal que α_S e α_M tenham valores mais extremos no espaço paramétrico, pode-se ainda ampliar mais a

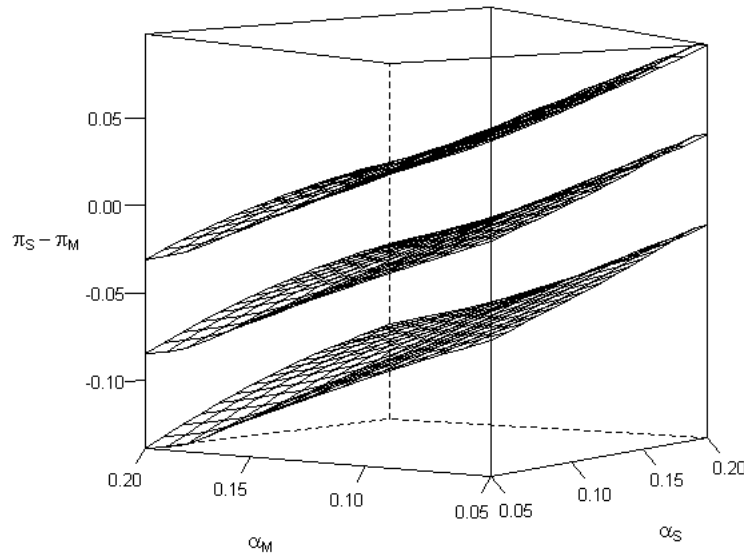


Figura 3.7: EMV e intervalos de 95% de confiança para $\pi_S - \pi_M$ variando (α_S, α_M)

faixa de variação do parâmetro de sensibilidade, permitindo que assumam valores $e^a/(1+e^a)$, com a variando de -5 a 5 em incrementos de 0.5 , que alarga os intervalos de ignorância e de 95% de incerteza para, respectivamente, $(-11.7; 7.9)\%$ e $(-16.8; 13.0)\%$. Note que o intervalo de ignorância praticamente coincide com o intervalo para o melhor-pior caso, o que teria acontecido exatamente se também fossem considerados α_S e α_M iguais a 0 e 1 .

Ludicamente, no sentido de um *post-mortem*, pois tal análise só é possível depois que a eleição já ocorreu, pode-se considerar as proporções de votos válidos do TRE como parâmetros populacionais para as intenções de voto da pesquisa do IBOPE (*i.e.*, π_S e π_M passam a ser considerados como conhecidos) e, assim, estimar as probabilidades condicionais de omissão dos candidatos com base nos dados da pesquisa de intenção de voto do IBOPE do dia 2. As estimativas (erros padrões) para α_S , α_M e α_O são 20.5% (3.2%), 0.0% (3.8%) e 5.0% (5.6%), em que a estrutura não tem um ajuste perfeito, com correspondente estatística de razão de verossimilhanças igual a 0.12 apesar de se ter zero graus de liberdade. A presença do viés de não-resposta pode ser verificada por meio do teste Wald para a hipótese $H : \alpha_S = \alpha_M = \alpha_O$ ($p=0.002$).

Os intervalos para o melhor-pior caso, de ignorância e de incerteza protegem as inferências pontual e por intervalo de possíveis vieses devido às omissões e ao desconhecimento de seu mecanismo ao custo de aumentar consideravelmente suas amplitudes. Sendo assim, é justo comparar a situação vista, em que talvez tenha ocorrido um viés de não-resposta, a uma outra

contrária, que se apresenta a seguir.

Continuando o acompanhamento das eleições municipais de 2004 da cidade de São Paulo, mas agora no 2º turno, a Tabela 3.19 apresenta os resultados das pesquisas²² de intenção de voto do Datafolha e do IBOPE, anunciados no dia 30 de outubro, da pesquisa de boca de urna do IBOPE além do resultado oficial do TRE, divulgados no dia seguinte.

Tabela 3.19: Comparação entre os resultados do Datafolha, IBOPE e TRE

Entidade	Datafolha	IBOPE	IBOPE	TRE
Dia	30	30	31	31
José Serra	54%	54%	55%	54.9%
Marta Suplicy	46%	46%	45%	45.1%
Margem de erro máxima	2%	2%	1.7%	
Nível de confiança	95%	95%	95%	
Nº de entrevistados	6 470	2 002	6 000	
Indecisos, brancos e nulos	7%	10%	6%	

Desta vez, nota-se a semelhança dos resultados das pesquisas em relação aos anunciados pelo TRE, indicando que, possivelmente, não houve nem viés de não-resposta nem os outros problemas discutidos previamente. É importante salientar que, igualmente ao 1º turno, as pesquisas do Datafolha e do IBOPE divulgadas no dia 30 tiveram seus trabalhos de campo iniciados no dia anterior e no mesmo dia 29, à noite, outro debate entre os candidatos foi realizado e televisionado pela Globo.

Por consistência, analisa-se novamente os resultados da pesquisa de intenção de voto do IBOPE divulgados no dia anterior à eleição. Na Tabela 3.20, apresenta-se uma comparação mais detalhada entre os resultados obtidos pelo IBOPE²³ e TRE.

Como nesse momento há apenas dois candidatos, $\pi_S = 1 - \pi_M$ e, conseqüentemente, $Corr(\hat{\pi}_S, \hat{\pi}_M) = -1$, então basta analisar π_S ou π_M , ao invés de $\pi_S - \pi_M$. O intervalo para

²²Pesquisas registradas no Tribunal Regional Eleitoral de São Paulo sob números 007800104-SPPE (Datafolha), 007400104-SPPE e 007600104-SPPE (IBOPE). Os resultados estão disponíveis em <http://www.datafolha.com.br> e <http://www.ibope.com.br>.

²³Fonte: Banco de Dados do CESOP – Centro de Estudos de Opinião Pública, UNICAMP (IBO/SPcap04. OUT-02328).

Tabela 3.20: Comparação entre os resultados do IBOPE e TRE

Candidato – nº (partido)	IBOPE (30/10/2004)				TRE (31/10/2004)			
	Intenções	%Tot.	%Resp.	%Vál.	Votos	%Tot.	%Resp.	%Vál.
José Serra – 45 (PSDB)	980	49.0	52.1	54.4	3 330 179	42.9	52.0	54.9
Marta Suplicy – 13 (PT)	821	41.0	43.6	45.6	2 740 152	35.3	42.8	45.1
Válidos	1 801	90.0	95.7	100.0	6 070 331	78.1	94.7	100.0
Branco e nulos	80	4.0	4.2		337 138	4.3	5.3	
Respondentes	1 881	94.0	100.0		6 407 469	82.4	100.0	
Não respondeu/compareceu	121	6.0			1 364 034	17.6		
Total	2 002	100.0			7 771 503	100.0		

o melhor-pior caso para π_S é²⁴ (49.0; 59.0)%. A função de verossimilhança para o vetor $\mathbf{N} = (n_S, n_M, n_N)'$, com as frequências observáveis, é explicitada por

$$L(\pi_S, \alpha_S, \alpha_M | \mathbf{N}) \propto \left(\pi_S [1 - \alpha_S]\right)^{n_S} \left([1 - \pi_S][1 - \alpha_M]\right)^{n_M} \left(\pi_S \alpha_S + [1 - \pi_S]\alpha_M\right)^{n_N}.$$

Ao variar o parâmetro de sensibilidade α_S de 0.1% a 99.9%, obtém-se, para π_S , o intervalo de ignorância (49.0; 59.0)% e o intervalo de 95% de incerteza (46.8; 61.6)%. Mesmo considerando α_S de 5.0% a 20.0%, uma vez que 10.0% dos entrevistados têm suas intenções de voto omissas, os intervalos de ignorância e de 95% de incerteza para π_S são reduzidos apenas para (51.5; 59.0)% e (49.2; 61.6)%, que é suficiente para que a previsão de vitória de Serra não seja completamente inquestionável. Essas conclusões podem ser observadas na Figura 3.8, que ilustra as EMV e intervalos de 95% de confiança para π_S em função de α_S .

Percebe-se que π_S cresce linearmente à medida que se aumenta α_S até cerca de $\alpha_S = 17.0\%$, quando π_S estabiliza em 59.0%. No entanto, o intervalo de confiança para π_S continua a se estreitar, indicando que a precisão da estimativa aumenta à medida que α_S tende a 100%. O ajuste do mecanismo de omissão é perfeito apenas para $\alpha_S \leq 17.0\%$. A estatística de razão de verossimilhanças do ajuste da estrutura cresce à medida que α_S se distancia de 17.0%, chegando ao valor de 12 462.07 quando $\alpha_S = 99.9\%$. Os intervalos de 95% de confiança para π_S contêm o valor 50% apenas para $\alpha_S < 6.5\%$, portanto, se houvesse razões para se suspeitar que $\alpha_S \geq 6.5\%$, então ainda assim poder-se-ia inferir que Serra ganharia as eleições.

²⁴Veja a coluna %Tot. (percentual em relação ao total) do IBOPE da Tabela 3.20 e note que este intervalo é construído primeiro supondo que todos os resultados com omissão (4.0% + 6.0% = 10.0%) são de eleitores com intenção de votar em Marta (caso em que π_S seria de apenas 49.0%) e, a seguir, em Serra (49.0% + 10.0%).

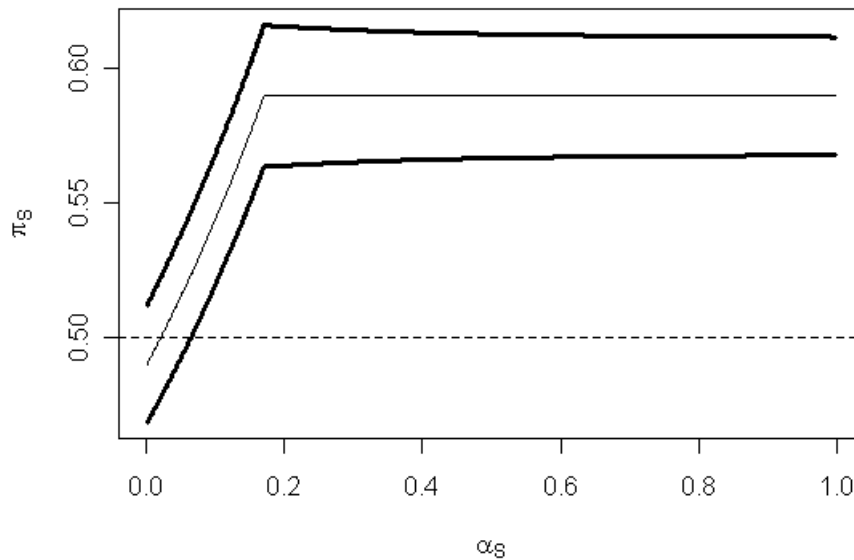


Figura 3.8: EMV e intervalos de 95% de confiança para π_S variando α_S

Ao considerar as proporções de votos válidos do TRE como parâmetros populacionais para as intenções de voto da pesquisa do IBOPE, as decorrentes estimativas (erros padrões) para α_S e α_M são 10.8% (2.0%) e 9.2% (2.4%), verificando-se um ajuste perfeito da estrutura e a não manifestação de evidências a favor de um viés de não-resposta ($p=0.966$ para o teste Wald de $H : \alpha_S = \alpha_M$).

■

Exemplo 5

Em um estudo conduzido na Faculdade de Medicina da Universidade de São Paulo, 219 pacientes foram avaliados por um ou mais métodos não-invasivos (ultra-sonografia, ressonância magnética e ecocolonoscopia retrocervicais) antes de serem submetidos ao procedimento de laparoscopia para diagnosticar endometriose (D). O objetivo do estudo é a estimação da precisão dos testes de diagnóstico, mensurada por parâmetros como sensibilidade, especificidade, valor preditivo positivo e valor preditivo negativo (Soares e Siqueira, 2002; Fleiss *et al.*, 2003), e a comparação destes parâmetros entre os métodos. As frequências dos pacientes com resultado positivo (+) e negativo (-) sob cada método são indicadas na Tabela 3.21.

Tabela 3.21: Frequências observadas dos pacientes

Ultra-sono- grafia (<i>US</i>)	Ressonância magnética (<i>RM</i>)	Ecocolonos- copia (<i>EC</i>)	Endometriose (<i>D</i>)	
			–	+
–	–	–	6	0
		+	1	0
		omisso	51	1
	+	–	0	0
		+	0	0
		omisso	4	1
+	omisso	–	3	1
		+	3	1
		omisso	51	2
	–	–	0	1
		+	0	2
		omisso	0	21
+	+	–	0	1
		+	0	2
		omisso	1	12
	omisso	–	0	4
		+	0	5
		omisso	2	43

Todos os 219 pacientes foram examinados via ultra-sonografia (*US*). Adicionalmente, 91 deles tiveram somente medidas por ressonância magnética (*RM*), 17 foram avaliados apenas via ecocolonoscopia (*EC*), 13 tiveram ambas (*RM* e *EC*) medidas e, para 98 pacientes, nenhuma das duas avaliações estava disponível.

Primeiramente, concentra-se na comparação da *RM* e *EC*. Os dados estão resumidos na Tabela 3.22.

O padrão de dados omissos nesse estudo permite que os testes investigados sejam comparados sob uma das seguintes estratégias:

- uma ACC considerando apenas os pacientes com medidas em ambos os testes (13, neste caso);

Tabela 3.22: Frequências observadas dos pacientes

Ressonância magnética (RM)	Ecocolonos- copia (EC)	Endometriose (D)	
		-	+
-	-	6	1
	+	1	2
	omisso	51	22
+	-	0	1
	+	0	2
	omisso	5	13
omisso	-	3	5
	+	3	6
	omisso	53	45

- uma análise com os pacientes examinados por apenas um dos dois testes ($91 + 17 = 108$); como as medidas nos pacientes avaliados via RM (91) são independentes das realizadas por EC (17), essa abordagem será chamada de *análise de casos independentes* (ACI);
- uma ACI que também inclua os resultados dos pacientes avaliados via ambos os testes ($91 + 13 = 104$ medidas por RM e $17 + 13 = 30$ por EC); como a possível dependência induzida pelas medidas repetidas nos 13 pacientes avaliados por ambos os testes não vai ser considerada por se assumir que se está na presença de duas amostras constituídas por elementos distintos, essa estratégia será referida por *análise de casos assumidos como independentes* (ACAI).

Sob a ACC, assume-se que os dados (completos) seguem uma distribuição multinomial com parâmetros $\pi_{ijk} = P(RM = i, EC = j, D = k)$, $i, j, k = -, +$. Sensibilidade ($Sens_{(l)}$), especificidade ($Espec_{(l)}$), valor preditivo positivo ($VPP_{(l)}$) e valor preditivo negativo ($VPN_{(l)}$) podem ser definidos em termos de $\{\pi_{ijk}\}$, para $l = RM, EC$, por

$$Sens_{(RM)} = P(RM = + | D = +) = \frac{\pi_{+-+} + \pi_{+++}}{\pi_{--+} + \pi_{-++} + \pi_{+-+} + \pi_{+++}},$$

$$Sens_{(EC)} = P(EC = + | D = +) = \frac{\pi_{-++} + \pi_{+++}}{\pi_{--+} + \pi_{-++} + \pi_{+-+} + \pi_{+++}},$$

$$\begin{aligned}
Espec_{(RM)} &= P(RM = -|D = -) = \frac{\pi_{---} + \pi_{-+-}}{\pi_{---} + \pi_{-+-} + \pi_{+--} + \pi_{+++}}, \\
Espec_{(EC)} &= P(EC = -|D = -) = \frac{\pi_{---} + \pi_{+--}}{\pi_{---} + \pi_{-+-} + \pi_{+--} + \pi_{+++}}, \\
VPP_{(RM)} &= P(D = +|RM = +) = \frac{\pi_{+-+} + \pi_{+++}}{\pi_{+--} + \pi_{+-+} + \pi_{+++} + \pi_{++++}}, \\
VPP_{(EC)} &= P(D = +|EC = +) = \frac{\pi_{-++} + \pi_{+++}}{\pi_{-+-} + \pi_{-++} + \pi_{+++} + \pi_{++++}}, \\
VPN_{(RM)} &= P(D = -|RM = -) = \frac{\pi_{---} + \pi_{-+-}}{\pi_{---} + \pi_{-+-} + \pi_{-+-} + \pi_{-+-}}, \\
VPN_{(EC)} &= P(D = -|EC = -) = \frac{\pi_{---} + \pi_{+--}}{\pi_{---} + \pi_{-+-} + \pi_{+--} + \pi_{+--}}.
\end{aligned}$$

Esses parâmetros, expressos na forma $\mathbf{F}(\boldsymbol{\pi}) = \mathbf{exp}[\mathbf{A}_2 \ln(\mathbf{A}_1 \boldsymbol{\pi})]$, com matrizes \mathbf{A}_1 e \mathbf{A}_2 apropriadas, podem ser estimados por MQG conforme descrito na Seção 2.4.3. Para as sensibilidades e especificidades tem-se

$$\mathbf{F}(\boldsymbol{\pi}) = (Sens_{(RM)}, Sens_{(EC)}, Espec_{(RM)}, Espec_{(EC)})',$$

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \quad \text{e} \quad \mathbf{A}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix},$$

e para os valores preditivos positivo e negativo tem-se

$$\mathbf{F}(\boldsymbol{\pi}) = (VPP_{(RM)}, VPP_{(EC)}, VPN_{(RM)}, VPN_{(EC)})',$$

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix} \quad \text{e} \quad \mathbf{A}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix},$$

em que $\boldsymbol{\pi} = (\pi_{---}, \pi_{-+-}, \pi_{-+-}, \pi_{-+-}, \pi_{+--}, \pi_{+--}, \pi_{+--}, \pi_{+--})'$.

As hipóteses de interesse são

$$\begin{aligned} H_S : Sens_{(RM)} &= Sens_{(EC)}, & H_E : Espec_{(RM)} &= Espec_{(EC)}, & H_{SE} : H_S &\cap H_E, \\ H_P : VPP_{(RM)} &= VPP_{(EC)}, & H_N : VP_N_{(RM)} &= VP_N_{(EC)}, & H_{PN} : H_P &\cap H_N, \end{aligned}$$

e podem ser re-expressadas como $H : \mathbf{C} \mathbf{F}(\boldsymbol{\pi}) = \mathbf{0}$, para convenientes escolhas de \mathbf{C} . Em particular, H_S e H_P podem ser avaliadas com $\mathbf{C} = \mathbf{C}_1 = (1, -1, 0, 0)$, H_E e H_N , com $\mathbf{C} = \mathbf{C}_2 = (0, 0, 1, -1)$ e H_{SE} e H_{PN} , com $\mathbf{C} = (\mathbf{C}'_1, \mathbf{C}'_2)'$.

Sob a ACI e a ACAI, supõe-se uma distribuição produto de multinomiais com cada multinomial independente correspondendo a cada um dos testes. Os resultados das três análises pela metodologia de MQG são apresentados na Tabela 3.23, com estimativas dos parâmetros, correspondentes erros padrões e valores-p dos testes das hipóteses descritas.

Tabela 3.23: Análises de ressonância magnética e ecocolonoscopia com parte dos dados

Parâmetro	ACC ($n = 13$)			ACI ($n = 108$)			ACAI ($n = 134$)		
	est.	e.p.	valor-p	est.	e.p.	valor-p	est.	e.p.	valor-p
$Sens_{(RM)}$	0.500	0.204	0.552 [†]	0.371	0.082	0.309 [†]	0.390	0.076	0.162 [†]
$Sens_{(EC)}$	0.667	0.192		0.469 [§]	0.545		0.150	0.084 [§]	
$Espec_{(RM)}$	1.000	0.006	0.281 [‡]	0.911	0.038	0.048 [‡]	0.921	0.034	0.085 [‡]
$Espec_{(EC)}$	0.857	0.132		0.500	0.204		0.692	0.128	
$VPP_{(RM)}$	0.999	0.015	0.265 [♣]	0.722	0.106	0.769 [♣]	0.762	0.093	0.755 [♣]
$VPP_{(EC)}$	0.800	0.179		0.362 [♣]	0.667		0.157	0.188 [♣]	
$VP_N_{(RM)}$	0.700	0.145	0.724 [◇]	0.699	0.054	0.071 [◇]	0.699	0.050	0.309 [◇]
$VP_N_{(EC)}$	0.750	0.153		0.375	0.171		0.562	0.124	

Obs.: os valores-p correspondem a [†] H_S , [‡] H_E , [§] H_{SE} , [♣] H_P , [◇] H_N e [♣] H_{PN} .

Tanto a ACC quando a ACI descartam importantes dados observados dos pacientes excluídos, enquanto que a ACAI também desconsidera a dependência entre as medidas repetidas realizadas nos indivíduos avaliados por ambos os testes, numa tentativa de se incorporar a máxima informação disponível.

Os resultados das análises sob os mecanismos MCAR e MAR usando a metodologia híbrida (MV/MQG) são exibidos na Tabela 3.24. Como esperado, os erros padrões das estimativas obtidas com a inclusão de todas as observações são substancialmente menores que os baseados em informações parciais, com exceção dos associados à $Espec_{(EC)}$ e $VPP_{(EC)}$ sob o mecanismo MAR. Isso é uma consequência das frequências nulas das celas ($RM = +, EC = -, D = -$) e ($RM = +, EC = +, D = -$) do cenário de ausência de omissão. De fato, como o mecanismo

MAR envolve um modelo saturado, contagens iguais a zero geram matrizes de covariâncias instáveis. Repete-se as análises sob ambos os mecanismos, substituindo as frequências nulas por um valor pequeno, *i.e.*, 0.001, que foi a mesma substituição realizada na ACC da Tabela 3.23. Os resultados da Tabela 3.25 praticamente coincidem com os apresentados na Tabela 3.24, principalmente para o mecanismo MCAR, com exceção dos erros padrões e valores-p associados às frequências substituídas, ou seja, $Espec_{(EC)}$, $VPP_{(EC)}$ e $VPN_{(EC)}$, sob o mecanismo MAR. Estes resultados estão mais próximos dos obtidos sob o mecanismo MCAR. Conforme mostrado nas Seções 1.2 e 2.3, as EMV de parâmetros associados às probabilidades de categorização sob o mecanismo MAR e qualquer outro mais restritivo são iguais (veja as Tabelas 3.24 e 3.25). Se a matriz de informação observada tivesse sido utilizada ao invés da matriz de informação de Fisher para a estimação da precisão, todos os resultados exibidos seriam iguais sob ambos os mecanismos.

Tabela 3.24: Análises de ressonância magnética e ecocolonoscopia com todos os dados

Parâmetro	MCAR ($n = 219$)			MAR ($n = 219$)		
	est.	e.p.	valor-p	est.	e.p.	valor-p
$Sens_{(RM)}$	0.390	0.072	0.196	0.390	0.076	0.164
$Sens_{(EC)}$	0.587	0.135		0.041	0.587	
$Espec_{(RM)}$	0.916	0.036	0.030	0.916	0.036	0.162
$Espec_{(EC)}$	0.674	0.110		0.030	0.674	
$VPP_{(RM)}$	0.788	0.082	0.101	0.788	0.082	0.204
$VPP_{(EC)}$	0.589	0.104		0.014	0.589	
$VPN_{(RM)}$	0.654	0.042	0.826	0.654	0.043	0.835
$VPN_{(EC)}$	0.672	0.086		0.091	0.672	

Comparando os resultados nas Tabelas 3.23 e 3.25, nota-se que as estimativas obtidas com todos os dados estão mais próximas da ACI e da ACAI que da ACC, pois a ACI e a ACAI englobam um maior número de pacientes, embora não considerem a dependência entre os resultados obtidos sob ambos os testes. Ao incorporar a correlação positiva entre as estimativas dos parâmetros (valores marcados com \dagger na Tabela 3.26), aumenta-se a precisão das suas diferenças. Por causa da dependência, mesmo quando não há evidências de que H_P e H_N não sejam marginalmente verdadeiros, sua igualdade conjunta, avaliada por H_{PN} , mostra que ambas afirmações não são simultaneamente válidas, visto que $\widehat{VPP}_{(RM)} - \widehat{VPP}_{(EC)}$ é > 0 ,

Tabela 3.25: Análises de ressonância magnética e ecocolonoscopia com todos os dados, substituindo contagens nulas por 0.001

Parâmetro	MCAR ($n = 219$)			MAR ($n = 219$)		
	est.	e.p.	valor-p	est.	e.p.	valor-p
$Sens_{(RM)}$	0.390	0.072	0.196	0.390	0.076	0.164
$Sens_{(EC)}$	0.587	0.135		0.042	0.587	
$Espec_{(RM)}$	0.916	0.036	0.030	0.916	0.036	0.056
$Espec_{(EC)}$	0.674	0.110		0.042	0.674	
$VPP_{(RM)}$	0.787	0.082	0.101	0.787	0.082	0.119
$VPP_{(EC)}$	0.589	0.104		0.014	0.589	
$VPN_{(RM)}$	0.654	0.042	0.826	0.654	0.043	0.818
$VPN_{(EC)}$	0.673	0.086		0.042	0.673	

$\widehat{VPN}_{(RM)} - \widehat{VPN}_{(EC)}$ é < 0 e ambas diferenças são positivamente correlacionadas: $r = 0.781$ sob a suposição MCAR e $r = 0.793$ sob a suposição MAR. Esta correlação também está presente sob a ACC ($r = 0.411$), mas o teste correspondente não permitiu a mesma conclusão, provavelmente devido ao pequeno tamanho da amostra. As estimativas do VPP e VPN de um mesmo teste não são correlacionadas sob a ACC, porque se baseiam, por definição dos respectivos parâmetros, em diferentes resultados do teste; contudo, essas estimativas se tornam correlacionadas sob os mecanismos MAR e MCAR quando o resultado do teste está omissos. Isso induz uma incerteza e, conseqüentemente, uma dependência entre os resultados (veja os valores marcados com § na Tabela 3.26). A mesma dependência seria induzida entre as estimativas da $Sens$ e $Espec$ se existissem pacientes com valores omissos em D .

Tabela 3.26: Correlações entre as estimativas dos parâmetros

Parâmetro	ACC			MCAR			MAR		
	$Sens_{(RM)}$	$Sens_{(EC)}$	$Espec_{(RM)}$	$Sens_{(RM)}$	$Sens_{(EC)}$	$Espec_{(RM)}$	$Sens_{(RM)}$	$Sens_{(EC)}$	$Espec_{(RM)}$
$Sens_{(EC)}$	0.000 [†]			0.008 [†]			0.015 [†]		
$Espec_{(RM)}$	0.000	0.000		0.000	0.000		0.000	0.000	
$Espec_{(EC)}$	0.000	0.000	0.017 [†]	0.000	0.000	0.117 [†]	0.000	0.000	0.075 [†]
Parâmetro	ACC			MCAR			MAR		
	$VPP_{(RM)}$	$VPP_{(EC)}$	$VPN_{(RM)}$	$VPP_{(RM)}$	$VPP_{(EC)}$	$VPN_{(RM)}$	$VPP_{(RM)}$	$VPP_{(EC)}$	$VPN_{(RM)}$
$VPP_{(EC)}$	0.027 [†]			0.170 [†]			0.141 [†]		
$VPN_{(RM)}$	0.000 [§]	-0.401		0.225 [§]	-0.211		0.244 [§]	-0.198	
$VPN_{(EC)}$	-0.014	0.000 [§]	0.549 [†]	-0.051	0.666 [§]	0.273 [†]	-0.065	0.663 [§]	0.280 [†]

Para ilustrar estabilidade dos resultados nas diferentes abordagens de análise quando a quantidade de dados com omissão é moderada, considera-se um exemplo com US , em vez de EC (veja Tabela 3.21). Neste caso, há omissão apenas no resultado da avaliação da RM . Os resultados sob a ACC e a ACAI estão resumidas na Tabela 3.27, enquanto que os resultados sob os mecanismos MCAR e MAR são exibidos na Tabela 3.28. Note que a ACI não pode ser realizada uma vez que não há pacientes avaliados apenas via RM .

Tabela 3.27: Análises de ressonância magnética e ultra-sonografia com parte dos dados

Parâmetro	ACC ($n = 104$)			ACAI ($n = 323$)		
	est.	e.p.	valor-p	est.	e.p.	valor-p
$Sens_{(RM)}$	0.390	0.076	< 0.001	0.390	0.076	< 0.001
$Sens_{(US)}$	0.951	0.034		0.938	0.024	
$Espec_{(RM)}$	0.921	0.034	0.039	0.921	0.034	0.137
$Espec_{(US)}$	0.984	0.016		0.975	0.014	
$VPP_{(RM)}$	0.762	0.093	0.012	0.762	0.093	0.029
$VPP_{(US)}$	0.975	0.025		0.968	0.018	
$VPN_{(RM)}$	0.699	0.050	< 0.001	0.699	0.050	< 0.001
$VPN_{(US)}$	0.969	0.022		0.952	0.019	

Tabela 3.28: Análises de ressonância magnética e ultra-sonografia com todos os dados

Parâmetro	MCAR ($n = 219$)			MAR ($n = 219$)		
	est.	e.p.	valor-p	est.	e.p.	valor-p
$Sens_{(RM)}$	0.392	0.072	< 0.001	0.392	0.076	< 0.001
$Sens_{(US)}$	0.938	0.024		0.938	0.024	
$Espec_{(RM)}$	0.912	0.034	0.049	0.912	0.033	0.039
$Espec_{(US)}$	0.975	0.014		0.975	0.014	
$VPP_{(RM)}$	0.781	0.078	0.010	0.781	0.077	0.009
$VPP_{(US)}$	0.968	0.018		0.968	0.018	
$VPN_{(RM)}$	0.654	0.042	< 0.001	0.654	0.043	< 0.001
$VPN_{(US)}$	0.952	0.019		0.952	0.019	

O mecanismo MCAR, condicionalmente à suposição MAR, não é rejeitado pelas estatísticas da razão de verossimilhanças e de Pearson ($p > 0.80$ em ambos) para todos os dados da Tabela 3.21. Estes resultados baseiam-se em testes assintóticos que podem ser imprecisos, devido à configuração esparsa da tabela de contingência. Contudo, a magnitude dos valores-p sugerem que a conclusão deve prevalecer. Por essa razão, e por causa do grande número de contagens iguais a zero, as análises a seguir são realizadas sob o mecanismo MCAR.

A análise apresentada previamente é facilmente estendida para os três testes de diagnóstico considerados simultaneamente incluindo um índice adicional às probabilidades (π) e alterando as matrizes \mathbf{A}_1 e \mathbf{A}_2 apropriadamente. Os resultados são apresentados na Tabela 3.29, onde se incluem os valores-p de testes de comparação entre pares das funções paramétricas indicadas. Conclui-se que a especificidade da *US* é estatisticamente maior que a da *RM* e subseqüentemente maior que a da *EC*. Adicionalmente, a sensibilidade e os valores preditivos negativo e positivo da *US* são estatisticamente maiores que os da *RM* e *EC*, ainda que não se possa afirmar se esses parâmetros possuem maiores valores para a *RM* ou *EC*. Ademais, os valores preditivos positivo e negativo da *RM* e *EC* não podem ser considerados simultaneamente iguais à luz dos resultados dos testes da Tabela 3.25.

Tabela 3.29: Análise de ressonância magnética, ecocolonosopia e ultra-sonografia sob MCAR

Parâmetro	Estimativa	Erro padrão	Correlações entre as estimativas dos parâmetros					Valores-p	
			<i>Sens</i> _(RM)	<i>Sens</i> _(EC)	<i>Sens</i> _(US)	<i>Espec</i> _(RM)	<i>Espec</i> _(EC)	<i>RM</i>	<i>EC</i>
<i>Sens</i> _(RM)	0.391	0.072							
<i>Sens</i> _(EC)	0.592	0.135	0.006					0.187	
<i>Sens</i> _(US)	0.938	0.024	-0.039	0.018				<0.001	0.011
<i>Espec</i> _(RM)	0.909	0.035	0.000	0.000	0.000				
<i>Espec</i> _(EC)	0.674	0.111	0.000	0.000	0.000	0.102		0.037	
<i>Espec</i> _(US)	0.975	0.014	0.000	0.000	0.000	0.372	0.023	0.041	0.007
			<i>VPP</i> _(RM)	<i>VPP</i> _(EC)	<i>VPP</i> _(US)	<i>VPN</i> _(RM)	<i>VPN</i> _(EC)		
<i>VPP</i> _(RM)	0.774	0.078							
<i>VPP</i> _(EC)	0.591	0.104	0.166					0.126	
<i>VPP</i> _(US)	0.968	0.018	0.381	0.091				0.008	<0.001
<i>VPN</i> _(RM)	0.653	0.042	0.218	-0.214	-0.097				
<i>VPN</i> _(EC)	0.675	0.086	-0.066	0.664	-0.070	0.268		0.790	
<i>VPN</i> _(US)	0.952	0.019	-0.103	-0.093	0.000	0.219	0.127	<0.001	0.001

Esse exemplo demonstra que quando há uma quantidade substancial de dados omissos, ACC ou outras abordagens alternativas que não consideram todos os dados podem levar a diferentes estimativas e conclusões mesmo quando o mecanismo MCAR é plausível, devido a uma combinação de viés e aumento da variabilidade. Por outro lado, quando o número de dados omissos é suficientemente pequeno, os benefícios das análises baseadas em mecanismos não dependentes das respostas (observadas ou não) não são tão aparentes, apesar de os erros padrões das estimativas serem usualmente menores.

Mesmo que as estatísticas da razão de verossimilhanças e de Pearson não indiquem evidências contra o mecanismo MCAR, deve-se atentar que isso é feito sob a alternativa do mecanismo MAR. Ao questionar a validade desta suposição, análises de sensibilidade com mecanismos MNAR como indicadas na Seção 1.3, e realizadas nas análises dos outros exemplos, devem ser consideradas. Todavia, se o ajuste de estruturas MNAR já é por si só problemático, a ocorrência das frequências nulas torna a tarefa ainda mais delicada, sendo o motivo pelo qual tais análises não são aqui levadas adiante.

Mecanismos MNAR podem ocorrer, por exemplo, quando nem todos os pacientes têm o estado da doença verificado, ou seja, quando há um viés de verificação (*verification bias*, como sugerido por Begg e Greenes, 1983), pois a ocorrência dos dados omissos geralmente está associada direta ou indiretamente (*e.g.*, variáveis não armazenadas) com os próprios valores omissos. Exemplos de análises com viés de verificação assumindo ignorabilidade do mecanismo de omissão são apresentadas por Zhou (1998), Barnhart e Kosinski (2003) e Alonzo (2005). Mecanismos não-ignoráveis foram considerados por Baker (1995a), Kosinski e Barnhart (2003a, 2003b) e Zhou e Castelluccio (2004). Compilações de discussões sobre métodos estatísticos de avaliação de testes de diagnóstico e problemas relacionados podem ser encontrados em Zhou, Obuchowski e McClish (2002) e Pepe (2003).



CAPÍTULO 4

Discussão

Neste trabalho abordou-se aspectos teóricos, computacionais e aplicados de análises clássicas de dados categorizados com omissão.

Uma revisão da literatura foi resumida no Capítulo 1, em que se introduziu os conceitos mais importantes da teoria por meio de um exemplo considerando apenas duas variáveis respostas dicotômicas e com o auxílio de estudos de simulação.

No Capítulo 2 ampliou-se a modelagem descrita em Paulino (1988, 1991) e Paulino e Singer (2006) para considerar variáveis explicativas. Os resultados foram desenvolvidos em formulação matricial adequada para a implementação computacional, consumada por meio de uma biblioteca para o ambiente estatístico R (R Development Core Team, 2006), conforme delineado na Seção 3.1.

Espera-se, com a disponibilização das rotinas, facilitar o traçado das inferências apresentadas nesta dissertação. As técnicas descritas foram ilustradas com cinco exemplos de características diversas, visto que se ajustou modelos estruturais lineares (homogeneidade marginal), log-lineares (independência, razão de chances adjacentes comum) e funcionais lineares (*kappa*, *kappa* ponderado, sensibilidade/especificidade, valor preditivo positivo/negativo) para as probabilidades de categorização. Os padrões de omissão também foram variados, com omissões em uma ou duas variáveis, confundimento de células vizinhas, sem ou com subpopulações.

O Exemplo 5 demonstrou que mesmo quando não há evidências contra o mecanismo MCAR, as ACC, que são consistentes sob tal suposição, podem ter conclusões bem diferentes das obtidas incorporando todos os dados disponíveis. Nesse sentido, conforme mostrou-se nas Seções 1.2.1 e 1.2.2, análises sob as estruturas MAR/MCAR protegem os resultados de possíveis vieses que poderiam ocorrer em ACC se o mecanismo verdadeiro for o MAR e ainda produzem estimadores mais eficientes se o mecanismo for o MCAR.

A Seção 1.2.3 revelou que as estruturas MNAR saturadas podem não ter ajustes perfeitos e que isto está associado a estimativas das probabilidades condicionais de omissão na fronteira do espaço paramétrico, embora isto também possa ocorrer com estruturas MNAR não saturadas e não seja um indício de que o mecanismo de omissão assumido não seja verdadeiro. Além disso, mostrou-se que os parâmetros das estruturas MNAR podem ser inidentificáveis. Tanto no caso de as estimativas das probabilidades condicionais de omissão estarem na fronteira do espaço paramétrico, quanto no caso de falta de identificabilidade, as propriedades assintóticas dos TRV são indesejáveis e os EMV são enviesados, embora ainda com viés menor do que se obtém sob a ACC se o mecanismo MNAR proposto for o verdadeiro.

Ao considerar estruturas MNAR, abre-se uma série de questões de sensibilidade, já que as frequências não-observáveis não permitem que se selecione satisfatoriamente o mecanismo que gerou os dados omissos, conforme discutido na Seção 1.3. Intervalos para o melhor-pior caso, de ignorância e incerteza constituem ferramentas inferenciais importantes nessas situações. Nos exemplos em que se utilizou tais análises, mostrou-se que o desconhecimento absoluto sobre o mecanismo de omissão resulta em tamanha ignorância sobre as inferências de interesse, que em geral se torna difícil chegar a uma conclusão incisiva. No entanto, não se acredita que isso seja uma deficiência das técnicas ou um motivo para o seu desuso, mas uma conseqüência natural da falta de informação acerca do mecanismo de omissão que deve, obrigatoriamente, ser refletida numa análise honesta. Isso não impede, é claro, que se restrinja a classe de estruturas para o mecanismo de omissão considerada desde que se tenha alguma informação, prévia ou obtida durante o estudo, para esse fim. Apenas no Exemplo 4 ilustrou-se essa prática ao restringir a faixa de variação do parâmetro de sensibilidade. Se a informação obtida sobre o mecanismo de omissão possibilitar, adicionalmente à proposição das estruturas, que se elicie distribuições *a priori* sobre os parâmetros, pode-se adotar o procedimento inferencial bayesiano, sobre o qual algumas referências são indicadas na Seção 1.5.

Possíveis trabalhos futuros envolvem, por exemplo, a realização de uma revisão bibliográfica mais abrangente e maiores investigações dos aspectos inferenciais relacionados às patologias das estruturas MNAR, discutidas na Seção 1.2.3 e evidenciadas nas análises dos exemplos (principalmente no Exemplo 2, Figura 3.4, p.107). Explicitamente com relação às patologias das estruturas MNAR, indaga-se o desenvolvimento de:

- correções de viés para os EMV e refinamentos para os TRV com aproximações assintóticas de ordem superior;

- estudos sobre o uso da fatoração de modelos de mistura de padrões (1.2) e/ou outras reparametrizações (quais?) como alternativas para transpor a inidentificabilidade dos parâmetros;
- formas mais práticas de inspeção dos pontos nos quais as estruturas saturadas ou reduzidas podem apresentar problemas de identificabilidade de seus parâmetros.

Conforme alertado na Seção 1.2.3, apesar de não ser possível decidir se determinada configuração amostral originou de uma estrutura paramétrica identificável, o conhecimento prévio dos casos em que os parâmetros são inidentificáveis é importante por destacar as condições que os resultados assintóticos não podem ser usufruídos (*e.g.*, sob a estrutura MNAR da Tabela 1.11, o caso em questão seria o de Y_1 e Y_2 não estarem associados).

Conclui-se a dissertação citando Meng (2000):

“Much remains to be done, however. The most pressing task, in my opinion, is placing further emphasis on the general recognition and understanding, at a conceptual level, of the necessity of properly dealing with the missing-data mechanism, as part of our ongoing emphasis on the importance of the data collection process in any meaningful statistical analysis. The missing-data mechanism is in the blood of statistics, and it is the nastiest and the most deceptive cell, especially for non-statisticians — why on earth should anyone be concerned with data that one does not even have?”

APÊNDICE A

Detalhes da obtenção de algumas expressões

A seguir, descreve-se detalhes da obtenção de alguns resultados apresentados no Capítulo 2.

A.1 Expressão (2.10)

De acordo com a terminologia utilizada por Dempster, Laird e Rubin (1977), (2.9) é a log-verossimilhança de dados observados (*observed-data log-likelihood*) e

$$\ln L_1^C(\boldsymbol{\theta} | \{y_{str}\}) = \sum_{s=1}^S \sum_{t=1}^{T_s} \sum_{r=1}^R y_{str} \ln \theta_{r(s)} \quad (\text{A.1})$$

é a log-verossimilhança de dados ampliados ou completos (*augmented/complete-data log-likelihood*).

Dado a estimativa de $\boldsymbol{\theta}_s$ na i -ésima iteração, $\hat{\boldsymbol{\theta}}_s^{(i)}$, o passo-E é definido por

$$\hat{y}_{str}^{(i)} = \hat{E} \left(y_{str} \mid \mathbf{N}, \hat{\boldsymbol{\theta}}_s^{(i)} \right) = n_{stc} \frac{\hat{\theta}_{r(s)}^{(i)}}{\mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s^{(i)}}, \quad (\text{A.2})$$

$s = 1, \dots, S$, $t = 1, \dots, T_s$, $r = 1, \dots, R$, $\{c : r \in \mathcal{C}_{stc}\}$. Este resultado é uma extensão de (1.6), refletindo o significado da omissão aleatória ou não-informativa e indicando que as unidades classificadas numa determinada classe estão distribuídas nas categorias de respostas pertencentes a essa classe da mesma forma como se não houvesse omissão.

Maximizando (A.1) com relação a $\{\theta_{r(s)}\}$ obtém-se o passo-M, a saber

$$\hat{\theta}_{r(s)} = \left(\sum_{t=1}^{T_s} \sum_{r=1}^R y_{str} \right)^{-1} \sum_{t=1}^{T_s} y_{str} = \frac{1}{n_{s++}} \sum_{t=1}^{T_s} y_{str}, \quad s = 1, \dots, S, \quad r = 1, \dots, R. \quad (\text{A.3})$$

Combinando (A.2) e (A.3) obtém-se

$$\hat{\theta}_{r(s)}^{(i+1)} = \frac{1}{n_{s++}} \left(n_{s1r} + \sum_{\substack{t=2 \\ \{c:r \in \mathcal{C}_{stc}\}}}^{T_s} n_{stc} \frac{\hat{\theta}_{r(s)}^{(i)}}{\mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s^{(i)}} \right), \quad s = 1, \dots, S, \quad r = 1, \dots, R,$$

que permite a construção de (2.10).

A.2 Expressões (2.12) e (2.14)

Seja

$$\begin{aligned} \ln L_{1st}(\boldsymbol{\theta}_s | \mathbf{N}_{st}) &= \sum_{c=1}^{R_{st}} n_{stc} \ln(\mathbf{z}'_{stc} \boldsymbol{\theta}_s) \\ &= \sum_{c=1}^{R_{st}-1} n_{stc} \ln \theta_{c(st)} + n_{stR_{st}} \ln(1 - \mathbf{1}'_{R_{st}-1} \bar{\boldsymbol{\theta}}_{st}) \end{aligned} \quad (\text{A.4})$$

o termo da log-verossimilhança de (2.9) associado ao t -ésimo cenário de omissão da s -ésima subpopulação visto como função de $\bar{\boldsymbol{\theta}}_{st}$, que por sua vez depende de $\bar{\boldsymbol{\theta}}_s$, $s = 1, \dots, S$, $t = 1, \dots, T_s$.

Derivando (A.4) em relação a $\bar{\boldsymbol{\theta}}_{st}$, obtém-se o vetor gradiente

$$\begin{aligned} \frac{\partial \ln L_{1st}(\boldsymbol{\theta}_s | \mathbf{N}_{st})}{\partial \bar{\boldsymbol{\theta}}_{st}} &= \left(\frac{n_{stc}}{\theta_{c(st)}} - \frac{n_{stR_{st}}}{1 - \sum_{d=1}^{R_{st}-1} \theta_{d(st)}}, c = 1, \dots, R_{st} - 1 \right)' \\ &= n_{st+} \left(\frac{p_{c(st)}}{\theta_{c(st)}} - \frac{1 - \sum_{d=1}^{R_{st}-1} p_{d(st)}}{1 - \sum_{d=1}^{R_{st}-1} \theta_{d(st)}}, c = 1, \dots, R_{st} - 1 \right)' \\ &= n_{st+} \left(\frac{p_{c(st)}}{\theta_{c(st)}} - \frac{1 - \sum_{d=1}^{R_{st}-1} (\theta_{d(st)} + p_{d(st)} - \theta_{d(st)})}{1 - \sum_{d=1}^{R_{st}-1} \theta_{d(st)}}, c = 1, \dots, R_{st} - 1 \right)' \end{aligned}$$

$$\begin{aligned}
&= n_{st+} \left(\frac{p_{c(st)}}{\theta_{c(st)}} - 1 + \frac{\sum_{d=1}^{R_{st}-1} (p_{d(st)} - \theta_{d(st)})}{1 - \sum_{d=1}^{R_{st}-1} \theta_{d(st)}}, c = 1, \dots, R_{st} - 1 \right)' \\
&= n_{st+} \left(\frac{p_{c(st)} - \theta_{c(st)}}{\theta_{c(st)}} + \frac{\sum_{d=1}^{R_{st}-1} (p_{d(st)} - \theta_{d(st)})}{1 - \sum_{d=1}^{R_{st}-1} \theta_{d(st)}}, c = 1, \dots, R_{st} - 1 \right)' \\
&= n_{st+} \left[\mathbf{D}_{\bar{\theta}_{st}}^{-1} + \frac{1}{1 - \mathbf{1}'_{R_{st}-1} \bar{\theta}_{st}} \mathbf{1}_{R_{st}-1} \mathbf{1}'_{R_{st}-1} \right] (\bar{\mathbf{p}}_{st} - \bar{\boldsymbol{\theta}}_{st}) \\
&= \left[\frac{1}{n_{st+}} (\mathbf{D}_{\bar{\theta}_{st}} - \bar{\boldsymbol{\theta}}_{st} \bar{\boldsymbol{\theta}}'_{st}) \right]^{-1} (\bar{\mathbf{p}}_{st} - \bar{\boldsymbol{\theta}}_{st}) \\
&= [\boldsymbol{\Sigma}_{st}(\bar{\boldsymbol{\theta}}_s)]^{-1} (\bar{\mathbf{p}}_{st} - \bar{\boldsymbol{\theta}}_{st}),
\end{aligned}$$

em que a penúltima igualdade é obtida utilizando um resultado apresentado em Rao (1973, p.33)¹.

Utilizando regras de diferenciação matricial, tem-se

$$\begin{aligned}
\frac{\partial \ln L_{1st}(\boldsymbol{\theta}_s | \mathbf{N}_{st})}{\partial \bar{\boldsymbol{\theta}}_s} &= \left[\frac{\partial \bar{\boldsymbol{\theta}}_{st}}{\partial \bar{\boldsymbol{\theta}}'_s} \right]' \frac{\partial \ln L_{1st}(\boldsymbol{\theta}_s | \mathbf{N}_{st})}{\partial \bar{\boldsymbol{\theta}}_{st}} = \left[\frac{\partial}{\partial \bar{\boldsymbol{\theta}}'_s} \bar{\mathbf{Z}}'_{st} \bar{\boldsymbol{\theta}}_s \right]' \frac{\partial \ln L_{1st}(\boldsymbol{\theta}_s | \mathbf{N}_{st})}{\partial \bar{\boldsymbol{\theta}}_{st}} \\
&= \bar{\mathbf{Z}}_{st} [\boldsymbol{\Sigma}_{st}(\bar{\boldsymbol{\theta}}_s)]^{-1} (\bar{\mathbf{p}}_{st} - \bar{\boldsymbol{\theta}}_{st}),
\end{aligned}$$

de onde se chega facilmente ao resultado (2.12).

A matriz hessiana de (A.4) em relação a $\bar{\boldsymbol{\theta}}_{st}$ tem os elementos da diagonal iguais a

$$\frac{\partial^2 \ln L_{1st}(\boldsymbol{\theta}_s | \mathbf{N}_{st})}{\partial \bar{\theta}_{c(st)}^2} = -\frac{n_{stc}}{\theta_{c(st)}^2} - \frac{n_{stR_{st}}}{\left(1 - \sum_{d=1}^{R_{st}-1} \theta_{d(st)}\right)^2}, \quad c = 1, \dots, R_{st} - 1$$

¹Seja \mathbf{A} uma matriz não singular e \mathbf{u} e \mathbf{v} , dois vetores colunas. Pode-se obter a inversa de $\mathbf{A} + \mathbf{u}\mathbf{v}'$ conhecendo apenas a inversa de \mathbf{A} da seguinte maneira

$$(\mathbf{A} + \mathbf{u}\mathbf{v}')^{-1} = \mathbf{A}^{-1} - \frac{(\mathbf{A}^{-1}\mathbf{u})(\mathbf{v}'\mathbf{A}^{-1})}{1 + \mathbf{v}'\mathbf{A}^{-1}\mathbf{u}}.$$

e os elementos localizados fora da diagonal iguais a

$$\frac{\partial^2 \ln L_{1st}(\boldsymbol{\theta}_s | \mathbf{N}_{st})}{\partial \bar{\theta}_{b(st)} \partial \bar{\theta}_{c(st)}} = - \frac{n_{stR_{st}}}{\left(1 - \sum_{d=1}^{R_{st}-1} \theta_{d(st)}\right)^2}, \quad b, c = 1, \dots, R_{st} - 1, \quad b \neq c,$$

podendo ser escrita como

$$\frac{\partial^2 \ln L_{1st}(\boldsymbol{\theta}_s | \mathbf{N}_{st})}{\partial \bar{\boldsymbol{\theta}}_{st} \partial \bar{\boldsymbol{\theta}}'_{st}} = - \left[\mathbf{D}_{\bar{\mathbf{N}}_{st}} \mathbf{D}_{\bar{\boldsymbol{\theta}}_{st}}^{-2} + \frac{n_{stR_{st}}}{\left(1 - \mathbf{1}'_{R_{st}-1} \bar{\boldsymbol{\theta}}_{st}\right)^2} \mathbf{1}_{R_{st}-1} \mathbf{1}'_{R_{st}-1} \right],$$

em que $\mathbf{D}_{\bar{\boldsymbol{\theta}}_{st}}^{-2} = \mathbf{D}_{\bar{\boldsymbol{\theta}}_{st}}^{-1} \mathbf{D}_{\bar{\boldsymbol{\theta}}_{st}}^{-1}$.

Utilizando regras de diferenciação matricial, chega-se a

$$\begin{aligned} \frac{\partial^2 \ln L_{1st}(\boldsymbol{\theta}_s | \mathbf{N}_{st})}{\partial \bar{\boldsymbol{\theta}}_s \partial \bar{\boldsymbol{\theta}}'_s} &= \left[\frac{\partial \bar{\boldsymbol{\theta}}_{st}}{\partial \bar{\boldsymbol{\theta}}'_s} \right]' \frac{\partial^2 \ln L_{1st}(\boldsymbol{\theta}_s | \mathbf{N}_{st})}{\partial \bar{\boldsymbol{\theta}}_{st} \partial \bar{\boldsymbol{\theta}}'_{st}} \frac{\partial \bar{\boldsymbol{\theta}}_{st}}{\partial \bar{\boldsymbol{\theta}}'_s} \\ &= -\bar{\mathbf{Z}}_{st} \left[\mathbf{D}_{\bar{\mathbf{N}}_{st}} \mathbf{D}_{\bar{\boldsymbol{\theta}}_{st}}^{-2} + \frac{n_{stR_{st}}}{\left(1 - \mathbf{1}'_{R_{st}-1} \bar{\boldsymbol{\theta}}_{st}\right)^2} \mathbf{1}_{R_{st}-1} \mathbf{1}'_{R_{st}-1} \right] \bar{\mathbf{Z}}'_{st}, \end{aligned}$$

de onde se obtém facilmente o resultado (2.14).

A.3 Expressão (2.17)

Por meio da verossimilhança do mecanismo MAR (2.4), nota-se que os estimadores de máxima verossimilhança de $\{\alpha_{t(cs)} \sum_{r \in \mathcal{C}_{stc}} \theta_{r(s)}\}$ são iguais a $\{n_{stc}/n_{st+}\}$. Após a obtenção das EMV $\{\hat{\theta}_{r(s)}\}$ de $\{\theta_{r(s)}\}$ por meio de algum método iterativo, utilizando a propriedade da invariância, obtém-se as EMV

$$\left\{ \hat{\alpha}_{t(cs)} = \frac{n_{stc}}{n_{st+} \sum_{r \in \mathcal{C}_{stc}} \hat{\theta}_{r(s)}} = \frac{n_{stc}}{n_{st+} \mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s} \right\}$$

de $\{\alpha_{t(cs)}\}$, que escritas matricialmente são iguais a (2.17).

A.4 Expressão (2.21)

Note que

$$\left(\hat{\alpha}_{t(cs)}, c = 1, \dots, R_{st}\right)' = \hat{\boldsymbol{\alpha}}_{st}^{\text{MAR}} = \frac{1}{n_{s++}} \mathbf{D}_{\mathbf{z}'_{st} \hat{\boldsymbol{\theta}}_s}^{-1} \mathbf{N}_{st} = \left(\frac{n_{stc}}{n_{s++} \mathbf{z}'_{stc} \hat{\boldsymbol{\theta}}_s}, c = 1, \dots, R_{st}\right)',$$

$s = 1, \dots, S$, $t = 1, \dots, T_s$. Substituindo $\{\hat{\alpha}_{t(cs)} = n_{stc}/(n_{s++}\mathbf{z}'_{stc}\hat{\boldsymbol{\theta}}_s)\}$ em $\mathcal{I}_1(\hat{\boldsymbol{\theta}}, \{\hat{\boldsymbol{\alpha}}_{st}^{\text{MAR}}\})$ obtém-se a igualdade (2.21).

A.5 Expressões (2.73) e (2.74)

Seja

$$\begin{aligned} \ln L_{1st}(\boldsymbol{\theta}_s(\boldsymbol{\beta})|\mathbf{N}_{st}) &= \sum_{c=1}^{R_{st}} n_{stc} \ln(\mathbf{z}'_{stc}\boldsymbol{\theta}_s(\boldsymbol{\beta})) = \mathbf{N}'_{st} \ln(\mathbf{Z}'_{st}\boldsymbol{\theta}_s(\boldsymbol{\beta})) \\ &= \mathbf{N}'_{st} \ln(\boldsymbol{\theta}_{st}(\boldsymbol{\theta}_s(\boldsymbol{\beta}))) \equiv \ln L_{1st}(\boldsymbol{\theta}_{st}(\boldsymbol{\theta}_s(\boldsymbol{\beta}))|\mathbf{N}_{st}) \end{aligned} \quad (\text{A.5})$$

o termo da log-verossimilhança de (2.9) associado ao t -ésimo cenário de omissão da s -ésima subpopulação visto como função de $\boldsymbol{\theta}_{st}(\boldsymbol{\theta}_s(\boldsymbol{\beta}))$, $s = 1, \dots, S$, $t = 1, \dots, T_s$, em que $\boldsymbol{\beta}$ indica que a log-verossimilhança é escrita sob o modelo log-linear. Recorde-se que $\boldsymbol{\theta}_s(\boldsymbol{\beta}) = \mathbf{D}_{\boldsymbol{\psi}_s}^{-1} \exp(\mathbf{X}_s\boldsymbol{\beta})$, em que $\boldsymbol{\psi}_s = (\mathbf{1}_R \mathbf{1}'_R) \exp(\mathbf{X}_s\boldsymbol{\beta})$, e note que a expressão pode ser reescrita como

$$\boldsymbol{\theta}_s(\boldsymbol{\beta}) = \frac{1}{\mathbf{1}'_R \exp(\mathbf{X}_s\boldsymbol{\beta})} \exp(\mathbf{X}_s\boldsymbol{\beta}). \quad (\text{A.6})$$

Utilizando regras de diferenciação matricial, tem-se

$$\frac{\partial \ln L_{1st}(\boldsymbol{\theta}_s(\boldsymbol{\beta})|\mathbf{N}_{st})}{\partial \boldsymbol{\beta}} = \left[\frac{\partial \boldsymbol{\theta}_s}{\partial \boldsymbol{\beta}'} \right]' \left[\frac{\partial \boldsymbol{\theta}_{st}}{\partial \boldsymbol{\theta}'_s} \right]' \frac{\partial \ln L_{1st}(\boldsymbol{\theta}_{st}|\mathbf{N}_{st})}{\partial \boldsymbol{\theta}_{st}}. \quad (\text{A.7})$$

O primeiro termo da expressão pode ser obtido derivando (A.6) em relação a $\boldsymbol{\beta}'$

$$\begin{aligned} \frac{\partial \boldsymbol{\theta}_s}{\partial \boldsymbol{\beta}'} &= \frac{\mathbf{D}_{\exp(\mathbf{X}_s\boldsymbol{\beta})} \mathbf{X}_s \mathbf{1}'_R \exp(\mathbf{X}_s\boldsymbol{\beta}) - \exp(\mathbf{X}_s\boldsymbol{\beta}) \mathbf{1}'_R \mathbf{D}_{\exp(\mathbf{X}_s\boldsymbol{\beta})} \mathbf{X}_s}{[\mathbf{1}'_R \exp(\mathbf{X}_s\boldsymbol{\beta})]^2} \\ &= \mathbf{D}_{\boldsymbol{\theta}_s(\boldsymbol{\beta})} \mathbf{X}_s - \frac{\exp(\mathbf{X}_s\boldsymbol{\beta}) [\exp(\mathbf{X}_s\boldsymbol{\beta})]'}{[\mathbf{1}'_R \exp(\mathbf{X}_s\boldsymbol{\beta})]^2} \mathbf{X}_s \\ &= \{\mathbf{D}_{\boldsymbol{\theta}_s(\boldsymbol{\beta})} - \boldsymbol{\theta}_s(\boldsymbol{\beta}) [\boldsymbol{\theta}_s(\boldsymbol{\beta})]'\} \mathbf{X}_s. \end{aligned} \quad (\text{A.8})$$

O segundo termo é obtido derivando $\boldsymbol{\theta}_{st} = \mathbf{Z}'_{st}\boldsymbol{\theta}_s$ em relação a $\boldsymbol{\theta}_s$

$$\frac{\partial \boldsymbol{\theta}_{st}}{\partial \boldsymbol{\theta}'_s} = \mathbf{Z}'_{st}, \quad (\text{A.9})$$

e o terceiro termo, derivando (A.5) em relação a $\boldsymbol{\theta}_{st}$

$$\frac{\partial \ln L_{1st}(\boldsymbol{\theta}_{st}|\mathbf{N}_{st})}{\partial \boldsymbol{\theta}_{st}} = \mathbf{D}_{\boldsymbol{\theta}_{st}}^{-1} \mathbf{N}_{st}. \quad (\text{A.10})$$

Substituindo (A.8), (A.9) e (A.10) em (A.7), obtém-se

$$\begin{aligned}
\frac{\partial \ln L_{1st}(\boldsymbol{\theta}_s(\boldsymbol{\beta})|\mathbf{N}_{st})}{\partial \boldsymbol{\beta}} &= \mathbf{X}'_s \left\{ \mathbf{D}_{\boldsymbol{\theta}_s(\boldsymbol{\beta})} - \boldsymbol{\theta}_s(\boldsymbol{\beta}) [\boldsymbol{\theta}_s(\boldsymbol{\beta})]'\right\} \mathbf{Z}_{st} \mathbf{D}_{\mathbf{Z}'_{st} \boldsymbol{\theta}_s(\boldsymbol{\beta})}^{-1} \mathbf{N}_{st} \\
&= \mathbf{X}'_s \left\{ \mathbf{D}_{\boldsymbol{\theta}_s(\boldsymbol{\beta})} \mathbf{Z}_{st} \mathbf{D}_{\mathbf{Z}'_{st} \boldsymbol{\theta}_s(\boldsymbol{\beta})}^{-1} \mathbf{N}_{st} - \boldsymbol{\theta}_s(\boldsymbol{\beta}) [\boldsymbol{\theta}_s(\boldsymbol{\beta})]'\right\} \mathbf{Z}_{st} \mathbf{D}_{\mathbf{Z}'_{st} \boldsymbol{\theta}_s(\boldsymbol{\beta})}^{-1} \mathbf{N}_{st} \\
&= \mathbf{X}'_s \left\{ \mathbf{D}_{\boldsymbol{\theta}_s(\boldsymbol{\beta})} \mathbf{Z}_{st} \mathbf{D}_{\mathbf{Z}'_{st} \boldsymbol{\theta}_s(\boldsymbol{\beta})}^{-1} \mathbf{N}_{st} - \boldsymbol{\theta}_s(\boldsymbol{\beta}) \mathbf{1}'_{R_{st}} \mathbf{N}_{st} \right\} \\
&= \mathbf{X}'_s \left\{ \mathbf{D}_{\boldsymbol{\theta}_s(\boldsymbol{\beta})} \mathbf{Z}_{st} \mathbf{D}_{\mathbf{Z}'_{st} \boldsymbol{\theta}_s(\boldsymbol{\beta})}^{-1} \mathbf{N}_{st} - n_{st+} \boldsymbol{\theta}_s(\boldsymbol{\beta}) \right\},
\end{aligned}$$

de onde se obtém o resultado (2.73), já que $\sum_{t=1}^{T_s} n_{st+} \boldsymbol{\theta}_s(\boldsymbol{\beta}) = n_{s++} \boldsymbol{\theta}_s(\boldsymbol{\beta})$ e, como $\mathbf{Z}_{s1} = \mathbf{I}_R$, $\mathbf{D}_{\boldsymbol{\theta}_s(\boldsymbol{\beta})} \mathbf{Z}_{s1} \mathbf{D}_{\mathbf{Z}'_{s1} \boldsymbol{\theta}_s(\boldsymbol{\beta})}^{-1} \mathbf{N}_{s1} = \mathbf{N}_{s1}$.

Seja

$$\mathbf{S}_{1LLs}(\boldsymbol{\beta}) = \mathbf{X}'_s \left\{ \mathbf{N}_{s1} + \sum_{t=2}^{T_s} \left[\mathbf{D}_{\boldsymbol{\theta}_s(\boldsymbol{\beta})} \mathbf{Z}_{st} \mathbf{D}_{\mathbf{Z}'_{st} \boldsymbol{\theta}_s(\boldsymbol{\beta})}^{-1} \mathbf{N}_{st} \right] - n_{s++} \boldsymbol{\theta}_s(\boldsymbol{\beta}) \right\} \quad (\text{A.11})$$

a contribuição para o vetor *score* (2.73) da *s*-ésima subpopulação.

Procedendo à diferenciação adicional de (A.11) em relação a $\boldsymbol{\beta}'$, obtém-se

$$\begin{aligned}
\frac{\partial^2 \ln L_{1s}(\boldsymbol{\theta}_s(\boldsymbol{\beta})|\mathbf{N}_{st})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} &= \frac{\partial^2 \ln L_{1s}(\boldsymbol{\theta}_s(\boldsymbol{\beta})|\mathbf{N}_{st})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\theta}'_s} \frac{\partial \boldsymbol{\theta}_s}{\partial \boldsymbol{\beta}'} = \frac{\partial \mathbf{S}_{1LLs}(\boldsymbol{\beta})}{\partial \boldsymbol{\theta}'_s} \frac{\partial \boldsymbol{\theta}_s}{\partial \boldsymbol{\beta}'} \\
&= \mathbf{X}'_s \left\{ \sum_{t=2}^{T_s} \left[\mathbf{D}_{\mathbf{Z}_{st} \mathbf{D}_{\mathbf{Z}'_{st} \boldsymbol{\theta}_s(\boldsymbol{\beta})}^{-1} \mathbf{N}_{st}} - \mathbf{D}_{\mathbf{D}_{\boldsymbol{\theta}_s(\boldsymbol{\beta})} \mathbf{Z}_{st} \mathbf{D}_{\mathbf{Z}'_{st} \boldsymbol{\theta}_s(\boldsymbol{\beta})}^{-2} \mathbf{N}_{st}} \mathbf{Z}_{st} \mathbf{Z}'_{st} \right] - n_{s++} \mathbf{I}_R \right\} \\
&\quad \left\{ \mathbf{D}_{\boldsymbol{\theta}_s(\boldsymbol{\beta})} - \boldsymbol{\theta}_s(\boldsymbol{\beta}) [\boldsymbol{\theta}_s(\boldsymbol{\beta})]'\right\} \mathbf{X}_s,
\end{aligned}$$

de onde se chega facilmente ao resultado (2.74).

APÊNDICE B

Alocações de casos extremos

As alocações extremas das unidades com omissão que originaram os intervalos para os melhores-piores casos apresentados nos Exemplos 2 e 3 do Capítulo 3 são descritas, respectivamente, nas Seções B.1 e B.2.

B.1 Exemplo 2

As alocações extremas das unidades com omissão da Tabela 2.1 são exibidas nas Tabelas B.2 e B.3. A Tabela B.1 faz a correspondência entre as alocações e os limites inferior e superior dos intervalos para o melhor-pior caso para $\{\omega_{ij(s)}\}$ apresentados na Tabela 3.1.

Tabela B.1: Correspondências entre as alocações apresentadas nas Tabelas B.2 e B.3 e os limites inferior e superior dos intervalos para o melhor-pior caso para $\{\omega_{ij(s)}\}$ exibidos na Tabela 3.1 (p.95)

Limite	$\omega_{11(1)}$	$\omega_{12(1)}$	$\omega_{21(1)}$	$\omega_{22(1)}$	$\omega_{11(2)}$	$\omega_{12(2)}$	$\omega_{21(2)}$	$\omega_{22(2)}$
inferior	B	D	D	C	B	D	D	C
superior	A	E	F	A	A	E	F	A

Note que algumas alocações (A e D) conseguem contemplar simultaneamente 2 parâmetros de cada cidade e possuem todas as unidades com omissão distribuídas em alguma categoria, enquanto que outras alocações (B, C, E e F) conseguem mirar apenas 1 parâmetro de cada cidade por vez. Nestas últimas alocações, as unidades com omissão que não afetam o limite do intervalo para o melhor-pior caso do parâmetro que se analisa não são alocadas e as categorias associadas permanecem indicadas com interrogações, resultando, então, em alocações parciais.

Tabela B.2: Alocações extremas para as unidades com omissão da Tabela 2.1

Alocação Cidade		$Y_1 \setminus Y_2$	$t = 1$			$t = 2$				$t = 3$			Total			
			1	2	3	1	2	3	total	1	2	3	1	2	3	total
Dados observados	Kingston-H.	1	167	17	19	?	?	?	176	?	?	?	?	?	?	
		2	10	1	3	?	?	?	24	?	?	?	?	?	?	
		3	52	10	11	?	?	?	121	?	?	?	?	?	?	
	total									28	10	12			661	
	Portage	1	120	22	19	?	?	?	103	?	?	?	?	?	?	
		2	8	5	1	?	?	?	3	?	?	?	?	?	?	
3		39	12	12	?	?	?	80	?	?	?	?	?	?		
total									31	8	14			477		
A	Kingston-H.	1	167	17	19	176	0	0	176	28	0	0	371	17	19	407
		2	10	1	3	0	24	0	24	0	10	0	10	35	3	48
		3	52	10	11	0	0	121	121	0	0	12	52	10	144	206
	total									28	10	12	433	62	166	661
	Portage	1	120	22	19	103	0	0	103	31	0	0	254	22	19	295
		2	8	5	1	0	3	0	3	0	8	0	8	16	1	25
3		39	12	12	0	0	80	80	0	0	14	39	12	106	157	
total									31	8	14	301	50	126	477	
B	Kingston-H.	1	167	17	19	0	176	0	176	0	10	?	167	203	?	?
		2	10	1	3	24	0	0	24	28	0	?	62	1	?	?
		3	52	10	11	?	?	?	121	0	0	?	?	?	?	?
	total									28	10	12	?	?	?	661
	Portage	1	120	22	19	0	103	0	103	0	8	?	120	133	?	?
		2	8	5	1	3	0	0	3	31	0	?	42	5	?	?
3		39	12	12	?	?	?	80	0	0	?	?	?	?	?	
total									31	8	14	?	?	?	477	
C	Kingston-H.	1	167	17	19	?	?	?	176	?	0	0	?	?	?	?
		2	10	1	3	0	0	24	24	?	0	12	?	1	39	?
		3	52	10	11	0	121	0	121	?	10	0	?	141	11	?
	total									28	10	12	?	?	?	661
	Portage	1	120	22	19	?	?	?	103	?	0	0	?	?	?	?
		2	8	5	1	0	0	3	3	?	0	14	?	5	18	?
3		39	12	12	0	80	0	80	?	8	0	?	100	12	?	
total									31	8	14	?	?	?	477	

Tabela B.3: Alocações extremas para as unidades com omissão da Tabela 2.1

Alocação	Cidade	$Y_1 \setminus Y_2$	$t = 1$			$t = 2$				$t = 3$			Total			
			1	2	3	1	2	3	total	1	2	3	1	2	3	total
D	Kingston-H.	1	167	17	19	0	0	176	176	0	0	12	167	17	207	391
		2	10	1	3	0	24	0	24	0	10	0	10	35	3	48
		3	52	10	11	121	0	0	121	28	0	0	201	10	11	222
		total								28	10	12	378	62	221	661
	Portage	1	120	22	19	0	0	103	103	0	0	14	120	22	136	278
		2	8	5	1	0	3	0	3	0	8	0	8	16	1	25
3		39	12	12	80	0	0	80	31	0	0	150	12	12	174	
total									31	8	14	278	50	149	477	
E	Kingston-H.	1	167	17	19	0	176	0	176	?	10	0	?	203	19	?
		2	10	1	3	0	0	24	24	?	0	12	?	1	39	?
		3	52	10	11	?	?	?	121	?	0	0	?	?	?	?
		total								28	10	12	?	?	?	661
	Portage	1	120	22	19	0	103	0	103	?	8	0	?	133	19	?
		2	8	5	1	0	0	3	3	?	0	14	?	5	18	?
3		39	12	12	?	?	?	80	?	0	0	?	?	?	?	
total								31	8	14	?	?	?	477		
F	Kingston-H.	1	167	17	19	?	?	?	176	0	0	?	?	?	?	?
		2	10	1	3	24	0	0	24	28	0	?	62	1	?	?
		3	52	10	11	0	121	0	121	0	10	?	52	141	?	?
		total								28	10	12	?	?	?	661
	Portage	1	120	22	19	?	?	?	103	0	0	?	?	?	?	?
		2	8	5	1	3	0	0	3	31	0	?	42	5	?	?
3		39	12	12	0	80	0	80	0	8	?	39	100	?	?	
total								31	8	14	?	?	?	477		

B.2 Exemplo 3

As alocações extremas das unidades com omissão da Tabela 2.3 são exibidas na Tabela B.5. A Tabela B.4 faz a correspondência entre as alocações e os limites inferior e superior dos intervalos para o melhor-pior caso para os parâmetros de interesse apresentados na Tabela 3.11.

Tabela B.4: Correspondências entre as alocações apresentadas na Tabela B.5 e os limites inferior e superior dos intervalos para o melhor-pior caso para os parâmetros de interesse exibidos na Tabela 3.11 (p.108)

Limite	$\pi_{1+} - \pi_{+1}$	$\pi_{2+} - \pi_{+2}$	κ	κ_{w1}	κ_{w2}
inferior	B	A	D	D	D
superior	A	B	C	C	C

Tabela B.5: Alocações extremas para as unidades com omissão da Tabela 2.3

Alocação	Métodos	t = 1			t = 2			t = 3			Total				
		simpl.\conv.	alto	médio	baixo	alto	médio	baixo	alto	médio	baixo	alto	médio	baixo	total
Dados obs.	alto		7	11	2	?	?	?	⊗	⊗	⊗	?	?	?	
	médio		3	9	5	?	?	?	?	?	?	?	?	?	
	baixo		0	10	4	⊗	⊗	⊗	?	?	?	?	?	?	
	total					8	7	3	7	14	7				97
A	alto		7	11	2	8	7	3	⊗	⊗	⊗	15	18	5	38
	médio		3	9	5	0	0	0	0	0	0	3	9	5	17
	baixo		0	10	4	⊗	⊗	⊗	7	14	7	7	24	11	42
	total					8	7	3	7	14	7	25	51	21	97
B	alto		7	11	2	0	0	0	⊗	⊗	⊗	7	11	2	20
	médio		3	9	5	8	7	3	7	14	7	18	30	15	63
	baixo		0	10	4	⊗	⊗	⊗	0	0	0	0	10	4	14
	total					8	7	3	7	14	7	25	51	21	97
C	alto		7	11	2	8	0	0	⊗	⊗	⊗	15	11	2	28
	médio		3	9	5	0	7	3	7	14	0	10	30	8	48
	baixo		0	10	4	⊗	⊗	⊗	0	0	7	0	10	11	21
	total					8	7	3	7	14	7	25	51	21	97
D	alto		7	11	2	0	7	3	⊗	⊗	⊗	7	18	5	30
	médio		3	9	5	8	0	0	0	0	7	11	9	12	32
	baixo		0	10	4	⊗	⊗	⊗	7	14	0	7	24	4	35
	total					8	7	3	7	14	7	25	51	21	97

⊗ representa zeros estruturais, simpl.: simplificado, conv.: convencional.

Referências bibliográficas

- AGRESTI, A. (2002). *Categorical data analysis*. 2^a ed. New York: John Wiley & Sons.
- ALBERT, P.S. (2000). A transitional model for longitudinal binary data subject to nonignorable missing data. *Biometrics* **56**, 602-608.
- ALBERT, P.S., FOLLMANN, D.A., WANG, S.A. e SUH, E.B. (2002). A latent autoregressive model for longitudinal binary data subject to informative missingness. *Biometrics* **58**, 631-642.
- ALLISON, P.D. (2001). *Missing data*. Thousand Oaks: Sage.
- ALMEIDA, A.C. (2002). *Como são feitas as pesquisas eleitorais e de opinião*. Rio de Janeiro: FGV.
- ALONZO, T.A. (2005). Verification bias-corrected estimators of the relative true and false positive rates of two binary screening tests. *Statistics in Medicine* **24**, 403-417.
- ANDERSEN, E.B. (1994). *The statistical analysis of categorical data*. 3^a ed. Berlin: Springer-Verlag.
- ANDERSON, T.W. (2003). *An introduction to multivariate statistical analysis*. 3^a ed. New York: John Wiley & Sons.
- ANDRADE, D.F., TAVARES, H.R. e VALLE, R.C. (2000). *Teoria da resposta ao item: conceitos e aplicações*. São Paulo: Associação Brasileira de Estatística.
- ANDRÉ, C.D.S., NEVES, M.M.C. e TSENG, T.H. (1990). *Relatório de análise estatística sobre o projeto: "Estudo comparativo entre os diferentes métodos de detecção de indivíduos com alto risco de cárie"*. São Paulo: IME-USP (RAE-CEA-90P08).
- ARTES, R. e BOTTER, D.A. (2005). *Funções de estimação em modelos de regressão*. São Paulo: Associação Brasileira de Estatística.
- AZEVEDO, C.L.N. (2003). *Métodos de estimação na teoria de resposta ao item*. Dissertação de mestrado. IME-USP.
- BAKER, S.G. (1992). A simple method for computing the observed information matrix when using the EM algorithm with categorical data. *Journal of Computational and Graphical Statistics* **1**, 63-73, 180 (correção).
- BAKER, S.G. (1994a). Missing data: composite linear models for incomplete multinomial data. *Statistics in Medicine* **13**, 609-622.
- BAKER, S.G. (1994b). Regression analysis of grouped survival data with incomplete covariates: nonignorable missing-data and censoring mechanisms. *Biometrics* **50**, 821-826.

- BAKER, S.G. (1995a). Evaluating multiple diagnostic tests with partial verification. *Biometrics* **51**, 330-337.
- BAKER, S.G. (1995b). Marginal regression for repeated binary data with outcome subject to nonignorable nonresponse. *Biometrics* **51**, 1042-1052.
- BAKER, S.G. (1996). The analysis of categorical case-control data subject to nonignorable nonresponse. *Biometrics* **52**, 362-369.
- BAKER, S.G. (2000). Analyzing a randomized cancer prevention trial with a missing binary outcome and an auxiliary variable and all-or-none compliance. *Journal of the American Statistical Association* **95**, 43-50.
- BAKER, S.G., KO, C.-W. e GRAUBARD, B.I. (2003). A sensitivity analysis for nonrandomly missing categorical data arising from a national health disability survey. *Biostatistics* **4**, 41-56.
- BAKER, S.G. e LAIRD, N.M. (1988). Regression analysis for categorical variables with outcome subject to nonignorable nonresponse. *Journal of the American Statistical Association* **83**, 62-69, 1232 (correção).
- BAKER, S.G., ROSENBERGER, W.F. e DERSIMONIAN, R. (1992). Closed-form estimates for missing counts in two-way contingency tables. *Statistics in Medicine* **11**, 643-657.
- BARNHART, H.X. e KOSINSKI, A.S. (2003). Evaluating medical diagnostic tests at the subunit level in the presence of verification bias. *Statistics in Medicine* **22**, 2161-2176.
- BASU, D. (1977). On the elimination of nuisance parameters. *Journal of the American Statistical Association* **72**, 355-366.
- BASU, D. e PEREIRA, C.A.B. (1982). On the bayesian analysis of categorical data: the problem of nonresponse. *Journal of Statistical Planning and Inference* **6**, 345-362.
- BEGG, C.B. e GREENES, R.A. (1983). Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics* **39**, 206-215.
- BIRMINGHAM, J. e FITZMAURICE, G.M. (2002). A pattern-mixture model for longitudinal binary responses with nonignorable nonresponse. *Biometrics* **58**, 989-996.
- BIRMINGHAM, J., ROTNITZKY, A. e FITZMAURICE, G.M. (2003). Pattern-mixture and selection models for analysing longitudinal data with monotone missing patterns. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **65**, 275-297.
- BISHOP, Y.M.M., FIENBERG, S.E. e HOLLAND, P.W. (1975). *Discrete multivariate analysis: theory and practice*. Cambridge: The MIT Press.
- BLACKHURST, D.W. e SCHLUCHTER, M.D. (1989). Logistic regression with a partially observed covariate. *Communications in Statistics - Simulation and Computation* **18**, 163-177.
- BLUMENTHAL, S. (1968). Multinomial sampling with partially categorized data. *Journal of the American Statistical Association* **63**, 542-551.
- BOLFARINE, H. e BUSSAB, W.O. (2005). *Elementos de amostragem*. São Paulo: Edgard Blücher.
- BOLLEN, K.A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.

- BONETTI, M., COLE, B.F. e GELBER, R.D. (1999). A method-of-moments estimation procedure for categorical quality-of-life data with nonignorable missingness. *Journal of the American Statistical Association* **94**, 1025-1034.
- BOTTAI, M. (2003). Confidence regions when the Fisher information is zero. *Biometrika* **90**, 73-84.
- BROWN, C.H. (1990). Protecting against nonrandomly missing data in longitudinal studies. *Biometrics* **46**, 143-156.
- CHAMBERS, R.L. e SKINNER, C.J. (2003). *Analysis of survey data*. New York: John Wiley & Sons.
- CHAMBERS, R.L. e WELSH, A.H. (1993). Log-linear models for survey data with non-ignorable non-response. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **55**, 157-170.
- CHEN, H.Y. e LITTLE, R.J.A. (1999). A test of missing completely at random for generalised estimating equations with missing data. *Biometrika* **86**, 1-13.
- CHEN, T.T. (1988). Modelling the liver function data in diabetes. *Statistics in Medicine* **7**, 671-683.
- CHEN, T.T. e FIENBERG, S.E. (1974). Two-dimensional contingency tables with both completely and partially cross-classified data. *Biometrics* **30**, 629-642.
- CHEN, T.T. e FIENBERG, S.E. (1976). The analysis of contingency tables with incompletely classified data. *Biometrics* **32**, 133-144.
- CHOI, S.C. e STABLEIN, D.M. (1982). Practical test for comparing two proportions with incomplete data. *Journal of the Royal Statistical Society. Series C: Applied Statistics* **31**, 256-262.
- CHOI, S.C. e STABLEIN, D.M. (1988). Comparing incomplete paired binomial data under non-random mechanisms. *Statistics in Medicine* **7**, 929-939.
- CICCHETTI, D.V. e ALLISON, T. (1971). A new procedure for assessing reliability of scoring EEG sleep recordings. *American Journal of EEG Technology* **11**, 101-109.
- CLARKE, P.S. (2002). On boundary solutions and identifiability in categorical regression with non-ignorable non-response. *Biometrical Journal* **44**, 701-717.
- CLARKE, P.S. e SMITH, P.W.F. (2004). Interval estimation for log-linear models with one variable subject to non-ignorable non-response. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **66**, 357-368.
- COCHRAN, W.G. (1977). *Sampling techniques*. 3^a ed. New York: John Wiley & Sons.
- COHEN, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* **20**, 37-46.
- COLOSIMO, E.A. e GIOLO, S.R. (2006). *Análise de sobrevivência aplicada*. São Paulo: Edgard Blücher.
- CONAWAY, M.R. (1992). The analysis of repeated categorical measurements subject to nonignorable nonresponse. *Journal of the American Statistical Association* **87**, 817-824.
- CONAWAY, M.R. (1993). Non-ignorable non-response models for time-ordered categorical variables. *Journal of the Royal Statistical Society. Series C: Applied Statistics* **42**, 105-115.

- CONAWAY, M.R. (1994). Causal nonresponse models for repeated categorical measurements. *Biometrics* **50**, 1102-1116.
- CONAWAY, M.R., WATERNAUX, C., ALFRED, E., BELLINGER, D. e LEVINTON, A. (1992). Pre-natal blood lead levels and learning difficulties in children: an analysis of non-randomly missing categorical data. *Statistics in Medicine* **11**, 799-811.
- COOK, R.J., ZENG, L. e YI, G.Y. (2004). Marginal analysis of incomplete longitudinal binary data: a cautionary note on LOCF imputation. *Biometrics* **60**, 820-828.
- COPAS, J. e EGUCHI, S. (2001). Local sensitivity approximations for selectivity bias. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **63**, 871-895.
- CORDEIRO, G.M. e LIMA NETO, E.A. (2004). *Modelos paramétricos*. São Paulo: Associação Brasileira de Estatística.
- DAVISON, A.C. e HINKLEY, D.V. (1997). *Bootstrap methods and their application*. Cambridge: Cambridge University Press.
- DARROCH, J.N. e RATCLIFF, D. (1972). Generalized iterative scaling for log-linear models. *The Annals of Mathematical Statistics* **43**, 1470-1480.
- DAWID, A.P. e DICKEY, J.M. (1977). Likelihood and bayesian inference from selectively reported data. *Journal of the American Statistical Association* **72**, 845-850.
- DELTOUR, I., RICHARDSON, S. e LE HESRAN, J.-Y. (1999). Stochastic algorithms for Markov models estimation with intermittent missing data. *Biometrics* **55**, 565-573.
- DEMPSTER, A.P., LAIRD, N.M. e RUBIN, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **39**, 1-38.
- DICKEY, J.M., JIANG, J.-M. e KADANE, J.B. (1987). Bayesian methods for censored categorical data. *Journal of the American Statistical Association* **87**, 773-781.
- DIGGLE, P.J., HEAGERTY, P., LIANG, K.-Y. e ZEGER, S.L. (2002). *Analysis of longitudinal data*. 2^a ed. Oxford: Oxford University Press.
- DIGGLE, P.J. e SHIH, W.J. (1993). On informative and random dropouts in longitudinal studies (letter to the editor). *Biometrics* **49**, 947-949.
- EFRON, B. e TIBSHIRANI, R.J. (1994). *An introduction to the bootstrap*. Boca Raton: Chapman & Hall / CRC.
- EKHOLM, A. e SKINNER, C.J. (1998). The muscatine children's obesity data reanalysed using pattern mixture models. *Journal of the Royal Statistical Society. Series C: Applied Statistics* **47**, 251-263.
- ELASHOFF, J.D. e ELASHOFF, R.M. (1974). Two sample problems for a dichotomous variable with missing data. *Journal of the Royal Statistical Society. Series C: Applied Statistics* **23**, 26-34.
- EMBRAPA (1997). Centro Nacional de Pesquisa Tecnológica em Informática para a Agricultura. *Ambiente de software NTIA, versão 4.2.2*. Campinas.
- FAY, R.E. (1986). Causal models for patterns of nonresponse. *Journal of the American Statistical Association* **81**, 354-365.

- FEELDERS, A.J. (2000). Credit scoring and reject inference with mixture models. *International Journal of Intelligent Systems in Accounting, Finance & Management* **9**, 1-8.
- FITZGERALD, P.E.B. (2002). Extended generalized estimating equations for binary familial data with incomplete families. *Biometrics* **58**, 718-726.
- FITZMAURICE, G.M., HEATH, A.F. e CLIFFORD, P. (1996). Logistic regression models for binary panel data with attrition. *Journal of the Royal Statistical Society. Series A: Statistics in Society* **159**, 249-263.
- FITZMAURICE, G.M., LAIRD, N.M. e LIPSITZ, S.R. (1994). Analyzing incomplete longitudinal binary responses: a likelihood-based approach. *Biometrics* **50**, 601-612.
- FITZMAURICE, G.M., LAIRD, N.M. e ZAHNER, G.E.P. (1996). Multivariate logistic models for incomplete binary responses. *Journal of the American Statistical Association* **91**, 99-108.
- FITZMAURICE, G.M., LIPSITZ, S.R., MOLENBERGHS, G. e IBRAHIM, J.G. (2001). Bias in estimating association parameters for longitudinal binary responses with drop-outs. *Biometrics* **57**, 15-21.
- FITZMAURICE, G.M., MOLENBERGHS, G. e LIPSITZ, S.R. (1995). Regression models for longitudinal binary responses with informative drop-outs. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **57**, 691-704.
- FLEISS, J.L. e COHEN, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement* **33**, 613-619.
- FLEISS, J.L., LEVIN, B. e PAIK, M.C. (2003). *Statistical methods for rates and proportions*. 3ª ed. New York: John Wiley & Sons.
- FORSTER, J.J. e SMITH, W.F. (1998). Model-based inference for categorical survey data subject to non-ignorable non-response. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **60**, 57-70.
- FORTHOFER, R.N. e LEHNEN, R.G. (1981). *Public program analysis: a new categorical data approach*. Belmont: Wadsworth.
- FUCHS, C. (1982). Maximum likelihood estimation and model selection in contingency tables with missing data. *Journal of the American Statistical Association* **77**, 270-278.
- GARCIA, M.G. (2003). *Imputação de dados categorizados usando o modelo multinomial*. Dissertação de mestrado. IME-USP.
- GIBBONS, L.E. e HOSMER, D.W. (1991). Conditional logistic regression with missing data. *Communications in Statistics - Simulation and Computation* **20**, 109-120.
- GLONEK, G.F.V. (1999). On identifiability in models for incomplete binary data. *Statistics & Probability Letters* **41**, 191-197.
- GLYNN, R.J., LAIRD, N.M. e RUBIN, D.B. (1986). Selection modeling versus mixture modeling with nonignorable nonresponse (with discussion). Em *Drawing inferences from self-selected samples*, ed.H.Wainer. Lawrence Erlbaum Associates, Mahwah. 115-151.

- GOETGHEBEUR, E. e MOLENBERGHS, G. (1996). Estimating efficacy in a placebo-controlled clinical trial with binary outcome and ordered compliance. *Journal of the American Statistical Association* **91**, 928-934.
- GOETGHEBEUR, E., MOLENBERGHS, G. e KATZ, J. (1998). Estimating the causal effect of compliance on binary outcome in randomized controlled trials. *Statistics in Medicine* **17**, 341-355.
- GOETGHEBEUR, E. e SHAPIRO, S. (1996). Analyzing non-compliance in clinical trials: ethical imperative or mission impossible?. *Statistics in Medicine* **15**, 2813-2826.
- GREEN, P.E. e PARK, T. (2003). A bayesian hierarchical model for categorical data with nonignorable nonresponse. *Biometrics* **59**, 886-896.
- GRIZZLE, J.E., STARMER, C.F. e KOCH, G.G. (1969). Analysis of categorical data by linear models. *Biometrics* **25**, 489-504.
- GROVES, R.M., DILLMAN, D.A., ELTINGE, J.L. e LITTLE, R.J.A. (2002). *Survey nonresponse*. New York: John Wiley & Sons.
- GUNEL, E. (1984). A bayesian analysis of the multinomial model for a dichotomous response with nonrespondents. *Communications in Statistics - Theory and Methods* **13**, 737-751.
- HABER, M., CHEN, C.C.H. e WILLIAMSON, G.D. (1991). Analysis of repeated categorical responses from fully and partially cross-classified data. *Communications in Statistics - Theory and Methods* **20**, 3293-3313.
- HAGENAARS, J.A. e MCCUTCHEON, A.L. (2002). *Applied latent class analysis*. Cambridge: Cambridge University Press.
- HANCOCK, G.R. e MUELLER, R.O. (2006). *Structural equation modeling: a second course*. Connecticut: Information Age Publishing.
- HARTLEY, H.O. e HOCKING, R.R. (1971). The analysis of incomplete data. *Biometrics* **27**, 783-823.
- HEITJAN, D.F. (1993). Ignorability and coarse data: some biomedical examples. *Biometrics* **49**, 1099-1109.
- HEITJAN, D.F. (1994). Ignorability in general incomplete-data models. *Biometrika* **81**, 701-708.
- HEITJAN, D.F. (1997). Ignorability, sufficiency and ancillarity. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **59**, 375-381.
- HEITJAN, D.F. e RUBIN, D.B. (1991). Ignorability and coarse data. *The Annals of Statistics* **19**, 2244-2253.
- HENS, N. (2005). *Non- and semi-parametric techniques for handling missing data*. Tese de doutorado. Facultait Wetenschappen, Limburgs Universitair Centrum.
- HEYTING, A., TOLBOOM, J.T.B.M. e ESSERS, J.G.A. (1992). Statistical handling of drop-outs in longitudinal clinical trials. *Statistics in Medicine* **11**, 2043-2061.
- HOCKING, R.R. e OXSPRING, H.H. (1971). Maximum likelihood estimation with incomplete multinomial data. *Journal of the American Statistical Association* **66**, 65-70.

- HOCKING, R.R. e OXSPRING, H.H. (1974). The analysis of partially categorized contingency data. *Biometrics* **30**, 469-483.
- HO, L.L. e SINGER, J.M. (2001). Generalized least squares methods for bivariate Poisson regression. *Communications in Statistics, Theory and Methods* **30**, 263-277.
- HOLMAN, R. e GLAS, C.A.W. (2005). Modelling non-ignorable missing-data mechanisms with item response theory models. *British Journal of Mathematical and Statistical Psychology* **58**, 1-17.
- HORTON, N.J. e FITZMAURICE, G.M. (2002). Maximum likelihood estimation of bivariate logistic models for incomplete responses with indicators of ignorable and non-ignorable missingness. *Journal of the Royal Statistical Society. Series C: Applied Statistics* **51**, 281-295.
- HORTON, N.J. e LAIRD, N.M. (1998). Maximum likelihood analysis of generalized linear models with missing covariates. *Statistical Methods in Medical Research* **8**, 37-50.
- HORTON, N.J. e LAIRD, N.M. (2001). Maximum likelihood analysis of logistic regression models with incomplete covariate data and auxiliary information. *Biometrics* **57**, 34-42.
- HOYLE, R.H. (1995). *Structural equation modeling: concepts, issues and applications*. Thousand Oaks: Sage.
- IBRAHIM, J.G. (1990). Incomplete data in generalized linear models. *Journal of the American Statistical Association* **85**, 765-769.
- IBRAHIM, J.G. e LIPSITZ, S.R. (1996). Parameter estimation from incomplete data in binomial regression when the missing data mechanism is non-ignorable. *Biometrics* **52**, 1071-1078.
- IBRAHIM, J.G., LIPSITZ, S.R. e CHEN, M.-H. (1999). Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **61**, 173-190.
- IBRAHIM, J.G., LIPSITZ, S.R. e HORTON, N.J. (2001). Using auxiliary data for parameter estimation with non-ignorably missing outcomes. *Journal of the Royal Statistical Society. Series C: Applied Statistics* **50**, 361-373.
- IMREY, P.B., KOCH, G.G., STOKES, M.E. *et al.* (1981). Categorical data analysis: some reflections on the log linear model and logistic regression. Part I: historical and methodological overview. *International Statistical Review* **49**, 265-283.
- IMREY, P.B., KOCH, G.G., STOKES, M.E. *et al.* (1982). Categorical data analysis: some reflections on the log linear model and logistic regression. Part II: data analysis. *International Statistical Review* **50**, 35-63.
- KADANE, J.B. (1993). Subjective bayesian analysis for surveys with missing data. *Journal of the Royal Statistical Society. Series D: The Statistician* **42**, 415-426.
- KATERI, M.M. e ILIOPOULOS, G.M. (2004). On collapsing categories in two-way contingency tables. *Statistics: A Journal of Theoretical and Applied Statistics* **37**, 443-455.
- KENWARD, M.G., GOETGHEBEUR, E. e MOLENBERGHS, G. (2001). Sensitivity analysis for incomplete categorical data. *Statistical Modelling* **1**, 31-48.

- KENWARD, M.G., LESAFFRE, E. e MOLENBERGHS, G. (1994). An application of maximum likelihood and estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random. *Biometrics* **50**, 945-953.
- KENWARD, M.G. e MOLENBERGHS, G. (1998). Likelihood based frequentist inference when data are missing at random. *Statistical Science* **13**, 236-247.
- KENWARD, M.G. e MOLENBERGHS, G. (1999). Parametric models for incomplete continuous and categorical longitudinal data. *Statistical Methods in Medical Research* **8**, 51-83.
- KING, G., HONAKER, J., JOSEPH, A. e SCHEVE, K. (2001). Analyzing incomplete political science data: an alternative algorithm for multiple imputation. *American Political Science Review* **95**, 49-69.
- KISH, L. (1965). *Survey sampling*. New York: John Wiley & Sons.
- KLEIN, J.P. e MOESCHBERGER, M.L. (2003). *Survival analysis: techniques for censored and truncated data*. 2ª ed. New York: Springer-Verlag.
- KLINE, R.B. (2005). *Principles and practice of structural equation modeling*. 2ª ed. New York: The Guilford Press.
- KOCH, G.G., IMREY, P.B. e REINFURT, D.W. (1972). Linear model analysis of categorical data with incomplete response vectors. *Biometrics* **28**, 663-692.
- KOCH, G.G., IMREY, P.B., SINGER, J.M., ATKINSON, S.S. e STOKES, M.E. (1985). *Analysis of categorical data*. Montréal: Les Presses de L'Université de Montréal.
- KOCH, G.G., SINGER, J.M. e AMARA, I.A. (1985). A two-stage procedure for the analysis of ordinal categorical data. Em *Biostatistics: Statistics in Biomedical, Public Health and Environmental Sciences*, ed. P.K. Sen. North Holland: Elsevier Science. 357-387.
- KOOREMAN, P. (1993). Bounds on the regression coefficients when a covariate is categorized. *Communications in Statistics - Theory and Methods* **22**, 2373-2380.
- KOSINSKI, A.S. e BARNHART, H.X. (2003a). Accounting for nonignorable verification bias in assessment of diagnostic tests. *Biometrics* **59**, 163-171.
- KOSINSKI, A.S. e BARNHART, H.X. (2003b). A global sensitivity analysis of performance of a medical diagnostic test when verification bias is present. *Statistics in Medicine* **22**, 2711-2721.
- LAIRD, N.M. (1988). Missing data in longitudinal studies. *Statistics in Medicine* **7**, 305-315.
- LATIF, S.A. (2000). *Modelagem de equações estruturais*. Dissertação de mestrado. IME-USP.
- LEHMANN, E.L. (1998). *Elements of large-sample theory*. New York: Springer-Verlag.
- LEHMANN, E.L. e CASELLA, G. (1998). *Theory of point estimation*. 2ª ed. New York: Springer-Verlag.
- LEHTONEN, R. e PAHKINEN, E. (2004). *Practical methods for design and analysis of complex surveys*. 2ª ed. New York: John Wiley & Sons.
- LESAFFRE, E., MOLENBERGHS, G. e DEWULF, L. (1996). Effect of dropouts in a longitudinal study: an application of a repeated ordinal model. *Statistics in Medicine* **15**, 1123-1141.

- LIANG, K.-Y. e QIN, J. (2000). Regression analysis under non-standard situations: a pairwise pseudolikelihood approach. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **62**, 773-786.
- LIANG, K.-Y. e ZEGER, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.
- LIN, I.-F. e PAIK, M.C. (2001). Matched case-control data analysis with selection bias. *Biometrics* **57**, 1106-1112.
- LIN, H., MCCULLOCH, C.E. e ROSENHECK, R.A. (2004). Latent pattern mixture models for informative intermittent missing data in longitudinal studies. *Biometrics* **60**, 295-305.
- LINDERN, W.J. e HAMBLETON, R.K. (1996). *Handbook of modern item response theory*. New York: Springer-Verlag.
- LINDSEY, J.K. (1996). *Parametric statistical inference*. Oxford: Oxford University Press.
- LINDSEY, J.K. (1999). *Models for repeated measurements*. 2^a ed. Oxford: Oxford University Press.
- LIPSITZ, S.R. e FITZMAURICE, G.M. (1996). The score test for independence in $R \times C$ contingency tables with missing data. *Biometrics* **52**, 751-762.
- LIPSITZ, S.R. e IBRAHIM, J.G. (1996). A conditional model for incomplete covariates in parametric regression models. *Biometrika* **83**, 916-922.
- LIPSITZ, S.R., IBRAHIM, J.G., CHEN, M.-H. e PETERSON, H. (1999). Non-ignorable missing covariates in generalized linear models. *Statistics in Medicine* **18**, 2435-2448.
- LIPSITZ, S.R., IBRAHIM, J.G. e FITZMAURICE, G.M. (1999). Likelihood methods for incomplete longitudinal binary responses with incomplete categorical covariates. *Biometrics* **55**, 214-223.
- LIPSITZ, S.R., IBRAHIM, J.G. e ZHAO, L.P. (1999). A weighted estimating equation for missing covariate data with properties similar to maximum likelihood. *Journal of the American Statistical Association* **94**, 1147-1160.
- LIPSITZ, S.R., LAIRD, N.M. e HARRINGTON, D.P. (1994). Weighted least squares analysis of repeated categorical measurements with outcomes subject to nonresponse. *Biometrics* **50**, 11-24.
- LIPSITZ, S.R., PARZEN, M. e EWELL, M. (1998). Inference using conditional logistic regression with missing covariates. *Biometrics* **54**, 295-303.
- LIPSITZ, S.R., ZHAO, L.P. e MOLENBERGHS, G. (1998). A semiparametric method of multiple imputation. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **60**, 127-144.
- LITTLE, R.J.A. (1982). Models for nonresponse in sample surveys. *Journal of the American Statistical Association* **77**, 237-250.
- LITTLE, R.J.A. (1983). Superpopulation models for nonresponse. Em *Incomplete Data in Sample Surveys 2: Theory and Bibliographies*, eds. W.G.Madow, I.Olin e D.B.Rubin. Academic Press, New York. 335-413.

- LITTLE, R.J.A. (1985). Nonresponse adjustments in longitudinal surveys: models for categorical data. *Bulletin of the International Statistical Institute*, Proceedings of the 45th Session: Invited Papers, Section 15, 1-17.
- LITTLE, R.J.A. (1992). Regression with missing X's: a review. *Journal of the American Statistical Association* **87**, 1227-1237.
- LITTLE, R.J.A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* **88**, 125-134.
- LITTLE, R.J.A. (1995). Modeling the drop-out mechanism in repeated measures studies. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **90**, 1112-1121.
- LITTLE, R.J.A. e RUBIN, D.B. (2002). *Statistical analysis with missing data*. 2^a ed. New York: John Wiley & Sons.
- LITTLE, R.J.A. e SCHLUCHTER, M.D. (1985). Maximum likelihood for mixed continuous and categorical data with missing values. *Biometrika* **72**, 497-512.
- LIU, C.H. e RUBIN, D.B. (1994). The ECME algorithm: a simple extension of EM and ECM with fast monotone convergence. *Biometrika* **81**, 633-648.
- LIU, C.H., RUBIN, D.B. e WU, Y.N. (1998). Parameter expansion to accelerate EM: the PX-EM algorithm. *Biometrika* **85**, 755-770.
- LIU, X., WATERNAUX, C. e PETKOVA, E. (1999). Influence of human immunodeficiency virus infection on neurological impairment: an analysis of longitudinal binary data with informative drop-out. *Journal of the Royal Statistical Society. Series C: Applied Statistics* **48**, 103-115.
- LOUIS, T.A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **44**, 226-233.
- LYLES, R.H. e ALLEN, A.S. (2003). Missing data in the 2×2 table: patterns and likelihood-based analysis for cross-sectional studies with supplemental sampling. *Statistics in Medicine* **22**, 517-534.
- MADOW, W.G. e OLKIN, I. (1983). *Incomplete data in sample surveys. Vol.3: proceedings of the symposium*. New York: Academic Press.
- MADOW, W.G., NISSELSON, H. e OLKIN, I. (1983). *Incomplete data in sample surveys. Vol.1: report and case studies*. New York: Academic Press.
- MADOW, W.G., OLKIN, I. e RUBIN, D.B. (1983). *Incomplete data in sample surveys. Vol.2: theory and bibliographies*. New York: Academic Press.
- MARK, S.D. e GAIL, M.H. (1994). A comparison of likelihood-based and marginal estimating equation methods for analyzing repeated ordered categorical responses with missing data: application to an intervention trial of vitamin prophylaxis for oesophageal dysplasia. *Statistics in Medicine* **13**, 479-493.
- MAYS, E. (2001). *Handbook of credit scoring*. Chicago: Glenlake Publishing Company.
- MAYS, E. (2003). *Credit scoring for risk managers: the handbook for lenders*. Mason: South-Western.

- MCCULLAGH, P. e NELDER, J.A. (1989). *Generalized linear models*. 2^a ed. Boca Raton: Chapman & Hall / CRC.
- MCLACHLAN, G.J. e KRISHNAN, T. (1996). *The EM algorithm and extensions*. New York: John Wiley & Sons.
- MEILIJSON, I. (1989). A fast improvement to the EM algorithm on its own terms. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **51**, 127-138.
- MELHADO, T.T. (2004). *Medidas de ajuste de modelos de equações estruturais*. Dissertação de mestrado. IME-USP.
- MENG, X.-L. (2000). Missing data: dial M for ????. *Journal of the American Statistical Association* **95**, 1325-1330.
- MENG, X.-L. e RUBIN, D.B. (1991). Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **86**, 899-909.
- MENG, X.-L. e RUBIN, D.B. (1993). Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* **80**, 267-278.
- MICHIELS, B. e MOLENBERGHS, G. (1997). Protective estimation of longitudinal categorical data with nonrandom dropout. *Communications in Statistics - Theory and Methods* **26**, 65-94.
- MICHIELS, B., MOLENBERGHS, G. e LIPSITZ, S.R. (1999). Selection models and pattern-mixture models for incomplete data with covariates. *Biometrics* **55**, 978-983.
- MILLER, M.E., TEN HAVE, T.R., REBOUSSIN, B.A., LOHMAN, K.K. e REJESKI, W.J. (2001). A marginal model for analyzing discrete outcomes from longitudinal surveys with outcomes subject to multiple-cause nonresponse. *Journal of the American Statistical Association* **96**, 844-857.
- MOLENBERGHS, G. e GOETGHEBEUR, E. (1997). Simple fitting algorithms for incomplete categorical data. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **59**, 401-414.
- MOLENBERGHS, G., GOETGHEBEUR, E., LIPSITZ, S.R. e KENWARD, M.G. (1999). Nonrandom missingness in categorical data: strengths and limitations. *The American Statistician* **53**, 110-118.
- MOLENBERGHS, G., KENWARD, M.G. e GOETGHEBEUR, E. (2001). Sensitivity analysis for incomplete contingency tables: the Slovenian plebiscite case. *Journal of the Royal Statistical Society. Series C: Applied Statistics* **50**, 15-29.
- MOLENBERGHS, G., KENWARD, M.G. e LESAFFRE, E. (1997). The analysis of longitudinal ordinal data with nonrandom drop-out. *Biometrika* **84**, 33-44.
- MOLENBERGHS, G., MICHIELS, B. e KENWARD, M.G. (1998). Pseudo-likelihood for combined selection and pattern-mixture models for incomplete data. *Biometrical Journal* **40**, 557-572.
- MOLENBERGHS, G., MICHIELS, B., KENWARD, M.G. e DIGGLE, P.J. (1998). Monotone missing data and pattern mixture models. *Statistica Neerlandica* **52**, 153-161.

- MOLENBERGHS, G. e VERBEKE, G. (2005). *Models for discrete longitudinal data*. New York: Springer-Verlag.
- MURRAY, G.D. e FINDLAY, J.G. (1988). Correcting for the bias caused by drop-outs in hypertension trials. *Statistics in Medicine* **7**, 941-946.
- NELDER, J.A. e WEDDERBURN, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A: Statistics in Society* **135**, 370-383.
- NORDHEIM, E.V. (1984). Inference from nonrandomly missing categorical data: an example from a genetic study on Turner's syndrome. *Journal of the American Statistical Association* **79**, 772-780.
- OLESON, J.J. e HE, C.Z. (2004). Hierarchical bayesian modeling in dichotomous processes in the presence of nonresponse. *Biometrics* **60**, 50-59.
- ORCHARD, T. e WOODBURY, M.A. (1972). A missing information principle: theory and applications. Em *Proceedings of the 6th Berkeley Symposium on Mathematics, Statistics and Probability* **1**, 697-715.
- PAIK, M.C. (1997). The generalized estimating equation approach when data are not missing completely at random. *Journal of the American Statistical Association* **92**, 1320-1329.
- PAIK, M.C. (2004). Nonignorable missingness in matched case-control data analyses. *Biometrics* **60**, 306-314.
- PAIK, M.C. e SACCO, R. (2000). Matched case-control data analyses with missing covariates. *Journal of the Royal Statistical Society. Series C: Applied Statistics* **49**, 145-156.
- PAIK, M.C., SACCO, R. e LIN, I.-F. (2000). Bivariate binary data analysis with nonignorablely missing outcomes. *Biometrics* **56**, 1145-1156.
- PARK, T. (1998). An approach to categorical data with nonignorable nonresponse. *Biometrics* **54**, 1579-1590.
- PARK, T. e BROWN, M.B. (1994). Models for categorical data with nonignorable nonresponse. *Journal of the American Statistical Association* **89**, 44-52.
- PAULA, G.A. (2004). *Modelos de regressão com apoio computacional*. São Paulo: IME-USP.
- PAULINO, C.D.M. (1988). *Análise de dados categorizados incompletos: fundamentos, métodos e aplicações*. Tese de doutorado. IME-USP.
- PAULINO, C.D.M. (1991). Analysis of incomplete categorical data: a survey of the conditional maximum likelihood and weighted least squares approaches. *Brazilian Journal of Probability and Statistics* **5**, 1-42.
- PAULINO, C.D.M. e PEREIRA, C.A.B. (1992). Bayesian analysis of categorical data informatively censored. *Communications in Statistics - Theory and Methods* **21**, 2689-2705.
- PAULINO, C.D.M. e PEREIRA, C.A.B. (1994). On identifiability of parametric statistical models. *Journal of the Italian Statistical Society* **3**, 125-151.
- PAULINO, C.D.M. e PEREIRA, C.A.B. (1995). Bayesian methods for categorical data under informative general censoring. *Biometrika* **82**, 439-446.

- PAULINO, C.D.M. e SINGER, J.M. (2006). *Análise de dados categorizados*. São Paulo: Edgard Blücher.
- PAULINO, C.D.M. e SOARES, P.J.J. (2003). Analysis of rates in incomplete Poisson data. *Journal of the Royal Statistical Society. Series D: The Statistician* **52**, 87-99.
- PEPE, M.S. (2003). *The statistical evaluation of medical tests for classification and prediction*. Oxford: Oxford University Press.
- PEREIRA, C.A.B. e WECHSLER, S. (1994). Sobre a intenção das pesquisas de intenção de voto. *O Estado de São Paulo*, 18 de março, 2-2.
- PEREIRA, G.H.A. (2004). *Modelos de risco de crédito de clientes: uma aplicação a dados reais*. Dissertação de mestrado. IME-USP.
- PHILIPS, M.J. (1993). Contingency tables with missing data. *Journal of the Royal Statistical Society. Series D: The Statistician* **42**, 9-18.
- PRADO, R.R. (2004). *Modelagem do coeficiente kappa ponderado*. Dissertação de mestrado. IME-USP.
- PREISSER, J.S., GALECKI, A.T., LOHMAN, K.K. e WAGENKNECHT, L.E. (2000). Analysis of smoking trends with longitudinal binary responses. *Journal of the American Statistical Association* **95**, 1021-1031.
- PREISSER, J.S., LOHMAN, K.K. e RATHOUZ, P.J. (2002). Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random. *Statistics in Medicine* **21**, 3035-3054.
- QIN, J., LEUNG, D. e SHAO, J. (2002). Estimation with survey data under nonignorable nonresponse or informative sampling. *Journal of the American Statistical Association* **97**, 193-200.
- RAAB, G.M. e DONNELLY, C.A. (1999). Information on sexual behaviour when some data are missing. *Journal of the Royal Statistical Society. Series C: Applied Statistics* **48**, 117-133.
- RAO, C.R. (1973). *Linear statistical inference and its applications*. 2ª ed. New York: John Wiley & Sons.
- R DEVELOPMENT CORE TEAM (2006). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- REBOUSSIN, B.A., MILLER, M.E., LOHMAN, K.K. e TEN HAVE, T.R. (2002). Latent class models for longitudinal studies of the elderly with data missing at random. *Journal of the Royal Statistical Society. Series C: Applied Statistics* **51**, 69-90.
- ROBINS, J.M. (1997). Non-response models for the analysis of non-monotone non-ignorable missing data. *Statistics in Medicine* **16**, 21-37.
- ROBINS, J.M. e GILL, R.D. (1997). Non-response models for the analysis of non-monotone ignorable missing data. *Statistics in Medicine* **16**, 39-56.
- ROBINS, J.M. e ROTNITZKY, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association* **90**, 122-129.

- ROBINS, J.M., ROTNITZKY, A. e ZHAO, L.P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846-866.
- ROBINS, J.M., ROTNITZKY, A. e ZHAO, L.P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90**, 106-121.
- RODRIGUES, I.M.A. (1996). *Implementação computacional de análises clássicas de dados categorizados incompletos*. Dissertação de mestrado. Instituto Superior Técnico, Universidade Técnica de Lisboa.
- ROSA, P.T.M. (2000). *Modelos de "credit scoring": regressão logística, CHAID e REAL*. Dissertação de mestrado. IME-USP.
- ROSENBAUM, P.R. e RUBIN, D.B. (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **45**, 212-218.
- ROTHENBERG, T.J. (1971). Identification in parametric models. *Econometrica* **39**, 577-591.
- ROTNITZKY, A., COX, D.R., BOTTAI, M. e ROBINS, J.M. (2000). Likelihood-based inference with singular information matrix. *Bernoulli* **6**, 243-284.
- ROTNITZKY, A. e ROBINS, J.M. (1997). Analysis of semi-parametric regression models with non-ignorable non-response. *Statistics in Medicine* **16**, 81-102.
- ROTNITZKY, A. e WYPIJ, D. (1994). A note on the bias of estimators with missing data. *Biometrics* **50**, 1163-1170.
- RUBIN, D.B. (1974). Characterizing the estimation of parameters in incomplete-data problems. *Journal of the American Statistical Association* **69**, 467-474.
- RUBIN, D.B. (1976). Inference and missing data. *Biometrika* **63**, 581-592.
- RUBIN, D.B. (1978). Multiple imputation in sample surveys: a phenomenological bayesian approach to nonresponse. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 20-34.
- RUBIN, D.B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons.
- RUBIN, D.B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association* **91**, 473-489.
- RUBIN, D.B., STERN, H. e VEHOVAR, V. (1995). Handling "don't know" survey responses: the case of the Slovenian plebiscite. *Journal of the American Statistical Association* **90**, 822-828.
- SANCHES, M.R. (2005). *Indicadores formativos em modelos de equações estruturais*. Dissertação de mestrado. IME-USP.
- SÄRNDAL, C.-E., SWENSSON, B. e WRETMAN, J.H. (1991). *Model assisted survey sampling*. New York: Springer-Verlag.
- SATTEN, G.A. e CARROLL, R.J. (2000). Conditional and unconditional categorical regression models with missing covariates. *Biometrics* **56**, 384-388.

- SAVALLI, C., PAULINO, C.D.M., SILVA, G.L., SINGER, J.M., CHICARINO, M.P.Z., CASTRO, M. e TAVARES, R.A. (1999). *Análise de dados categorizados no sistema EMBRAPA-CNPTIA-NTIA, versão 1.0*. São Paulo: IME-USP.
- SCHAFFER, J.L. (1997). *Analysis of incomplete multivariate data*. Boca Raton: Chapman & Hall / CRC.
- SCHAFFER, J.L. (1999). Multiple imputation: a primer. *Statistical Methods in Medical Research* **8**, 3-15.
- SCHAFFER, J.L. e GRAHAM, J.W. (2002). Missing data: our view of the state of the art. *Psychological Methods* **7**, 147-177.
- SCHARFSTEIN, D.O. e IRIZARRY, R.A. (2003). Generalized additive selection models for the analysis of studies with potentially nonignorable missing outcome data. *Biometrics* **59**, 601-613.
- SCHLUCHTER, M.D. e JACKSON, K.L. (1989). Log-linear analysis of censored survival data with partially observed covariates. *Journal of the American Statistical Association* **84**, 42-52.
- SHEINER, L.B., BEAL, S.L. e DUNNE, A. (1997). Analysis of nonrandomly censored ordered categorical longitudinal data from analgesic trials (with discussion). *Journal of the American Statistical Association* **92**, 1235-1255.
- SHIH, W.J. (1987). Maximum likelihood estimation and likelihood ratio test for square tables with missing data. *Statistics in Medicine* **6**, 91-97.
- SHIH, W.J. (1992). On informative and random dropouts in longitudinal studies (letter to the editor). *Biometrics* **48**, 970-972.
- SMITH, P.W.F., SKINNER, C.J. e CLARKE, P.S. (1999). Allowing for non-ignorable non-response in the analysis of voting intention data. *Journal of the Royal Statistical Society. Series C: Applied Statistics* **48**, 563-577.
- SOARES, J.F. e SIQUEIRA, A.L. (2002). *Introdução à estatística médica*. 2ª ed. Belo Horizonte: COOPMED.
- SOARES, P.J.J. (2004). *Análise bayesiana de dados deficientemente categorizados*. Tese de doutorado. Instituto Superior Técnico, Universidade Técnica de Lisboa.
- SOARES, P.J.J. e PAULINO, C.D.M. (2001). Incomplete categorical data analysis: a bayesian perspective. *Journal of Statistical Computation and Simulation* **69**, 157-170.
- SPITZER, R.L., COHEN, J., FLEISS, J.L. e ENDICOTT, J. (1967). Quantification of agreement in psychiatric diagnosis. *Archives of General Psychiatry* **17**, 83-87.
- SPROTT, D.A. (2000). *Statistical inference in science*. New York: Springer-Verlag.
- STASNY, E.A. (1986). Estimating gross flows using panel data with nonresponse: an example from the canadian labour force survey. *Journal of the American Statistical Association* **81**, 42-47.
- STEPHAN, F.F. e MCCARTHY, P.J. (1958). *Sampling opinions*. New York: John Wiley & Sons.
- STRAM, D.O., WEI, L.J. e WARE, J.H. (1988). Analysis of repeated ordered categorical outcomes with possibly missing observations and time-dependent covariates. *Journal of the American Statistical Association* **83**, 631-637.

- SUNDBERG, R. (1974). Maximum likelihood theory for solution of the incomplete data from an eponential family. *Scandinavian Journal of Statistics* **1**, 49-58.
- SUNDBERG, R. (1976). An interative method for solution of the likelihood equations for incomplete data from exponential families. *Communications in Statistics - Simulation and Computation* **5**, 55-64.
- TANG, M.-L. e TANG, N.S. (2004). Exact tests for comparing two paired proportions with incomplete data. *Biometrical Journal* **46**, 72-82.
- TANNER, M.A. (1996). *Tools for statistical inference: methods for the exploration of posterior distributions and likelihood functions*. 3^a ed. New York: Springer-Verlag.
- TEN HAVE, T.R., JOFFE, M. e CARY, M. (2003). Causal logistic models for non-compliance under randomized treatment with univariate binary response. *Statistics in Medicine* **22**, 1255-1283.
- TEN HAVE, T.R., KUNSELMAN, A.R., PULKSTENIS, E.P. e LANDIS, J.R. (1998). Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics* **54**, 367-383.
- TEN HAVE, T.R., KUNSELMAN, A.R. e TRAN, L. (1999). A comparison of mixed effects logistic regression models for binary response data with two nested levels of clustering. *Statistics in Medicine* **18**, 947-960.
- TEN HAVE, T.R., MILLER, M.E., REBOUSSIN, B. e JAMES, M.K. (2000). Mixed effects logistic regression models for longitudinal ordinal functional response data with multiple-cause drop-out from the longitudinal study of aging. *Biometrics* **56**, 279-287.
- THOMAS, L.C., EDELMAN, D.B. e CROOK, J.N. (2002). *Credit scoring and its applications*. Philadelphia: Society for Industrial and Applied Mathematics.
- THOMAS, L.C., EDELMAN, D.B. e CROOK, J.N. (2004). *Readings in credit scoring: foundations, developments and aims*. Oxford: Oxford University Press.
- TOLEDANO, A.Y. e GATSONIS, C. (1999). Generalized estimating equations for ordinal categorical data: arbitrary patterns of missing responses and missingness in a key covariate. *Biometrics* **55**, 488-496.
- TROXEL, A.B., LIPSITZ, S.R. e BRENNAN, T.A. (1997). Weighted estimating equations with nonignorably missing response data. *Biometrics* **53**, 857-869.
- VACH, W. (1994). *Logistic regression with missing values in the covariates*. New York: Springer-Verlag.
- VACH, W. (1997). Some issues in estimating the effect of prognostic factors from incomplete covariate data. *Statistics in Medicine* **16**, 57-72.
- VACH, W. e BLETTNER, M. (1995). Logistic regression with incompletely observed categorical covariates - investigating the sensitivity against violation of the missing at random assumption. *Statistics in Medicine* **14**, 1315-1329.
- VACH, W. e SCHUMACHER, M. (1993). Logistic regression with incompletely observed categorical covariates: a comparison of three approaches. *Biometrika* **80**, 353-362.

- VANSTEELANDT, S. e GOETGHEBEUR, E. (2001). Analyzing the sensitivity of generalized linear models to incomplete outcomes via the IDE algorithm. *Journal of Computational and Graphical Statistics* **10**, 656-672.
- VANSTEELANDT, S. e GOETGHEBEUR, E. (2005). Sense and sensitivity when correcting for observed exposures in randomized clinical trials. *Statistics in Medicine* **24**, 191-210.
- VANSTEELANDT, S., GOETGHEBEUR, E., KENWARD, M.G. e MOLENBERGHS, G. (2003). *Ignorance and uncertainty regions as inferential tools in a sensitivity analysis*. Relatório técnico. Centrum voor Statistiek, Ghent University.
- VENEZUELA, M.K. (2003). *Modelos lineares generalizados para análise de dados com medidas repetidas*. Dissertação de mestrado. IME-USP.
- VERBEKE, G. e MOLENBERGHS, G. (2000). *Linear mixed models for longitudinal data*. New York: Springer-Verlag.
- VERMUNT, J.K. (1996). *Log-linear event history analysis: a general approach with missing data, latent variables and unobserved heterogeneity*. Tese de doutorado. Tilburg University.
- VERZILLI, C.J. e CARPENTER, J.R. (2002). Assessing uncertainty about parameter estimates with incomplete repeated ordinal data. *Statistical Modelling* **2**, 203-215.
- WALKER, S. (1996). A bayesian maximum a posteriori algorithm for categorical data under informative general censoring. *Journal of the Royal Statistical Society. Series D: The Statistician* **45**, 293-298.
- WANG, Y.-G. (1999). Estimating equations with nonignorably missing response data. *Biometrics* **55**, 984-989.
- WEST, C.P. e DAWSON, J.D. (2002). Complete imputation of missing repeated categorical data: one-sample applications. *Statistics in Medicine* **21**, 203-217.
- WILLIAMSON, G.D. e HABER, M. (1994). Models for three-dimensional contingency tables with completely and partially cross-classified data. *Biometrics* **49**, 194-203.
- WOODBURY, M.A. (1971). Discussion of the paper by Hartley and Hocking. *Biometrics* **27**, 808-813.
- WOOLSON, R.F. e CLARKE, W.R. (1984). Analysis of categorical incomplete longitudinal data. *Journal of the Royal Statistical Society. Series A: Statistics in Society* **147**, 87-99.
- ZHAO, L.P., LIPSITZ, S.R. e LEW, D. (1996). Regression analysis with missing covariate data using estimating equations. *Biometrics* **52**, 1165-1182.
- ZHOU, X.-H. (1998). Comparing accuracies of two screening tests in a two-phase study for dementia. *Journal of the Royal Statistical Society. Series C: Applied Statistics* **47**, 135-147.
- ZHOU, X.-H. e CASTELLUCCIO, P. (2004). Adjusting for non-ignorable verification bias in clinical studies for Alzheimer's disease. *Statistics in Medicine* **23**, 221-230.
- ZHOU, X.-H., OBUCHOWSKI, N.A. e MCCLISH, D.K. (2002). *Statistical methods in diagnostic medicine*. New York: John Wiley & Sons.

Índice de autores

- Agresti, A., 1, 108
Albert, P.S., 26, 52
Alfred, E., 26
Allen, A.S., 41
Allison, P.D., 49
Allison, T., 77
Almeida, A.C., 118
Alonzo, T.A., 135
Amara, I.A., 87
Andersen, E.B., 51
Anderson, T.W., 28
Andrade, D.F., 51
André, C.D.S., 58
Artes, R., 49
Atkinson, S.S., 71, 79
Azevedo, C.L.N., 51

Baker, S.G., 1, 21, 25, 26, 41, 48, 50, 52, 53, 97, 135
Barnhart, H.X., 135
Basu, D., 16, 51
Beal, S.L., 50
Begg, C.B., 135
Bellinger, D., 26
Birmingham, J., 7, 26
Bishop, Y.M.M., 1, 100
Blackhurst, D.W., 51
Blettner, M., 51

Blumenthal, S., 2
Bolfarine, H., 118
Bollen, K.A., 52
Bonetti, M., 30
Bottai, M., 31
Botter, D.A., 49
Brennan, T.A., 50
Brown, C.H., 27
Brown, M.B., 26
Bussab, W.O., 118

Carpenter, J.R., 43
Carroll, R.J., 51
Cary, M., 50
Casella, G., 28
Castelluccio, P., 135
Castro, M., 94
Chambers, R.L., 26, 121
Chen, C.C.H., 3
Chen, H.Y., 50
Chen, M.-H., 51
Chen, T.T., 2, 3, 21, 26
Chicarino, M.P.Z., 94
Choi, S.C., 3, 21
Cicchetti, D.V., 77
Clarke, P.S., 30, 31, 97
Clarke, W.R., 49
Clifford, P., 50

- Cochran, W.G., 118
Cohen, J., 74, 76, 77
Cole, B.F., 30
Colosimo, E.A., 50
Conaway, M.R., 26
Cook, R.J., 50
Copas, J., 42
Cordeiro, G.M., 51
Cox, D.R., 31
Crook, J.N., 52
- Davison, A.C., 30
Dawid, A.P., 1
Dawson, J.D., 51
Deltour, I., 26
Dempster, A.P., 3, 26, 65, 141
DerSimonian, R., 1, 26, 48
Dewulf, L., 50
Dickey, J.M., 1, 51
Diggle, P.J., 6, 49
Dillman, D.A., 49
Donnelly, C.A., 26
Dunne, A., 50
- Edelman, D.B., 52
Efron, B., 30
Eguchi, S., 42
Ekholm, A., 7
Elashoff, J.D., 21
Elashoff, R.M., 21
Eltinge, J.L., 49
EMBRAPA, 94
Endicott, J., 76
Essers, J.G.A., 26
Ewell, M., 51
Fay, R.E., 19
- Feelders, A.J., 52
Fienberg, S.E., 1–3, 21, 100
Findlay, J.G., 40
FitzGerald, P.E.B., 50
Fitzmaurice, G.M., 3, 7, 26, 30, 49–52, 56
Fleiss, J.L., 1, 49, 75–77, 126
Follmann, D.A., 52
Forster, J.J., 51
Forthofer, R.N., 1, 71
- Gail, M.H., 49
Galecki, A.T., 50
Garcia, M.G., 26
Gatsonis, C., 50
Gelber, R.D., 30
Gibbons, L.E., 51
Gill, R.D., 26
Giolo, S.R., 50
Glas, C.A.W., 51
Glonck, G.F.V., 30
Glynn, R.J., 7, 41
Goetghebeur, E., 26, 30, 42, 43, 53
Graham, J.W., 49, 51
Graubard, B.I., 41
Green, P.E., 51
Greenes, R.A., 135
Grizzle, J.E., 49, 69
Groves, R.M., 49
Gunel, E., 51
- Haber, M., 3, 84
Hagenaars, J.A., 51
Hambleton, R.K., 51
Hancock, G.R., 52
Harrington, D.P., 49
Hartley, H.O., 2, 3

- He, C.Z., 51
Heagerty, P., 49
Heath, A.F., 50
Heitjan, D.F., 11
Hens, N., 31
Hesran, J.-Y.Le, 26
Heyting, A., 26
Hinkley, D.V., 30
Ho, L.L., 87
Hocking, R.R., 2, 3
Holland, P.W., 1, 100
Holman, R., 51
Honaker, J., 51
Horton, N.J., 50–52
Hosmer, D.W., 51
Hoyle, R.H., 52

Ibrahim, J.G., 50, 51
Iliopoulos, G.M., 100
Imrey, P.B., 49, 68, 70, 71, 79, 87
Irizarry, R.A., 52

Jackson, K.L., 50
James, M.K., 50
Jiang, J.-M., 51
Joffe, M., 50
Joseph, A., 51

Kadane, J.B., 51
Kateri, M.M., 100
Katz, J., 26
Kenward, M.G., 6, 7, 10, 30, 42, 43, 49, 50
King, G., 51
Kish, L., 118, 119
Klein, J.P., 50
Kline, R.B., 52
Ko, C.-W., 41

Koch, G.G., 49, 68–71, 79, 87
Kosinski, A.S., 135
Krishnan, T., 11
Kunselman, A.R., 50

Laird, N.M., 3, 7, 21, 26, 30, 41, 49–51, 65, 97,
141
Landis, J.R., 50
Latif, S.A., 52
Lehmann, E.L., 28
Lehnen, R.G., 1, 71
Lehtonen, R., 121
Lesaffre, E., 49, 50
Leung, D., 52
Levin, B., 1, 49, 75, 126
Levinton, A., 26
Lew, D., 50
Liang, K.-Y., 49, 52
Lima Neto, E.A., 51
Lin, H., 51
Lin, I.-F., 52
Lindern, W.J., 51
Lindsey, J.K., 16, 49
Lipsitz, S.R., 3, 7, 30, 49–51, 56
Little, R.J.A., 7, 11, 19, 26, 49–51, 62
Liu, C.H., 25
Liu, X., 26
Lohman, K.K., 50, 51
Louis, T.A., 25
Lyles, R.H., 41

Madow, W.G., 49
Mark, S.D., 49
Mays, E., 52
McCarthy, P.J., 119, 120
McClish, D.K., 135

- McCullagh, P., 51
 McCulloch, C.E., 51
 McCutcheon, A.L., 51
 McLachlan, G.J., 11
 Meilijson, I., 25
 Melhado, T.T., 52
 Meng, X.-L., 25, 139
 Michiels, B., 6, 7, 26, 96
 Miller, M.E., 50, 51
 Moeschberger, M.L., 50
 Molenberghs, G., 6, 7, 10, 26, 27, 30, 42, 43, 49–51, 53, 96
 Mueller, R.O., 52
 Murray, G.D., 40

 Nelder, J.A., 51
 Neves, M.M.C., 58
 Nisselson, H., 49
 Nordheim, E.V., 42

 Obuchowski, N.A., 135
 Oleson, J.J., 51
 Olkin, I., 49
 Orchard, T., 3
 Oxspring, H.H., 2

 Pahkinen, E., 121
 Paik, M.C., 1, 49–52, 75, 126
 Park, T., 26, 51
 Parzen, M., 51
 Paula, G.A., 51
 Paulino, C.D.M., ix, xi, 1, 30, 49, 51, 53, 58, 71, 73, 78, 79, 87, 94, 108, 137
 Pepe, M.S., 135
 Pereira, C.A.B., 30, 51, 58, 118
 Pereira, G.H.A., 52
 Peterson, H., 51

 Petkova, E., 26
 Philips, M.J., 42
 Prado, R.R., 108
 Preisser, J.S., 50
 Pulkstenis, E.P., 50

 Qin, J., 52

 R Development Core Team, 4, 91, 137
 Raab, G.M., 26
 Rao, C.R., 143
 Rathouz, P.J., 50
 Reboussin, B., 50
 Reboussin, B.A., 50, 51
 Reinfurt, D.W., 49, 68, 70, 87
 Rejeski, W.J., 50
 Richardson, S., 26
 Robins, J.M., 26, 31, 49, 50
 Rodrigues, I.M.A., 58, 94
 Rosa, P.T.M., 52
 Rosenbaum, P.R., 42
 Rosenberger, W.F., 1, 26, 48
 Rosenheck, R.A., 51
 Rothenberg, T.J., 28
 Rotnitzky, A., 26, 31, 49, 50
 Rubin, D.B., 3, 7, 10, 11, 16, 25, 26, 41, 42, 49, 51, 65, 141

 Sacco, R., 52
 Sanches, M.R., 52
 Särndal, C.-E., 118
 Satten, G.A., 51
 Savalli, C., 94
 Schafer, J.L., 49, 51
 Scharfstein, D.O., 52
 Scheve, K., 51
 Schluchter, M.D., 50, 51

- Schumacher, M., 51
Shao, J., 52
Shapiro, S., 26
Sheiner, L.B., 50
Shih, W.J., 3, 26
Silva, G.L., 94
Singer, J.M., 1, 49, 53, 58, 71, 73, 78, 79, 87,
94, 108, 137
Siqueira, A.L., 126
Skinner, C.J., 7, 30, 97, 121
Smith, P.W.F., 30, 31, 97
Smith, W.F., 51
Soares, J.F., 126
Soares, P.J.J., 51, 53, 58
Spitzer, R.L., 76
Sprott, D.A., 108
Stablein, D.M., 3, 21
Starmer, C.F., 49, 69
Stasny, E.A., 21
Stephan, F.F., 119, 120
Stern, H., 26
Stokes, M.E., 71, 79, 87
Stram, D.O., 26
Suh, E.B., 52
Sundberg, R., 3
Swensson, B., 118

Tang, M.-L., 3
Tang, N.S., 3
Tanner, M.A., 11
Tavares, H.R., 51
Tavares, R.A., 94
Ten Have, T.R., 50, 51
Thomas, L.C., 52
Tibshirani, R.J., 30
Tolboom, J.T.B.M., 26

Toledano, A.Y., 50
Troxel, A.B., 50
Tseng, T.H., 58

Vach, W., 51
Valle, R.C., 51
Vansteelandt, S., 42, 43
Vehovar, V., 26
Venezuela, M.K., 49
Verbeke, G., 49
Vermunt, J.K., 51
Verzilli, C.J., 43

Wagenknecht, L.E., 50
Walker, S., 51
Wang, S.A., 52
Wang, Y.-G., 50
Ware, J.H., 26
Waternaux, C., 26
Wechsler, S., 118
Wedderburn, R.W.M., 51
Wei, L.J., 26
Welsh, A.H., 26
West, C.P., 51
Williamson, G.D., 3, 84
Woodbury, M.A., 3
Woolson, R.F., 49
Wretman, J.H., 118
Wu, Y.N., 25
Wypij, D., 49

Yi, G.Y., 50

Zahner, G.E.P., 30, 50
Zeger, S.L., 49
Zeng, L., 50
Zhao, L.P., 49–51
Zhou, X.-H., 135

Índice de assuntos

- abandono, 1, 11, 27, 62
- ACC (análise de casos completos), **2**, 3, 21, 137, 138
 - exemplos/análises, em, 45, 100, 101, 106, 111, 113, 118, 127–133
 - MAR, sob, 9, 12–14, 41
 - MCAR, sob, 16–19, 41
 - MNAR, sob, 32–40, 42
- ambiente computacional estatístico
 - Minitab, 1
 - NTIA, 94
 - R, 1, 4, 91–94, 137
 - S-Plus, 1, 94
 - SAS, 1, 94
 - SPSS, 1
 - Statistica, 1
- amostragem
 - aleatória simples, 54, 117, 119, 120
 - conglomerado, por, 120
 - dois estágios, em, 118, 120
 - estratificada, 54, 117
 - quotas, por, 117–120
- análise de sensibilidade, 41–43, 48, 108, 115, 122, 138
- associação, ver razão de chances
 - linear por linear, 73
- Bayes empírico, 26
- bayesiana, inferência, 51, 138
- bootstrap*, 30, 31
- classes latentes, análise de, 51
- coarse data*, ver dados grosseiros
- complete case analysis*, ver ACC
- concordância, medida de, ver *kappa*
- conglomerado, ver amostragem
- credit scoring*, ver risco de crédito, modelos de
- dados grosseiros, 11
- delta*, método, 67, 80, 81, 83, 89
- diagnóstico, precisão de teste de, 126–135
- distribuição
 - binomial, 16, 119
 - degenerada, 28–30
 - Dirichlet, 26
 - multinomial, 5, **7**, 13, 16, 17, 27, 31, 53, 62, 69, 120–122, 128, 130
 - Poisson, de, 53
 - produto de multinomiais, 16, **60**, 62, 69, 130
 - score*, do vetor, 28–30
- dropout*, ver abandono
- EM, algoritmo, 3, 11, 22–26, 65, 67
 - MAR/MCAR, sob, 64, 141
 - MNAR, sob, 24
- equações de estimação generalizadas, 49–50, 108

- equações estruturais, modelagem de, 52
- erro de classificação, 51
- especificidade, 126–134, 137
- estatística
- ancilar, 16
 - ancilar específica, 16
 - ancilar parcial (S-ancilar), 16
 - ignorância, **42**, 43, 48, 95, 101, 108, 138
 - imprecisão, **42**, 43, 135
 - incerteza, **42**, 95, 101, 108
 - suficiente, 16
 - suficiente específica, 16
 - suficiente parcial (S-suficiente), 16
- estimador protetor, **27**, 96, 109, 121
- estrato, 54, 94, 120
- Fisher, algoritmo *scoring* de, 11, 25, 65–67, 80, 83
- formulação
- equações livres, em, 71–73, 78, 81, 88, 93
 - restrições, em termos de, 71, 73, 74, 78, 82, 88
- freqüência nula, ver zero amostral
- GSK, método, 49, 69
- identificabilidade, ver modelo identificável
- ignorabilidade, ver omissão ignorável
- dados grosseiros, em, 11
- imputação múltipla, 51, 94
- inferência sobre rejeitados, 52
- inidentificabilidade, ver modelo inidentificável
- intervalo
- ignorância/incerteza, de, **43**, 47, 106, 113, 115, 122, 123, 125, 138
 - melhor-pior caso, para o, **43**, 44, 95, 100, 101, 108, 115, 121, 123–125, 138
- kappa*, medida
- Cohen, de, 74–76, 90, 108–115
 - ponderada, 76–77, 90, 108–115
- last observation carried forward* (LOCF), 50
- logito, 20, 25, 44, 47
- referência, de, 81
- longitudinais, estudos/dados, 11, 19, 26, 27, 50
- MAR (*missing at random*), **7**, 8, 15, 16, 20, 40–42, **61**, 62, 137
- ACC, 9, 12–14, 41
 - EM, algoritmo, 64, 141
 - estimação, 10, 11, 49, 63–68, 71, 78–94
 - exemplos/análises, em, 44–47, 95–113, 121, 130–135
 - simulação, estudo de, 13, 14
 - verossimilhança, função de, 9, 10, 61
- MCAR (*missing completely at random*), **15**, 20, 41, 42, **62**, 137
- ACC, 16–19, 41
 - EM, algoritmo, 64, 141
 - estimação, 15, 16, 49, 63–71, 78–94
 - exemplos/análises, em, 44–46, 95–106, 109, 111, 122, 130–135
 - simulação, estudo de, 17–19
 - verossimilhança, função de, 15, 62
- mecanismo de omissão, ver omissão
- misclassification*, ver erro de classificação
- missing(ness)*, ver omissão
- MNAR (*missing not at random*), **19–21**, **63**, 138–139
- ACC, 32–40, 42
 - ajuste imperfeito de modelos saturados, 21–25, 28, 31–41, 43, 97, 103, 123, 138
 - EM, algoritmo, 24

- estimação, 21–27, 71, 87–90, 92–94
- exemplos/análises, em, 44–48, 95–115, 121–125, 135
- modelo saturado inidentificável, 27–40, 42
- simulação, estudo de, 31–40
- verossimilhança, função de (log-), 29, 87, 88, 122, 125
- modelo
 - funcional linear, 53, 71, **87–90**, 92–94, 137
 - kappa* de Cohen, para a medida, 74
 - kappa* ponderada, para a medida, 76
 - sensibilidade/especificidade, para a, 129
 - valor preditivo positivo/negativo, para o, 129
 - identificável, 7, 27, 28, 30, 31, 37, 42, 51, 61, 63, 139
 - inidentificável, 5, **27**, 28, 30, 32, 37, 40, 42, 61, 138, 139
 - linear, 53, 71, **78–80**, 80, 90, 92–94, 137
 - homogeneidade marginal, de, 73
 - linear generalizado, 51
 - log-linear, 3, 50, 53, 71, **81–83**, 90, 92–94, 121, 137
 - independência, de, 71
 - razão de chances adjacentes comum, de, 72, 73
 - logístico, 50, 51, 121
 - mistura de padrões, de, 6, 7, 139
 - seleção, de, 6, 7
 - sobre-parametrizado/-saturado, 5, 20, 27, 43, 47, 48, 61, 106, 113, 122
 - sobrevivência, de análise de, 50
- multiple imputation*, ver imputação múltipla
- não-cumprimento, 1
- não-ignorabilidade, ver omissão não-ignorável
- não-resposta, 1, 19
 - pesquisas, em, 49, 117–126
 - viés de, 117–126
- Newton-Raphson, algoritmo de, 11, 25, 65, 67
- non-compliance*, ver não-cumprimento e recusa
- non-response*, ver não-resposta
- nuisance*, ver parâmetro perturbador
- observed at random* (OAR), 16
- omissão, **1**, 139
 - aleatória, ver MAR
 - completamente aleatória, ver MCAR
 - ignorável, **3**, **10**, 16, 50, 51, 69, 135
 - informativa, ver MNAR
 - não-aleatória, ver MNAR
 - não-ignorável, **3**, **10**, 19, 26, 31, 50, 51, 87, 135
 - não-informativa, ver MAR
 - padrão monótono, 11
 - variáveis explicativas, em, 50, 51
- parâmetro
 - distinto, 10
 - estimável, **43**, 47, 106, 113, 122
 - perturbador ou de incômodo, 16, 79
 - sensibilidade, de, **43**, 47, 106, 113, 122, 123, 125, 138
- pattern-mixture model*, ver modelo de mistura de padrões
- precisão de teste de diagnóstico, 126–135
- princípio generalizado da condicionalidade, 16
- probabilidade
 - condicional de omissão, 6
 - marginal de categorização, 6
- probitto, 25
- processo de omissão, ver omissão

- programa, ver ambiente computacional
- protective estimator*, ver estimador protetor
- quotas, ver amostragem
- razão de chances, (logaritmo da), 13, 44–48, 71
adjacentes, 73, 100, 137
- recusa, 1
- região de ignorância/incerteza, **42**, **43**, 113, 114
- regressão funcional assintótica, 87
- reject inference*, ver inferência sobre rejeitados
- risco de crédito, modelos de, 52
- selection model*, ver modelo de seleção
- sensibilidade, 126–134, 137
análise de, ver análise de sensibilidade
parâmetro de, ver parâmetro de sensibilidade
- simulação, estudo de
MAR, 13, 14
MCAR, 17–19
MNAR, 31–40
- software*, ver ambiente computacional
- teoria de resposta ao item, 51
- teste de diagnóstico, precisão de, 126–135
- valores preditivos positivo e negativo, 126–134, 137
- verification bias*, ver viés de verificação
- verossimilhança, função de (log-), 7
MAR, sob, 9, 10, 61
MCAR, sob, 15, 62
MNAR, sob, 29, 87, 88, 122, 125
- viés
não-resposta, de, ver não-resposta
verificação, de, 135
- zero amostral/frequência nula, 26, 65, 68, 70, 109, 130–132, 135