Análise de dados categorizados com omissão em variáveis explicativas e respostas

Frederico Zanqueta Poleto

Tese apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo
para
obtenção do título
de
Doutor em Ciências

Programa: Estatística

Orientador: Prof. Dr. Julio da Motta Singer

Coorientador: Prof. Dr. Carlos Daniel Paulino

Durante o desenvolvimento deste trabalho o autor
recebeu auxílio financeiro da FAPESP e da CAPES

São Paulo, abril de 2011

Análise de dados categorizados com omissão em variáveis explicativas e respostas

Este exemplar corresponde à redação
final da tese devidamente corrigida
e defendida por Frederico Zanqueta Poleto
e aprovada pela Banca Examinadora.

São Paulo, 08 de abril de 2011.

Banca Examinadora:

- Prof. Dr. Julio da Motta Singer – IME-USP.
- Prof. Dr. Carlos Daniel Paulino – IST-Universidade Técnica de Lisboa.
- Prof. Dr. Fernando Antônio da Silva Moura – IM-UFRJ.
- Profa. Dra. Rosangela Helena Loschi – ICEx-UFMG.
- Prof. Dr. Enrico Antônio Colosimo – ICEx-UFMG.

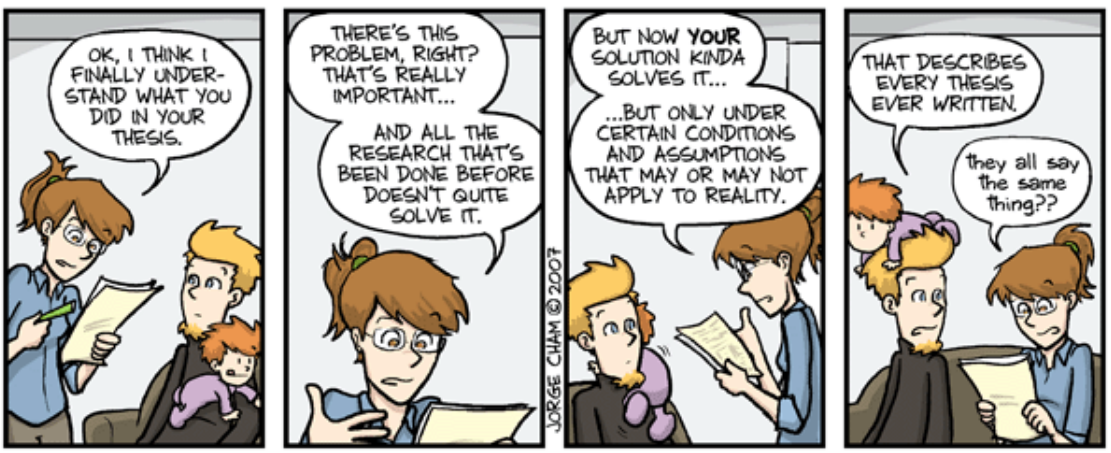
www.phdcomics.com

JORGE CHAM ©THE STANFORD DAILY phd.stanford.edu/comics

JORGE CHAM ©THE STANFORD DAILY phd.stanford.edu/comics

JORGE CHAM ©THE STANFORD DAILY phd.stanford.edu/comics

JORGE CHAM ©THE STANFORD DAILY phd.stanford.edu/comics



WWW.PHDCOMICS.COM

À minha esposa Lílian,

aos meus pais Dagmar e Sérgio,

à minha avó Aurora

e aos meus irmãos Natália e Guilherme

Resumo

Nesta tese apresentam-se desenvolvimentos metodológicos para analisar dados com omissão e também estudos delineados para compreender os resultados de tais análises.

Escrutinam-se análises de sensibilidade bayesiana e clássica para dados com respostas categorizadas sujeitas a omissão. Mostra-se que as componentes subjetivas de cada abordagem podem influenciar os resultados de maneira não-trivial, independentemente do tamanho da amostra, e que, portanto, as conclusões devem ser cuidadosamente avaliadas. Especificamente, demonstra-se que distribuições *a priori* comumente consideradas como não-informativas ou levemente informativas podem, na verdade, ser bastante informativas para parâmetros inidentificáveis, e que a escolha do modelo sobreparametrizado também tem um papel importante.

Quando há omissão em variáveis explicativas, também é necessário propor um modelo marginal para as covariáveis mesmo se houver interesse apenas no modelo condicional. A especificação incorreta do modelo para as covariáveis ou do modelo para o mecanismo de omissão leva a inferências enviesadas para o modelo de interesse. Trabalhos anteriormente publicados têm-se dividido em duas vertentes: ou utilizam distribuições semiparamétricas/não-paramétricas, flexíveis para as covariáveis, e identificam o modelo com a suposição de um mecanismo de omissão não-informativa, ou empregam distribuições paramétricas para as covariáveis e permitem um mecanismo mais geral, de omissão informativa. Neste trabalho analisam-se respostas binárias, combinando um mecanismo de omissão informativa com um modelo não-paramétrico para as covariáveis contínuas, por meio de uma mistura induzida pela distribuição *a priori* de processo de Dirichlet.

No caso em que o interesse recai apenas em momentos da distribuição das respostas, propõe-se uma nova análise de sensibilidade sob o enfoque clássico para respostas incompletas que evita suposições distribucionais e utiliza parâmetros de sensibilidade de fácil interpretação. O procedimento tem, em particular, grande apelo na análise de dados contínuos, campo que tradicionalmente emprega suposições de normalidade e/ou utiliza parâmetros de sensibilidade de difícil interpretação.

Todas as análises são ilustradas com conjuntos de dados reais.

Palavras-chave: Análise de sensibilidade; Dados faltantes ou incompletos; Identificabilidade; Intervalos de ignorância e de incerteza; MAR, MCAR e MNAR; Mecanismo de omissão não-informativa e informativa; Modelos de seleção e de mistura de padrões; Processo de Dirichlet; Sobreparametrização.

Abstract

We present methodological developments to conduct analyses with missing data and also studies designed to understand the results of such analyses.

We examine Bayesian and classical sensitivity analyses for data with missing categorical responses and show that the subjective components of each approach can influence results in non-trivial ways, irrespectively of the sample size, concluding that they need to be carefully evaluated. Specifically, we show that prior distributions commonly regarded as slightly informative or non-informative may actually be too informative for non-identifiable parameters, and that the choice of over-parameterized models may drastically impact the results.

When there is missingness in explanatory variables, we also need to consider a marginal model for the covariates even if the interest lies only on the conditional model. An incorrect specification of either the model for the covariates or of the model for the missingness mechanism leads to biased inferences for the parameters of interest. Previously published works are commonly divided into two streams: either they use semi-/non-parametric flexible distributions for the covariates and identify the model via a non-informative missingness mechanism, or they employ parametric distributions for the covariates and allow a more general informative missingness mechanism. We consider the analysis of binary responses, combining an informative missingness model with a non-parametric model for the continuous covariates via a Dirichlet process mixture.

When the interest lies only in moments of the response distribution, we consider a new classical sensitivity analysis for incomplete responses that avoids distributional assumptions and employs easily interpreted sensitivity parameters. The procedure is particularly useful for analyses of missing continuous data, an area where normality is traditionally assumed and/or relies on hard-to-interpret sensitivity parameters.

We illustrate all analyses with real data sets.

Key words: Sensitivity analysis; Incomplete or missing data; Identifiability; Ignorance and uncertainty intervals; MAR, MCAR and MNAR; Informative and non-informative missingness mechanism; Selection and pattern-mixture models; Dirichlet process; Overparameterization.

Agradecimentos

É com imensa alegria que olho para trás e vejo o tamanho amparo e auxílio que tive no decorrer desta longínqua jornada do meu doutorado e até mesmo da trajetória mais abrangente de minha vida que possibilitou, de fato, chegar até aqui.

Ao Julio da Motta Singer, Carlos Daniel Paulino e Geert Molenberghs, que me orientaram, ensinaram, guiaram e inspiraram, que tiveram paciência e entenderam minhas limitações, que suportaram minhas escolhas, que foram compreensíveis e companheiros, só posso deixar o meu agradecimento de todo o coração! As possibilidades que me forneceram certamente mudaram minha vida.

Sou muito grato (1) à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), que me forneceu bolsa de doutorado (processo 06/06860-8) e, assim, me suportou durante o desenvolvimento da pesquisa no Brasil, além de ter possibilitado a realização de viagens de estágio e de participação em congressos nacionais e internacionais, (2) à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), que me disponibilizou bolsa do Programa de Doutorado no País com Estágio no Exterior (PDEE, vulgo doutorado sanduíche, processo BEX 1901/08-7), abrindo as portas para uma interação mais próxima junto ao meu orientador estrangeiro e (3) outros auxílios recebidos para idas a congressos e realizações de cursos, especificamente, do projeto CAPES PROEX e do *Section on Bayesian Statistical Science* (SBSS) da *American Statistical Association* (ASA).

Agradeço ao Dr. Arnaud Perrier e ao Dr. Henri Bounameaux, da *Division of General Internal Medicine* do *Geneva University Hospital*, por me fornecerem o conjunto de dados que utilizo no Capítulo 4, bem como a amigos e colegas que tentaram obter autorização de uso de conjuntos de dados de suas instituições financeiras e, principalmente, ao Mauro Correia Alves, ao Abraham Laredo Sicsú, à Maria José Pegorin e demais envolvidos no processo que efetivamente possibilitou receber autorização para utilização de uma amostra da Serasa Experian. Infelizmente, dificuldades do emprego da abordagem do Capítulo 4 em grandes tamanhos de amostra na fase final de desenvolvimento da tese impossibilitaram que o conjunto de dados fosse, de fato, empregado como exemplo.

Os *clusters* de computadores do Departamento de Matemática do Instituto Superior Técnico (DM-IST) e do Laboratório de Computação Científica Avançada da Universidade de São Paulo (LCCA-USP) foram essenciais para a condução dos gigantescos trabalhos computacionais das Seções 3.2, 4.4 e 5.4. Sou grato a todos que possibilitaram que eu utilizasse esses sistemas, mais especificamente, ao professor Francisco Miguel Dionísio, Edgar Costa, João Luis Dimas Santos, Francisco

Ribacionka e Ettore Enrico Delfino Ligorio.

As infraestruturas da USP e da universidade do exterior, com disponibilização de livros e acesso a artigos de periódicos relevantes, subsidiaram significativamente o desenvolvimento dos trabalhos, portanto, agradeço muito terem me permitido o acesso a essa gama de informações.

Aos membros suplentes, sou grato por se prontificarem em completar a composição da banca, e, aos titulares, Rosangela Helena Loschi, Enrico Antônio Colosimo e Fernando Antônio da Silva Moura, adicionalmente, pela atenciosa leitura da tese e pelas contribuições.

À minha esposa Lílian Nati, agradeço muito por sempre me encorajar e me apoiar, por ser incrivelmente companheira, iluminando meus dias com amor, humor, compreensão e alegria, e pela sua força e determinação.

Aos meus pais e à minha avó, sou grato por sempre terem me dado o exemplo de que nada na vida se consegue sem esforço e dedicação, por mais adversas que as situações sejam, e também por terem buscado o melhor para mim, muitas vezes com muito sacrifício.

Aos meus irmãos, agradeço por me inspirarem e humildemente me ensinarem que a vida proporciona muitos bons caminhos.

Aos meus sogros, sou grato pelo acolhimento caloroso, compreensivo, o suporte, o respeito e a torcida.

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

Seja na USP, no Instituto Superior Técnico da Universidade Técnica de Lisboa (IST-UTL), no Centro de Aplicações da Universidade de Lisboa (CEAUL-FCUL) ou em outros âmbitos, dentro ou fora do Brasil, amigos, colegas, professores, dentre outros profissionais de setores administrativos, me receberam calorosamente e auxiliaram minha pesquisa direta ou indiretamente: sou grato a todos. Peço desculpas pela lista não ser completa, mas faço questão de tentar mencionar algumas pessoas que foram importantes nestes últimos anos: Afrânio Márcio Corrêa Vieira, Airlane Alencar, Alejandro Jara, Alexandra Calabrez, Alexandra Mello Schimidt, Ana Poletto, André Gomes de Azevedo, André Luiz Silva Samartini, Angel Rodriguez, Ângela Tavares Paes, Antonio Carlos Pedroso de Lima, Antonio Carlos Zanchetta, Ariel Alonso Abad, Caio Lucidius Naberezny Azevedo, Carine Savalli Redígolo, Carlos Alberto Auricchio Jr, Carlos Eduardo Crespo, Carolina Pacífico, Cátia Petri Nakano, Cibele Maria Russo, Clarice Garcia Borges Démetrio, Cléber da Costa Figueiredo, Clélia Maria Castro de Tolói, Conny Aerts, Cristiano Alves, Daniela Estaregue, Danillo Nakano, Danilo Clemente Coelho, David da Silva Pires, Dione Maria Valença, Dóris Satie Maruyama Fontes, Edson Roberto da Silva, Egídio Moretti, Élen Natis Gomes de Azevedo, Elias Teixeira Krainski, Elisabeti Kira, Emilene Parlato, Erika Vandersmissen, Fabio Carlos Pereira da Silva, Fábio de Marqui, Fábio Prates Machado, Fábio Veríssimo Gonçalves, Felipe Villarino Prieto, Fernando Correa Lima, Fernando Frota Redígolo, Fernando Henrique Ferraz Pereira da Rosa, Fernando Valvano Cerezetti, Francisco José de Azevêdo Cysneiros, Francisco Marcelo Monteiro da Rocha, Gilberto A. Paula,

Giovani Loiola Silva, Gisela Tunes da Silva, Gustavo Castilhos, Gustavo Cruz, Gustavo Henrique de Araujo Pereira, Heleno Bolfarine, Hélio Arizono, Inês Dias, Isa Debonnez, Jacqueline Sant'Eufemia David, João Fernando Serrajordia Rocha de Mello, João Paulo dos Santos Pacífico, Joel Mauricio Correa da Rosa, Jorge Alberto Achcar, José Cortiñas Abrahantes, Josep Argelich Romà, Juan A. Acébron, Juvêncio Santos Nobre, Karina Pretto, Kárita Cristina Francisco, Kátia Fontana, Leonardo Soares Bastos, Liliam Pereira de Lima, Lina Thomas, Livia Perancini Fernandes, Lizandra Castilho Fabio, Lourdes Contreras Montenegro, Lúcia Pereira Barroso, Luiz Carlos Moretti, Marc Thoelen, Marcel de Toledo Vieira, Marcel Frederico de Lima Taga, Marcelo Castro, Marcelo Hiroshi Ogava, Marcelo Leal Santos, Márcia D'Elia Branco, Marco César dos Santos Barbosa, Marcos Antônio Vincenzi, Marcos Nascimento Magalhães, Maria Jacqueline Batista, Maria Kelly Venezuela, Maria Lucia Sobral Singer, Maria Paula Zanardi Chicarino, Mariana Teixeira Carballo, Michel Helcias Montoril, Mônica Carneiro Sandoval, Mônica S. Zanchetta, Nelson Ithiro Tanaka, Nuno Sepúlveda, Paulo Justiniano Ribeiro Junior, Paulo Ricardo Magalhães Rocha, Paulo Soares, Pedro Alberto Morettin, Pledson Guedes de Medeiros, Rafael Bráz Azevedo Farias, Rafael Izbicki, Raquel da Cunha Valle, Renato Martins Assunção, Renato Fadel Fava, Ricardo Fernandes, Ricardo Raitz, Ricardo Tadashi Takeyama, Rodrigo Andrade Tavares, Rodrigo Franco Polacco, Rodrigo Nati, Rogério Ruscitto do Prado, Rosana Francisco Alves, Saskia Litière, Sebastião Pinho, Silvia Emiko Shimakura, Silvia Lopes de Paula Ferrari, Silvio Teruo Watanabe, Stela Valeska Kretchetoff, Susana Miyuki Okaze Tomazela, Tatiana Salomão Miamoto, Thiago Vasques Moretti, Thomas Woerly, Victor Fossaluza, Victorine Broekmans e Willy Nielsen Patané.

Frederico Zanqueta Poleto

São Paulo, abril de 2011

Sumário

Resumo **v**

Abstract **vii**

Agradecimentos **ix**

1 Introdução **1**

2 Conceitos de modelagem de dados com omissão **5**

2.1 Notação e terminologia 5

2.2 Fatorações 6

2.3 Mecanismos de omissão 8

2.3.1 Modelos de seleção 8

2.3.2 Modelos de mistura de padrões 12

2.3.3 Modelos de parâmetros compartilhados 15

2.3.4 Resumo e outras abordagens 17

2.4 Análises de sensibilidade 18

3 Modelos sobreparametrizados para a análise de dados categorizados com respostas omissas e suas implicações inferenciais **21**

3.1 Abordagens inferenciais 23

3.1.1 Descrição do problema, notação e verossimilhança 23

3.1.2 Abordagem bayesiana 25

3.1.3 Análise de sensibilidade clássica 28

3.1.4 Diferenças entre as abordagens bayesiana e clássica e extensões 30

3.2 Estudo de simulação 31

3.2.1 Descrição das análises 31
3.2.2 Resultados 33
3.3 Reanálise dos dados do *Collaborative Perinatal Project* 40
3.4 Discussão 42

4 Análise bayesiana semiparamétrica de respostas binárias com uma covariável contínua sujeita a omissão informativa 45

4.1 Dados de embolia pulmonar 46
4.2 Modelos não-paramétricos para variáveis contínuas completas 48
4.3 Modelo semiparamétrico para respostas binárias com uma variável explicativa sujeita a omissão informativa 54
4.4 Estudo de simulação 55
4.5 Análise dos dados de embolia pulmonar 61
4.6 Discussão 64

5 Análise de sensibilidade clássica para a estimação de momentos em dados com omissão 67

5.1 Dados de faculdades americanas 68
5.2 Caso univariado 68
5.3 Caso multivariado 74
5.4 Avaliação dos intervalos de incerteza 77
5.5 Discussão 81

6 Considerações finais 83

6.1 Contribuições e conclusões deste trabalho 83
6.2 Sugestões de temas para pesquisas futuras 84

Referências bibliográficas 87

Índice de autores 97

Capítulo 1

Introdução

A análise de dados categorizados é um dos campos da análise multivariada que visa estudar métodos específicos para descrição e inferência estatística baseadas em variáveis naturalmente dispostas em categorias ou categorizadas a partir de variáveis contínuas. Como no processo de coleta de informações, especialmente em estudos longitudinais, algumas observações planejadas podem não ser obtidas, esses dados omissos ou faltantes (*missing*) também requerem metodologia especializada para a sua análise. Diversos autores têm abordado esse assunto. Em particular, Poleto (2006) considerou técnicas estatísticas para a análise de dados categorizados com omissão em variáveis respostas. Neste trabalho, desenvolve-se metodologias para analisar dados com omissão, procurando compreender melhor sua influência nos resultados das análises e considerando especialmente situações em que há omissão tanto nas variáveis respostas quanto nas variáveis explicativas. Especificamente, as contribuições dividem-se em:

1. Mostra-se que, independentemente do tamanho da amostra, análises de modelos sobreparametrizados requerem maior cuidado tanto na especificação das distribuições *a priori*, porque distribuições consideradas não-informativas ou levemente informativas podem na verdade ser bastante informativas, quanto na adoção de modelos sobreparametrizados parcimoniosos, porque não há informação amostral para justificar a redução da dimensão dos modelos estruturais e, se as suposições não representarem a verdade, podem levar a conclusões equivocadas.
2. Propõe-se uma análise semiparamétrica bayesiana para respostas binárias sempre observadas com uma covariável contínua sujeita a omissão informativa e ilustra-se que a mistura induzida pelo processo de Dirichlet é suficientemente flexível para acomodar uma certa variedade de distribuições para a covariável e que se pode realizar também diversas suposições para os mecanismos de omissão evitando, assim, vieses nas inferências de interesse.
3. Deriva-se uma análise de sensibilidade clássica para a estimação de momentos em dados com omissão evitando a utilização de suposições de normalidade para as respostas e possibilitando que se empregue parâmetros de sensibilidade de fácil interpretação, como médias, desvios padrões e correlações, ou funções deles, como a de relacionar médias não-observadas a quantis das distribuições observadas.

Um exemplo de dados categorizados com omissão é apresentado na Tabela 1.1. Ele é oriundo de um estudo prospectivo do Departamento de Saúde dos Estados Unidos, em que mães grávidas, fumantes ou não, foram acompanhadas até ao parto e os seus recém-nascidos foram classificados conforme o peso fosse < 2.5 kg ou ≥ 2.5 kg (Baker, Rosenberger & DerSimonian, 1992). O objetivo do estudo foi avaliar a associação entre o hábito de fumo da mãe e o peso do recém-nascido. 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.

Tabela 1.1: Frequências observadas do *Collaborative Perinatal Project*.

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

Outras situações em que estatísticos aplicados se deparam com dados omissos são (Vach, 1994):

i. Entrevistas e preenchimentos de questionários, em que a não-resposta (*non-response*) é uma causa típica de dados faltantes. Isso pode acontecer devido à falta de conhecimento da resposta, como quando o entrevistado é questionado sobre doenças de sua infância, recusa intencional, principalmente em questões embaraçosas como consumo de álcool ou drogas, atividades sexuais ou renda, ou indecisão sobre o assunto.

ii. Estudos longitudinais, em que pacientes abandonam (*drop out, attrition*) ou não comparecem às consultas em alguns dos instantes previstos. As razões podem ser mudança de endereço, ineficiência do tratamento, impressão de cura, surgimento de efeitos colaterais, morte, *etc.*

iii. Estudos retrospectivos, em que informações são geralmente obtidas de documentos, como registros hospitalares. A incompletude desses documentos também pode gerar dados omissos. Em estudos prospectivos, a inclusão de novas unidades experimentais pode perdurar por vários anos. Durante esse período, novos fatores de risco podem ser descobertos, o que pode fazer com que se decida coletar informações adicionais. As unidades que ingressaram antes dessa decisão terão resultados omissos para essas variáveis.

iv. Casos em que a mensuração de uma variável é muito cara e, por isso, restrita a um subconjunto de todas as unidades experimentais.

v. Estudos bem planejados e conduzidos, em que acidentes ocorrem. Um tubo de ensaio pode quebrar, um questionário pode ser perdido, um exame pode ser esquecido, a imprecisão de um instrumento pode ser descoberta muito tarde, *etc.* Cada acidente pode gerar um valor omisso.

vi. Casos em que informações estão limitadas a um subconjunto das alternativas, diferentemente dos casos tradicionais em que a resposta de uma variável é ou não observada. O respondente de

um inquérito pode ficar indeciso com relação a duas ou mais opções, pode haver dúvida quanto à classificação de uma resposta entre duas categorias subjacentes, um registro pode ter sido preenchido dubiamente e ter sua resposta confundida entre algumas das alternativas, *etc.*

Em virtude das inúmeras possibilidades de se encontrar dados com omissão, não se deve estranhar que Wilks (1932) já discutisse as dificuldades que dados faltantes causam no processo de inferência estatística. O aumento contínuo de interesse científico sobre o tema nas últimas décadas pode ser observado pelo número crescente de artigos, pelo aparecimento de cada vez mais livros específicos (Rubin, 1987; Schafer, 1997; Vach, 1994; Allison, 2001; Little & Rubin, 2002; Groves, Dillman, Eltinge & Little, 2002; Van Der Laan & Robins, 2002; Särndal & Lundström, 2005; Tsiatis, 2006; Molenberghs & Kenward, 2007; Daniels & Hogan, 2007; Tan, Tian & Ng, 2010) e até mesmo pela ampliação da quantidade de livros de outros assuntos que passam a incluir um ou mais capítulos especializados (*e.g.*, Särndal, Swensson & Wretman, 1991; Halloran & Berry, 1999; Korn & Graubard, 1999; Levy & Lemeshow, 1999; Lindsey, 1999; Verbeke & Molenberghs, 2000; Harrell, 2001; Diggle, Heagerty, Liang & Zeger, 2002; Hagenaars & McCutcheon, 2002; Srivastava, 2002; Twisk, 2002; Chambers & Skinner, 2003; Fleiss, Levin & Paik, 2003; Fitzmaurice, Laird & Ware, 2004; Gelman & Meng, 2004; Lehtonen & Pahkinen, 2004; Congdon, 2005, 2006; Molenberghs & Verbeke, 2005; Weiss, 2005; Gelman & Hill, 2006; Hedeker & Gibbons, 2006; Paulino & Singer, 2006; De Leeuw & Meijer, 2008; Fitzmaurice, Davidian, Verbeke & Molenberghs, 2008).

Quando existem dados omissos, uma abordagem simples e por isso largamente empregada é a de realizar uma análise convencional utilizando apenas as unidades experimentais em que todas as variáveis foram observadas, denominada análise de casos completos (ACC, *complete case analysis*). Quando o interesse momentâneo recai apenas sobre as distribuições marginais (univariadas ou bivariadas), a ACC pode desperdiçar muita informação. Nestas circunstâncias, pode ser mais indicado utilizar todas as unidades experimentais que tiveram respostas observadas para a variável (ou para as duas variáveis); esta abordagem é chamada de análise de casos disponíveis (ACD, *available case analysis*)[1]. Para a Tabela 1.1, por exemplo, a ACC utiliza 93% das unidades experimentais, enquanto que a ACD referente ao hábito de fumo da mãe abrange 97% (= 93% + 4%) das respostas.

Métodos estatísticos específicos desenvolvidos para lidar com o problema de dados com omissão podem ser grosseiramente agrupados em três classes.

1. A extensão dos métodos tradicionais de estimação em pesquisas de levantamento de dados (*surveys*), com a utilização de pesos. A maior parte dos livros de amostragem apresenta capítulos sobre o assunto; uma revisão mais abrangente é apresentada em Särndal & Lundström (2005), enquanto que uma introdução pode ser encontrada em Little & Rubin (2002, Seção 3.3).

2. A adoção de procedimentos de imputação de dados para, de alguma forma, substituir os dados faltantes por estimativas com o intuito de utilizar os métodos tradicionais para a análise de

[1] Os *softwares* estatísticos costumam realizar a ACC como padrão. Em alguns casos, análises exploratórias podem ser realizadas tanto com ACC quanto com ACD e, geralmente, são designadas *listwise* e *pairwise*, respectivamente.

dados completos. Little & Rubin (2002, Cap. 4) revisaram as abordagens mais utilizadas, como a imputação de médias incondicionais ou condicionais ou o sorteio de valores de distribuições preditivas de modelos explícitos ou implícitos para os dados faltantes. Em estudos longitudinais, a substituição dos dados omissos pelo último valor observado (*last observation carried forward*) é também muito comum, embora esta prática esteja sendo largamente combatida, conforme indicam Molenberghs & Kenward (2007).

As abordagens de imputação de um único valor (*single imputation*) para cada dado omisso não levam em conta a incerteza que se tem dos valores faltantes. Para contornar esse inconveniente, o método de imputação múltipla (*multiple imputation*) consiste na substituição de cada dado omisso por dois ou mais valores, seguida da análise de cada conjunto de dados completado e, finalmente, da combinação das inferências por meio de ajustes especiais (Rubin, 1987; Schafer, 1997; Little & Rubin, 2002). Embora algumas das abordagens descritas até aqui pareçam simples, as suposições sob as quais esses procedimentos *ad hoc* fornecem análises válidas nem sempre ficam claras.

3. A alternativa adotada neste trabalho, considera explicitamente um mecanismo gerador dos dados omissos, *i.e.*, mecanismo de omissão (*missingness mechanism*) no processo de modelagem probabilística e estrutural (Little & Rubin, 2002; Molenberghs & Kenward, 2007). Inferências frequentistas, verossimilhancistas ou bayesianas podem ser utilizadas para tirar conclusões apenas sobre o fenômeno de interesse, somente estabelecendo quais foram as suposições incluídas no mecanismo de omissão, e/ou pode-se tratar de questões sobre o processo gerador dos dados faltantes.

No Capítulo 2, revisam-se alguns conceitos e termos da modelagem de dados com omissão sob os diferentes paradigmas inferenciais, evidenciando-se que os modelos probabilísticos para dados com omissão são, em geral, sobreparametrizados. Por esta razão, no Capítulo 3 escrutinam-se análises de sensibilidade bayesiana e clássica com modelos sobreparametrizados para o caso de respostas categorizadas sujeitas a omissão. No Capítulo 4, propõe-se metodologia para analisar respostas binárias com variáveis explicativas contínuas sujeitas a omissão. No Capítulo 5, deriva-se uma nova análise de sensibilidade clássica para respostas incompletas quando o interesse recai apenas em momentos da distribuição das respostas. No Capítulo 6, resumem-se as principais contribuições e conclusões obtidas e sugerem-se alguns temas para pesquisas futuras.

A notação e os acrônimos utilizados são apresentados na primeira vez em que aparecem em cada um dos capítulos, para que estes possam ser lidos independentemente. Com exceção do pacote computacional OpenBUGS (Lunn, Spiegelhalter, Thomas & Best, 2009) que se usa no Capítulo 4 para obter amostras das distribuições *a posteriori*, todos os demais cálculos da tese foram efetuados no ambiente estatístico R (R Development Core Team, 2010). Para a análise da convergência das distribuições *a posteriori* nos Capítulos 3 e 4 utilizou-se, respectivamente, os pacotes boa (*Bayesian output analysis*, Smith, 2007) e coda (*convergence diagnosis and output analysis*, Plummer, Best, Cowles & Vines, 2006), do R.

Capítulo 2

Conceitos de modelagem de dados com omissão

No capítulo anterior introduziram-se as principais classes de procedimentos para lidar com dados incompletos. Neste capítulo revisa-se a modelagem de dados com omissão com o intuito de apresentar um panorama da literatura sobre esta temática. A natureza categorizada das variáveis respostas é evidenciada no uso de somatórios, que seriam substituídos por integrais no caso de respostas contínuas. Poleto (2006, Cap. 1) apresenta uma introdução similar, particularizada para o caso de duas variáveis respostas dicotômicas.

2.1 Notação e terminologia

Suponha que, para cada uma de n unidades experimentais, planeja-se obter o resultado de J variáveis ou de uma mesma variável medida em J ocasiões, Y_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, J$. Então, para a i-ésima unidade, pode-se agrupar as informações no vetor $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iJ})'$. Adicionalmente, definem-se as variáveis aleatórias

$$R_{ij} = \begin{cases} 1, & \text{se } Y_{ij} \text{ é observado,} \\ 0, & \text{em caso contrário.} \end{cases}$$

Agrupam-se estes indicadores de dados observados[1] no vetor $\mathbf{R}_i = (R_{i1}, \ldots, R_{iJ})'$, com dimensão igual à de $\mathbf{Y}_i$. Pode-se subdividir $\mathbf{Y}_i$ em duas partes, $\mathbf{Y}_i^o$ e $\mathbf{Y}_i^m$, de maneira que $\mathbf{Y}_i^o$ contenha os componentes observados, $\{R_{ij} = 1\}$, e $\mathbf{Y}_i^m$ inclua os omissos, $\{R_{ij} = 0\}$.

O vetor $\mathbf{Y}_i$ de observações planejadas é designado por **dados completos** (*complete data*). Esse seria o vetor observado se não houvesse omissão. O processo gerador do vetor $\mathbf{R}_i$ é chamado de **mecanismo de omissão**. Os dados completos juntamente com os indicadores de dados observados, $(\mathbf{Y}_i, \mathbf{R}_i)$, são designados por **dados totais** (*full data*). Os dados totais são observados apenas quando todos os componentes de $\mathbf{R}_i$ são iguais a 1. Nos outros casos, observa-se apenas $(\mathbf{Y}_i^o, \mathbf{R}_i)$, chamados de **dados observados** (*observed data*). Para deixar evidente que parte dos dados totais está em

[1] Note-se que essas variáveis não contemplam as formas de omissão menos comuns descritas no item (vi) da página 2, em que há informação ou censura parcial. Esses casos podem ser incluídos definindo categorias adicionais para R_{ij}. Modificações similares permitem que se trate adicionalmente de erros de classificação (Soares, 2004).

falta, alguns autores preferem designá-los por **dados ampliados** (*augmented data*).

Cada configuração de $\{\mathbf{R}_i, i = 1, \ldots, n\}$ define um **padrão de omissão**. Destaque especial é dedicado ao **padrão de omissão monótono** (Rubin, 1974), em que as variáveis Y_{ij} são ordenadas em $\mathbf{Y}_i$, numa mesma disposição para todo i, de tal maneira que os vetores $\mathbf{R}_i$ de todos os indivíduos são divididos em dois grupos, os primeiros com elementos iguais a 1 e os últimos com elementos iguais a 0; por exemplo, para $J = 3$ poderíamos ter $\mathbf{R}_i \in \{(1,1,1)', (1,1,0)', (1,0,0)', (0,0,0)'\}$. Sob este padrão, $R_{ij} = 1$ indica que pelo menos a sequência $Y_{i1}, \ldots, Y_{ij}$ foi observada e $R_{ij} = 0$ indica que a sequência $Y_{ij}, \ldots, Y_{iJ}$ está omissa. Este costuma ser o padrão de omissão mais comum em estudos longitudinais, em que os indivíduos são observados até um determinado instante, mas em virtude de algum motivo acabam abandonando-o (*drop out, attrition*). Essa é a motivação para se chamar o escalar $D_i = 1 + \sum_j R_{ij}$ de indicador (do instante) de abandono, que pode substituir o vetor $\mathbf{R}_i$ sem perda de informação. Algumas vezes, pode ser mais conveniente utilizar o indicador alternativo $T_i = D_i - 1$, que indica o número de observações obtidas.

No que segue, realizações de variáveis aleatórias são representadas pelas suas letras minúsculas correspondentes, *e.g.*, y_{ij} e $\mathbf{y}_i$ para Y_{ij} e $\mathbf{Y}_i$, respectivamente.

2.2 Fatorações

Para levar em conta o mecanismo gerador dos dados omissos bem como o processo de mensuração/categorização dos dados completos, um ponto de partida é a função de probabilidade conjunta dos dados totais

$$f(\mathbf{y}_i, \mathbf{r}_i | \boldsymbol{\theta}, \boldsymbol{\psi}), \tag{2.1}$$

em que $\boldsymbol{\theta}$ é o vetor de parâmetros relacionados com a distribuição marginal do processo de mensuração/categorização e $\boldsymbol{\psi}$ é o vetor que inclui parâmetros associados à distribuição marginal do mecanismo de omissão e também parâmetros associados à dependência entre $\mathbf{Y}_i$ e $\mathbf{R}_i$. Embora os parâmetros possam também depender de variáveis explicativas $\mathbf{X}_i$, estas são suprimidas para simplificar a notação. Neste capítulo, supõe-se que as variáveis de $\mathbf{X}_i$ são sempre observadas.

Modelos obtidos por meio da fatoração

$$f(\mathbf{y}_i, \mathbf{r}_i | \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{y}_i | \boldsymbol{\theta}) f(\mathbf{r}_i | \mathbf{y}_i, \boldsymbol{\psi}) \tag{2.2}$$

são chamados de **modelos de seleção** (**SeM**, *selection models*); o primeiro termo é a função de probabilidade marginal do mecanismo de mensuração/categorização e o segundo é a função de probabilidade do processo de omissão condicional aos resultados. Essa nomenclatura, com origem na área econométrica (Heckman, 1976, 1979), se deve ao fato de $f(\mathbf{r}_i | \mathbf{y}_i, \boldsymbol{\psi})$ poder ser encarada como um mecanismo individual de autosseleção das componentes de $\mathbf{Y}_i$ observadas ou não, condicionalmente aos seus valores.

Outra opção de fatoração, nomeadamente

$$f(\mathbf{y}_i, \mathbf{r}_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{y}_i|\mathbf{r}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) f(\mathbf{r}_i|\boldsymbol{\psi}), \tag{2.3}$$

especifica um modelo condicional para o processo de mensuração/categorização dados os padrões de omissão e um modelo marginal para o mecanismo de omissão (Glynn, Laird & Rubin, 1986; Little, 1993). A designação de **modelos de mistura de padrões** (**PMM**, *pattern-mixture models*) justifica-se pelo fato de a distribuição marginal dos dados completos resultar numa mistura das distribuições condicionais aos diferentes padrões de omissão, *i.e.*,

$$f(\mathbf{y}_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = \sum_{\mathbf{r}_i} f(\mathbf{y}_i|\mathbf{r}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) f(\mathbf{r}_i|\boldsymbol{\psi}). \tag{2.4}$$

As ideias oriundas das duas fatorações podem ser combinadas nos **modelos de mistura de conjuntos de padrões** (*pattern-set mixture models*) por meio de

$$f\left(\mathbf{y}_i, \mathbf{r}_i \middle| \boldsymbol{\theta}, \boldsymbol{\psi}\right) = f\left(\mathbf{y}_i \middle| \mathbf{r}_i^{(1)}, \boldsymbol{\theta}, \boldsymbol{\psi}\right) f\left(\mathbf{r}_i^{(2)} \middle| \mathbf{y}_i, \mathbf{r}_i^{(1)}, \boldsymbol{\theta}, \boldsymbol{\psi}\right) f\left(\mathbf{r}_i^{(1)} \middle| \boldsymbol{\psi}\right) \tag{2.5}$$

após subdividir as variáveis indicadoras de omissão em duas partes, *i.e.*, $\mathbf{R}_i = (\mathbf{R}_i^{(1)}, \mathbf{R}_i^{(2)})$, por exemplo, com $\mathbf{R}_i^{(1)} = (R_{i1}, \ldots, R_{ij})'$ e $\mathbf{R}_i^{(2)} = (R_{i,j+1}, \ldots, R_{iJ})'$ (Little, 1993; Little & Rubin, 2002).

Os **modelos de parâmetros compartilhados** (**SPM**, *shared-parameter models*), por sua vez, consideram um vetor de efeitos aleatórios ou latentes, $\mathbf{b}_i$, para cada unidade experimental, do qual um ou mais componentes são compartilhadas pelos processos de mensuração/categorização e de omissão, juntamente com uma das fatorações apresentadas (Wu & Carroll, 1988; Wu & Bailey, 1988, 1989). Por exemplo, ao utilizar SeM, a fatoração toma a forma

$$f(\mathbf{y}_i, \mathbf{r}_i|\boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{b}_i) = f(\mathbf{y}_i|\boldsymbol{\theta}, \mathbf{b}_i) f(\mathbf{r}_i|\mathbf{y}_i, \boldsymbol{\psi}, \mathbf{b}_i). \tag{2.6}$$

Em muitas situações, supõe-se que $\mathbf{Y}_i$ e $\mathbf{R}_i$ são condicionalmente independentes, dados os efeitos aleatórios $\mathbf{b}_i$, caso em que tanto a utilização da fatoração SeM quanto PMM leva a

$$f(\mathbf{y}_i, \mathbf{r}_i|\boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{b}_i) = f(\mathbf{y}_i|\boldsymbol{\theta}, \mathbf{b}_i) f(\mathbf{r}_i|\boldsymbol{\psi}, \mathbf{b}_i). \tag{2.7}$$

Os parâmetros compartilhados de $\mathbf{b}_i$ podem ser encarados como traços latentes governando ambos os processos de mensuração/categorização e de omissão.

Por fim, utilizando a partição dos dados completos com componentes $\mathbf{Y}_i^o$ e $\mathbf{Y}_i^m$, a fatoração

$$f\left(\mathbf{y}_i, \mathbf{r}_i \middle| \boldsymbol{\theta}, \boldsymbol{\psi}\right) = f\left(\mathbf{y}_i^m \middle| \mathbf{y}_i^o, \mathbf{r}_i, \boldsymbol{\theta}, \boldsymbol{\psi}\right) f\left(\mathbf{y}_i^o \middle| \mathbf{r}_i, \boldsymbol{\theta}, \boldsymbol{\psi}\right) f\left(\mathbf{r}_i \middle| \boldsymbol{\psi}\right) \tag{2.8}$$

é uma extensão bastante natural de PMM, sendo utilizada em algumas situações para obtenção de

distribuições *a posteriori* (Paulino, 1988; Paulino & Pereira, 1992, 1995; Soares & Paulino, 2001; Soares, 2004). O primeiro termo do segundo membro de (2.8) é denominado de modelo de imputação para os dados faltantes a partir dos dados observados e desempenha um papel fundamental na etapa de imputação da abordagem de imputação múltipla (Rubin, 1987).

2.3 Mecanismos de omissão

Nas próximas subseções apresentam-se definições de mecanismos de omissão sob as três principais fatorações introduzidas.

2.3.1 Modelos de seleção

Três famílias de mecanismos de omissão, discutidas por Rubin (1976), desempenham um papel fundamental neste contexto. Esses mecanismos são mais facilmente expressos por meio da fatoração SeM, baseando-se no segundo fator do segundo membro de (2.2) e atentando-se para a distinção do que foi ou não observado, ou seja,

$$f(\mathbf{r}_i|\mathbf{y}_i, \boldsymbol{\psi}) = f(\mathbf{r}_i|\mathbf{y}_i^o, \mathbf{y}_i^m, \boldsymbol{\psi}). \tag{2.9}$$

Sob um mecanismo de **omissão não-informativa** ou **aleatória** (**MAR**, *missing at random*), a probabilidade de um resultado estar em falta é condicionalmente independente do valor omisso, dados os resultados observados, *i.e.*,

$$f(\mathbf{r}_i|\mathbf{y}_i, \boldsymbol{\psi}) = f(\mathbf{r}_i|\mathbf{y}_i^o, \boldsymbol{\psi}). \tag{2.10}$$

Essa nomenclatura advém do fato de os dados omissos poderem ser extrapolados a partir dos dados observados sem qualquer influência do mecanismo de omissão, conforme indicado pelo modelo de imputação

$$\begin{aligned} f(\mathbf{y}_i^m|\mathbf{y}_i^o, \mathbf{r}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) &= \frac{f(\mathbf{y}_i^o, \mathbf{y}_i^m, \mathbf{r}_i|\boldsymbol{\theta}, \boldsymbol{\psi})}{f(\mathbf{y}_i^o, \mathbf{r}_i|\boldsymbol{\theta}, \boldsymbol{\psi})} = \frac{f(\mathbf{y}_i^o, \mathbf{y}_i^m|\boldsymbol{\theta}) f(\mathbf{r}_i|\mathbf{y}_i^o, \mathbf{y}_i^m, \boldsymbol{\psi})}{\sum_{\mathbf{y}_i^m} f(\mathbf{y}_i^o, \mathbf{y}_i^m|\boldsymbol{\theta}) f(\mathbf{r}_i|\mathbf{y}_i^o, \mathbf{y}_i^m, \boldsymbol{\psi})} \\ &\overset{\text{MAR}}{=} \frac{f(\mathbf{y}_i^o, \mathbf{y}_i^m|\boldsymbol{\theta}) f(\mathbf{r}_i|\mathbf{y}_i^o, \boldsymbol{\psi})}{\sum_{\mathbf{y}_i^m} f(\mathbf{y}_i^o, \mathbf{y}_i^m|\boldsymbol{\theta}) f(\mathbf{r}_i|\mathbf{y}_i^o, \boldsymbol{\psi})} = \frac{f(\mathbf{y}_i^o, \mathbf{y}_i^m|\boldsymbol{\theta})}{f(\mathbf{y}_i^o|\boldsymbol{\theta})} = f(\mathbf{y}_i^m|\mathbf{y}_i^o, \boldsymbol{\theta}) \end{aligned} \tag{2.11}$$

que deixa de depender dos padrões de omissão sob MAR.

Um caso particular do mecanismo MAR é o mecanismo de **omissão completamente aleatória** (**MCAR**, *missing completely at random*), em que a probabilidade de uma resposta não ser observada independe dos resultados (observados e faltantes), levando a

$$f(\mathbf{r}_i|\mathbf{y}_i, \boldsymbol{\psi}) = f(\mathbf{r}_i|\boldsymbol{\psi}). \tag{2.12}$$

Os mecanismos que não são MAR são chamados de mecanismos de **omissão informativa** ou **não-aleatória** (**MNAR**, *missing not at random*). Nesses casos, a probabilidade de um resultado estar em falta depende de algum modo das respostas não observadas. Consequentemente, a função de probabilidade (2.9) depende de $\mathbf{y}_i^m$ e também pode depender de $\mathbf{y}_i^o$.

Como as inferências baseiam-se no que é observado, o foco das inferências verossimilhancistas recai sobre a função de verossimilhança dos dados observados, em que a contribuição da i-ésima unidade é dada por

$$L(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}_i^o, \mathbf{r}_i) \propto f(\mathbf{y}_i^o, \mathbf{r}_i|\boldsymbol{\theta}, \boldsymbol{\psi}), \tag{2.13}$$

com

$$f(\mathbf{y}_i^o, \mathbf{r}_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = \sum_{\mathbf{y}_i^m} f(\mathbf{y}_i, \mathbf{r}_i|\boldsymbol{\theta}, \boldsymbol{\psi}). \tag{2.14}$$

Sob um mecanismo MAR (2.10), a função de probabilidade é simplificada para

$$\begin{aligned} f(\mathbf{y}_i^o, \mathbf{r}_i|\boldsymbol{\theta}, \boldsymbol{\psi}) &= \sum_{\mathbf{y}_i^m} f(\mathbf{y}_i|\boldsymbol{\theta}) f(\mathbf{r}_i|\mathbf{y}_i^o, \boldsymbol{\psi}) \\ &= f(\mathbf{y}_i^o|\boldsymbol{\theta}) f(\mathbf{r}_i|\mathbf{y}_i^o, \boldsymbol{\psi}), \end{aligned} \tag{2.15}$$

e, sob o seu caso particular MCAR (2.12), ela é reduzida para

$$f(\mathbf{y}_i^o, \mathbf{r}_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{y}_i^o|\boldsymbol{\theta}) f(\mathbf{r}_i|\boldsymbol{\psi}). \tag{2.16}$$

Portanto, as verossimilhanças sob os mecanismos MAR e MCAR se fatoram em duas partes, em que uma depende apenas do processo de mensuração/categorização e a outra, do mecanismo de omissão. Por conseguinte, se $\boldsymbol{\theta}$ e $\boldsymbol{\psi}$ forem funcionalmente não-relacionados (ou distintos), no sentido de o espaço paramétrico de $(\boldsymbol{\theta}', \boldsymbol{\psi}')'$ ser o produto dos espaços paramétricos de $\boldsymbol{\theta}$ e $\boldsymbol{\psi}$, tanto sob o mecanismo MAR quanto sob o MCAR, inferências sobre $\boldsymbol{\theta}$ baseadas unicamente na verossimilhança podem ser realizadas utilizando apenas o primeiro termo do lado direito de (2.15) e (2.16). Como o mecanismo de omissão é ignorado nesses casos para se fazer inferências sobre $\boldsymbol{\theta}$, muitas vezes chama-se o mecanismo MAR de mecanismo de **omissão ignorável**. Esse mecanismo de omissão também pode ser ignorado para a realização de inferências bayesianas sobre $\boldsymbol{\theta}$ se, adicionalmente, as distribuições *a priori* de $\boldsymbol{\theta}$ e $\boldsymbol{\psi}$ forem independentes.

Contudo, como o mecanismo MAR depende do que é observado, a distribuição dos dados observados continua a depender de $\boldsymbol{\psi}$ mesmo após ser condicionada nos padrões de omissão obtidos, *i.e.*,

$$f(\mathbf{y}_i^o|\mathbf{r}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) = \frac{f(\mathbf{y}_i^o|\boldsymbol{\theta}) f(\mathbf{r}_i|\mathbf{y}_i^o, \boldsymbol{\psi})}{\sum_{\mathbf{y}_i^o} f(\mathbf{y}_i^o|\boldsymbol{\theta}) f(\mathbf{r}_i|\mathbf{y}_i^o, \boldsymbol{\psi})}. \tag{2.17}$$

Por esse motivo, o mecanismo MAR não pode ser ignorado para se fazer inferências frequentistas sobre $\boldsymbol{\theta}$. Por outro lado, sob a suposição MCAR, $\mathbf{r}_i$ é uma estatística suficiente parcial (S-suficiente) para

$\boldsymbol{\psi}$[2], ou equivalentemente, ancilar parcial (S-ancilar) para $\boldsymbol{\theta}$ (Lindsey, 1996). Portanto, pelo princípio generalizado da condicionalidade (Basu, 1977) pode-se analisar os dados com base na função de probabilidade $f(\mathbf{y}_i^o|\boldsymbol{\theta})$, que não inclui o parâmetro perturbador (*nuisance*) $\boldsymbol{\psi}$, sem incorrer em perda de informação relevante sobre o parâmetro de interesse, $\boldsymbol{\theta}$. Consequentemente, sob o mecanismo MCAR, os dados observados podem ser analisados como se os padrões de omissão tivessem sido fixados pelo planejamento amostral e o processo de omissão é também ignorável sob o ponto de vista das inferências frequentistas sobre $\boldsymbol{\theta}$. Neste contexto, quando se pretende utilizar o método da máxima verossimilhança e a teoria assintótica convencional, o mecanismo MAR ainda pode ser ignorado se a matriz de informação de Fisher for substituída pela matriz de informação observada de $f(\mathbf{y}_i^o|\boldsymbol{\theta})$ na obtenção da matriz de covariância assintótica do estimador de $\boldsymbol{\theta}$ e nas estatísticas de teste de Wald e *score* de Rao (Kenward & Molenberghs, 1998).

A análise de casos completos (ACC) utiliza apenas as unidades experimentais que não tiveram dados omissos, ou seja, condiciona os dados observados ao padrão de omissão $\mathbf{r}_i = (1, 1, \ldots, 1)'$. O resultado (2.17) permite antever que essas inferências sob o mecanismo MAR continuam a depender de $\boldsymbol{\psi}$, sendo, assim, inapropriadas para fazer inferências apenas sobre $\boldsymbol{\theta}$. Portanto, apenas sob o mecanismo MCAR, o condicionamento realizado pela ACC permite ignorar o mecanismo de omissão. Na prática, isso quer dizer que as inferências obtidas numa ACC são, em geral, enviesadas se o verdadeiro mecanismo de omissão for MAR e são sempre não-enviesadas se for MCAR. Contudo, as inferências obtidas sob o mecanismo MCAR são, em geral, mais eficientes do que as obtidas na ACC, pois a última estratégia desconsidera as informações das unidades amostrais parcialmente observadas. Poleto, Singer & Paulino (2011a) ilustram essa perda de informação quando há interesse em se comparar a precisão de testes diagnósticos e mostram que mesmo num caso em que o mecanismo MCAR é plausível, as conclusões obtidas na ACC, e outras análises que não levam em conta toda a informação disponível, podem ser equivocadas. Além de exemplificar esse fato, Poleto, Singer & Paulino (2011b) apresentam exceções em que, para algumas funções paramétricas de interesse, os estimadores obtidos na ACC podem ser (1) consistentes sob um mecanismo MAR e (2) tão eficientes quanto os do verdadeiro mecanismo MCAR.

Sob um mecanismo MNAR, a função de probabilidade dos dados observados não pode ser fatorada em duas partes, como nos mecanismos anteriores, conforme evidenciado em

$$f(\mathbf{y}_i^o, \mathbf{r}_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = \sum_{\mathbf{y}_i^m} f(\mathbf{y}_i|\boldsymbol{\theta}) f(\mathbf{r}_i|\mathbf{y}_i^o, \mathbf{y}_i^m, \boldsymbol{\psi}). \tag{2.18}$$

Logo, sob qualquer das três abordagens inferenciais para $\boldsymbol{\theta}$ o mecanismo MNAR é um mecanismo de **omissão não-ignorável**. Além disso, enquanto os modelos MAR e MCAR são identificáveis, os modelos MNAR mais gerais são inidentificáveis. Assim, inferências frequentistas e verossimilhancistas

[2]Suficiente para $\boldsymbol{\psi}$ para cada valor de $\boldsymbol{\theta}$ (*i.e.*, suficiente específica para $\boldsymbol{\psi}$) e ancilar para $\boldsymbol{\theta}$ para cada valor de $\boldsymbol{\psi}$ (*i.e.*, ancilar específica para $\boldsymbol{\theta}$).

só podem ser realizadas após a inclusão de suposições sobre as partes dos dados observados e omissos de que o mecanismo de omissão depende, deixando, assim, o modelo identificável. Tais requisitos não são exigidos sob a ótica bayesiana, que viabiliza a análise sem a inclusão de qualquer restrição estrutural no mecanismo de omissão embora, em geral, requeira a utilização de distribuições *a priori* próprias. É de se esperar, todavia, que os dados observados não sejam suficientes para atualizar toda a informação *a priori*, o que implica que certas funções paramétricas continuem, *a posteriori*, dependentes apenas da distribuição *a priori* escolhida (Paulino, 1988; Paulino & Pereira, 1992, 1995; Walker, 1996; Soares & Paulino, 2001; Soares, 2004; Jiang & Dickey, 2008).

Em análises clássicas de dados categorizados com omissão, modelos MNAR saturados podem não ter ajustes perfeitos e isto está associado à obtenção de estimativas de $\boldsymbol{\psi}$ na fronteira do espaço paramétrico (Baker & Laird, 1988), embora isto também possa ocorrer no caso de modelos estruturais reduzidos (Smith, Skinner & Clarke, 1999). Além disso, modelos MNAR saturados também podem ter seus parâmetros inidentificáveis conforme indica Glonek (1999). Molenberghs, Goetghebeur, Lipsitz & Kenward (1999), Clarke (2002) e Clarke & Smith (2004) também discutem essas patologias. Por meio de um estudo de simulação, Poleto *et al.* (2011b) mostraram que, tanto no caso de as estimativas de $\boldsymbol{\psi}$ estarem na fronteira do espaço paramétrico quanto no caso de falta de identificabilidade, os estimadores de máxima verossimilhança são enviesados, embora com viés menor do que aquele obtido sob a ACC se o mecanismo MNAR proposto for verdadeiro. Esses autores também mostraram que a obtenção de estimativas na fronteira do espaço paramétrico não é um indício de que o mecanismo de omissão seja falso e, dependendo da estrutura para o mecanismo de omissão, pode ainda ocorrer com probabilidade razoável para amostras grandes (*e.g.*, probabilidade de 18% para $n = 10\,000$). Vários dos trabalhos citados neste parágrafo também chamaram a atenção para as dificuldades de se obter convergência nos processos iterativos que visam maximizar a função de verossimilhança nas situações em que as estimativas de $\boldsymbol{\psi}$ estão na fronteira do espaço paramétrico. Em alguns destes casos, se os algoritmos de otimização não incluírem restrições, pode-se obter estimativas negativas ou maiores do que um para as probabilidades! Estas dificuldades computacionais não ocorrem quando se utiliza a fatoração PMM ao invés da fatoração SeM.

Em estudos longitudinais com padrão de omissão monótono é razoável admitir que um abandono não depende de valores futuros. Diggle & Kenward (1994) exploraram esta ideia, que no contexto da fatoração SeM pode ser traduzida por

$$f\left(t_i = s \,|y_{i1}, y_{i2}, \ldots, y_{iJ}, \boldsymbol{\theta}, \boldsymbol{\psi}\right) = f\left(t_i = s \,|y_{i1}, y_{i2}, \ldots, y_{i,s+1}, \boldsymbol{\theta}, \boldsymbol{\psi}\right), \quad \forall s. \tag{2.19}$$

Kenward, Molenberghs & Thijs (2003) chamaram este mecanismo de omissão não-dependente do futuro (MNFD, *missing non-future dependence*) porque num estudo longitudinal (2.19) traduz a suposição de que a probabilidade de se observar apenas os s primeiros instantes de tempo depende apenas dessas s respostas observadas e da primeira omissa, mas não das demais $J-s-1$ que também não foram observadas. O mecanismo MNFD por si só não traduz um modelo identificável, embora

seja um modelo MNAR mais reduzido que o MNAR geral (sem qualquer restrição); detalhes serão discutidos na próxima seção.

2.3.2 Modelos de mistura de padrões

Embora as famílias de mecanismos de omissão discutidas por Rubin (1976) sejam mais facilmente expressas por meio da fatoração SeM, vários autores têm utilizado a fatoração PMM como alternativa. Os mecanismos de omissão são, neste caso, especificados majoritariamente utilizando o primeiro fator do segundo membro de (2.3), *i.e.*, $f(\mathbf{y}_i|\mathbf{r}_i, \boldsymbol{\theta}, \boldsymbol{\psi})$. Note-se que ao obter a distribuição dos dados observados a partir de (2.8), nomeadamente

$$f\left(\mathbf{y}_i^o, \mathbf{r}_i \,|\boldsymbol{\theta}, \boldsymbol{\psi}\right) = \sum_{\mathbf{y}_i^m} f\left(\mathbf{y}_i^m \,|\mathbf{y}_i^o, \mathbf{r}_i, \boldsymbol{\theta}, \boldsymbol{\psi}\right) f\left(\mathbf{y}_i^o \,|\mathbf{r}_i, \boldsymbol{\theta}, \boldsymbol{\psi}\right) f\left(\mathbf{r}_i \,|\boldsymbol{\psi}\right) = f\left(\mathbf{y}_i^o \,|\mathbf{r}_i, \boldsymbol{\theta}, \boldsymbol{\psi}\right) f\left(\mathbf{r}_i \,|\boldsymbol{\psi}\right),$$

fica evidente que o modelo de imputação, *i.e.*, a distribuição dos dados faltantes condicionalmente aos dados observados, é o termo para o qual não há informação amostral. Logo, funções de $(\boldsymbol{\theta}, \boldsymbol{\psi})$ que estejam presentes apenas nesse componente não aparecerão na verossimilhança dos dados observados. Numa ótica bayesiana, se estas funções paramétricas forem independentes *a priori* das demais, elas não terão suas distribuições *a priori* atualizadas; em caso contrário, a atualização será induzida apenas pela dependência suposta *a priori* entre essas funções e as que aparecem na verossimilhança (Scharfstein, Daniels & Robins, 2003).

Tal como na fatoração SeM, os modelos mais gerais são inidentificáveis e, portanto, inferências clássicas só podem ser realizadas após a inclusão de restrições identificadoras. Com esta finalidade, Little (1993, 1994) incorpora tais restrições igualando os parâmetros de padrões com alguma omissão inestimáveis a (funções de) parâmetros que descrevem as distribuições das unidades completamente categorizadas. Por exemplo, nas restrições de valores omissos (oriundas) de casos completos (CCMV, *complete case missing values*), informações indisponíveis são tomadas por empréstimo dos casos completamente categorizados, ou seja, podem ser definidas como

$$f\left(\mathbf{y}_i^m \,|\mathbf{y}_i^o, \mathbf{r}_i = \mathbf{s}, \boldsymbol{\theta}, \boldsymbol{\psi}\right) = f\left(\mathbf{y}_i^m \,\big|\mathbf{y}_i^o, \mathbf{r}_i = (1,1,\ldots,1)', \boldsymbol{\theta}, \boldsymbol{\psi}\right), \quad \forall \mathbf{s} \neq (1,1,\ldots,1)', \tag{2.20}$$

em que as subdivisões $\mathbf{y}_i^o$ e $\mathbf{y}_i^m$ do segundo membro são iguais às do primeiro; no caso do padrão de omissão monótono, essas restrições podem ser equivalentemente expressas por[3]

$$f\left(y_{ij} \,|y_{i1}, y_{i2}, \ldots, y_{i,j-1}, t_i = s, \boldsymbol{\theta}, \boldsymbol{\psi}\right) = f\left(y_{ij} \,|y_{i1}, y_{i2}, \ldots, y_{i,j-1}, t_i = J, \boldsymbol{\theta}, \boldsymbol{\psi}\right), \quad j = s{+}1, s{+}2, \ldots, J. \tag{2.21}$$

Para ficar mais clara a equivalência entre (2.20) e (2.21), note-se que, para $J = 3$, se tem (veja a

[3]Obviamente, caso se deseje utilizar o indicador D_i ao invés de T_i, basta substituir t_i por $d_i - 1$ nas expressões correspondentes.

Tabela 2.1: Distribuições condicionais identificáveis *vs.* inidentificáveis da fatoração PMM (num padrão de omissão monótono com $J = 3$) indicadas, respectivamente, a esquerda/acima e a direita/abaixo da linha que cruza a tabela.

				$f\left(y_j \,\vert y_1, y_2, \ldots, y_{j-1}, t = s\right)$		
$t = s$	$\mathbf{r}$	$\mathbf{y}^o$	$\mathbf{y}^m$	$j = 1$	$j = 2$	$j = 3$
3	(1,1,1)	(y_1, y_2, y_3)	$\emptyset$	$f\left(y_1 \,\vert t = 3\right)$	$f\left(y_2 \,\vert y_1, t = 3\right)$	$f\left(y_3 \,\vert y_1, y_2, t = 3\right)$
2	(1,1,0)	(y_1, y_2)	y_{i3}	$f\left(y_1 \,\vert t = 2\right)$	$f\left(y_2 \,\vert y_1, t = 2\right)$	$f\left(y_3 \,\vert y_1, y_2, t = 2\right)$
1	(1,0,0)	y_1	(y_{i2}, y_{i3})	$f\left(y_1 \,\vert t = 1\right)$	$f\left(y_2 \,\vert y_1, t = 1\right)$	$f\left(y_3 \,\vert y_1, y_2, t = 1\right)$
0	(0,0,0)	$\emptyset$	(y_{i1}, y_{i2}, y_{i3})	$f\left(y_1 \,\vert t = 0\right)$	$f\left(y_2 \,\vert y_1, t = 0\right)$	$f\left(y_3 \,\vert y_1, y_2, t = 0\right)$

Obs.: por simplicidade, os parâmetros e o índice i foram omitidos das expressões.

Tabela 2.1), por exemplo,

$$
\begin{aligned}
&f\left(y_{i2}, y_{i3} \,\vert y_{i1}, \mathbf{r}_i = (1,0,0)\right) = f\left(y_{i2}, y_{i3} \,\vert y_{i1}, t_i = 1\right) = \\
&\quad f\left(y_{i2} \,\vert y_{i1}, t_i = 1\right) f\left(y_{i3} \,\vert y_{i1}, y_{i2}, t_i = 1\right) \overset{\text{CCMV}}{=} f\left(y_{i2} \,\vert y_{i1}, t_i = 3\right) f\left(y_{i3} \,\vert y_{i1}, y_{i2}, t_i = 3\right) = \\
&\qquad f\left(y_{i2}, y_{i3} \,\vert y_{i1}, t_i = 3\right) = f\left(y_{i2}, y_{i3} \,\vert y_{i1}, \mathbf{r}_i = (1,1,1)\right),
\end{aligned}
$$

em que, por simplicidade, se omitiu os parâmetros das expressões. Para padrões de omissão não-monótonos, esse esquema de identificação é mais facilmente aplicável do que outros esquemas, apresentados a seguir. Além disso, ele é talvez mais razoável nos casos em que a maior parte das unidades amostrais é completamente categorizada e apenas poucas apresentam cada um dos demais padrões de omissão. Contudo, essas unidades com informações completas podem estar "distantes" em algum sentido dos padrões com omissão, especialmente nos casos em que o abandono ocorre nos primeiros instantes. Por esse motivo, Verbeke & Molenberghs (2000) sugerem que pode ser mais adequado tomar a informação emprestada de algum outro padrão de omissão mais próximo, ou mesmo de todos em que há informação sobre a variável omissa; no primeiro caso, as restrições de valores omissos de casos vizinhos (NCMV, *neighboring case missing values*) são expressas por

$$
f\left(y_{ij} \,\vert y_{i1}, y_{i2}, \ldots, y_{i,j-1}, t_i = s, \boldsymbol{\theta}, \boldsymbol{\psi}\right) = f\left(y_{ij} \,\vert y_{i1}, y_{i2}, \ldots, y_{i,j-1}, t_i = j, \boldsymbol{\theta}, \boldsymbol{\psi}\right), \quad j = s{+}1, s{+}2, \ldots, J, \tag{2.22}
$$

enquanto que no último, as restrições de valores omissos de casos disponíveis (ACMV, *available case missing values*) são dadas por

$$
f\left(y_{ij} \,\vert y_{i1}, y_{i2}, \ldots, y_{i,j-1}, t_i = s, \boldsymbol{\theta}, \boldsymbol{\psi}\right) = f\left(y_{ij} \,\vert y_{i1}, y_{i2}, \ldots, y_{i,j-1}, t_i \geq j, \boldsymbol{\theta}, \boldsymbol{\psi}\right), \quad j = s{+}1, s{+}2, \ldots, J. \tag{2.23}
$$

Thijs, Molenberghs, Michiels, Verbeke & Curran (2002) derivaram uma estratégia computacional prática para incorporar as restrições ACMV, notando que uma expressão geral para restrições identificadoras que tomam informações faltantes emprestadas de padrões em que estas tenham sido

observadas é

$$f\left(y_{ij}\left|y_{i1},y_{i2},\ldots,y_{i,j-1},t_i=s,\boldsymbol{\theta},\boldsymbol{\psi}\right.\right)=\sum_{k=j}^{J}\omega_{jk}f\left(y_{ij}\left|y_{i1},y_{i2},\ldots,y_{i,j-1},t_i=k,\boldsymbol{\theta},\boldsymbol{\psi}\right.\right), \tag{2.24}$$

$j=s+1,s+2,\ldots,J$, em que $\sum_{k=j}^{J}\omega_{jk}=1$. Devido à utilização de distribuições observáveis para identificar distribuições não-observáveis, Kenward *et al.* (2003) chamaram esta família de restrições de "Interior"; ela engloba CCMV, quando $\omega_{jJ}=1$ e demais ω_{jk}'s iguais a zero, NCMV, quando $\omega_{jj}=1$ e demais ω_{jk}'s iguais a zero, e abrange ACMV, quando

$$\omega_{jk}=\frac{f\left(t_i=k\left|\boldsymbol{\psi}\right.\right)f\left(y_{i1},y_{i2},\ldots,y_{i,j-1}\left|t_i=k,\boldsymbol{\theta},\boldsymbol{\psi}\right.\right)}{\sum_{l=j}^{J}f\left(t_i=l\left|\boldsymbol{\psi}\right.\right)f\left(y_{i1},y_{i2},\ldots,y_{i,j-1}\left|t_i=l,\boldsymbol{\theta},\boldsymbol{\psi}\right.\right)}.$$

Molenberghs, Michiels, Kenward & Diggle (1998) mostraram que as restrições ACMV são a contrapartida do mecanismo MAR em padrões de omissão monótonos, mas que com padrões de omissão como o da Tabela 1.1 equivaleriam ao mecanismo MCAR. Uma importante implicação deste resultado em padrões de omissão monótonos é que outras restrições pertencentes à família "Interior" que não sejam a ACMV traduzem mecanismos MNAR. Apesar da equivalência, Kenward *et al.* (2003) chamam a atenção de que modelos MAR e ACMV podem não apresentar os mesmos resultados quando há variáveis explicativas, pois elas são incorporadas de diferentes formas nas fatorações SeM e PMM, resultando em diferentes modelos para $\mathbf{Y}_i$.

O resultado (2.11) em conjunto com as expressões (2.8) e (2.3) evidencia que, nos PMM, as suposições do mecanismo MAR traduzem-se em

$$f(\mathbf{y}_i|\mathbf{r}_i,\boldsymbol{\theta},\boldsymbol{\psi})=f\left(\mathbf{y}_i^m\left|\mathbf{y}_i^o,\boldsymbol{\theta}\right.\right)f\left(\mathbf{y}_i^o\left|\mathbf{r}_i,\boldsymbol{\theta},\boldsymbol{\psi}\right.\right). \tag{2.25}$$

Como $\boldsymbol{\psi}$ não é ignorável para se fazer inferências sobre o processo de mensuração/categorização sob as restrições equivalentes à MAR na fatoração PMM, conclui-se que a ignorabilidade do mecanismo MAR é uma característica particular da fatoração SeM. Em compensação, as expressões (2.2), (2.3) e (2.12) indicam que a independência entre $\mathbf{Y}_i$ e $\mathbf{R}_i$ sob o mecanismo MCAR leva a que ele seja igualmente expressável na fatoração PMM, ou seja,

$$f(\mathbf{y}_i|\mathbf{r}_i,\boldsymbol{\theta},\boldsymbol{\psi})=f(\mathbf{y}_i|\boldsymbol{\theta}). \tag{2.26}$$

Kenward *et al.* (2003) mostraram que o mecanismo MNFD (2.19), no caso do padrão de omissão monótono, é equivalente às restrições de valores omissos não-futuros (NFMV, *non-future missing*

values), para a fatoração PMM, expressas por

$$f\left(y_{ij}\left|y_{i1},\ldots,y_{i,j-1},t_i=s,\boldsymbol{\theta},\boldsymbol{\psi}\right.\right)=f\left(y_{ij}\left|y_{i1},\ldots,y_{i,j-1},t_i\geq j-1,\boldsymbol{\theta},\boldsymbol{\psi}\right.\right),\quad j=s+2,s+3,\ldots,J. \tag{2.27}$$

Contudo, MNFD/NFMV não são suficientes para obter modelos identificáveis. Note-se, por exemplo, no caso das restrições NFMV, que cada um dos padrões de omissão continua com uma distribuição inidentificável, nomeadamente

$$f\left(y_{i,s+1}\left|y_{i1},\ldots,y_{i,s},t_i=s,\boldsymbol{\theta},\boldsymbol{\psi}\right.\right). \tag{2.28}$$

Ou seja, a distribuição da resposta faltante "atual" condicionalmente às respostas anteriores permanece irrestrita. Por esta razão, a família NFMV contém membros fora da família "Interior". Por outro lado, a família NFMV exclui membros como CCMV e NCMV. Contudo, a escolha de (2.28) com mesma forma funcional de (2.27) leva à ACMV, mostrando que existe uma interseção entre as duas famílias. Kenward *et al.* (2003) sugerem que (2.28) pode ser facilmente especificada com restrições do tipo CCMV ou NCMV ou pode ser utilizada numa análise de sensibilidade, conforme se descreve na Seção 2.4.

2.3.3 Modelos de parâmetros compartilhados

Recentemente, alguma atenção tem sido dedicada também à fatoração SPM. Por esta razão, há interesse em se estabelecer conexões entre suposições do mecanismo de omissão incluídas nessa fatoração, na SeM e na PMM. Descrevem-se, portanto, os principais resultados apresentados por Creemers *et al.* (2009), que se debruçam sobre o caso (2.7), em que existe independência condicional entre os mecanismos de mensuração/categorização e de omissão dados os efeitos aleatórios. Como, em geral, ocorre em modelos com efeitos aleatórios, admite-se que $\mathbf{b}_i$ segue alguma distribuição indexada por parâmetros incluídos no vetor $\boldsymbol{\xi}$. Uma vez que os efeitos aleatórios não são observados, a distribuição dos dados observados, neste caso, é obtida após integração de $\mathbf{b}_i$, ou seja,

$$f(\mathbf{y}_i^o,\mathbf{r}_i|\boldsymbol{\theta},\boldsymbol{\psi},\boldsymbol{\xi})=\sum_{\mathbf{y}_i^m}\int f(\mathbf{y}_i,\mathbf{r}_i|\boldsymbol{\theta},\boldsymbol{\psi},\mathbf{b}_i)f(\mathbf{b}_i|\boldsymbol{\xi})d\mathbf{b}_i.$$

No caso de $\mathbf{b}_i$ ser discreto, a integral é substituída por um somatório, resultando num modelo de mistura ou de classes latentes.

Por simplicidade, os parâmetros são omitidos na notação empregada no desenvolvimento a seguir. Além disso, convém reescrever o primeiro termo do segundo membro de (2.7) utilizando a partição dos dados completos e indicar quais vetores de efeitos aleatórios pertencem a cada distribuição conforme evidenciado em

$$f(\mathbf{y}_i,\mathbf{r}_i|\mathbf{b}_i)=f(\mathbf{y}_i^o|\mathbf{b}_i^o,\mathbf{b}_i^{om},\mathbf{b}_i^{or},\mathbf{b}_i^{omr})f(\mathbf{y}_i^m|\mathbf{y}_i^o,\mathbf{b}_i^m,\mathbf{b}_i^{om},\mathbf{b}_i^{mr},\mathbf{b}_i^{omr})f(\mathbf{r}_i|\mathbf{b}_i^r,\mathbf{b}_i^{or},\mathbf{b}_i^{mr},\mathbf{b}_i^{omr}), \tag{2.29}$$

em que os índices o, m e r nos subvetores $\mathbf{b}_i$'s indicam que os efeitos aleatórios estão associados, respectivamente, às distribuições de $\mathbf{y}_i^o$, $\mathbf{y}_i^m$ dado $\mathbf{y}_i^o$ e $\mathbf{r}_i$, $\mathbf{b}_i = (\mathbf{b}_i^{o\prime}, \mathbf{b}_i^{m\prime}, \mathbf{b}_i^{r\prime}, \mathbf{b}_i^{om\prime}, \mathbf{b}_i^{or\prime}, \mathbf{b}_i^{mr\prime}, \mathbf{b}_i^{omr\prime})'$ e os 7 subvetores de $\mathbf{b}_i$ são mutuamente independentes. Utilizando (2.11), Creemers *et al.* (2009) mostram que um membro da família (2.29) é MAR se, e somente se,

$$\frac{\int f(\mathbf{y}_i^o|\mathbf{b}_i^o, \mathbf{b}_i^{om}, \mathbf{b}_i^{or}, \mathbf{b}_i^{omr}) f(\mathbf{y}_i^m|\mathbf{y}_i^o, \mathbf{b}_i^m, \mathbf{b}_i^{om}, \mathbf{b}_i^{mr}, \mathbf{b}_i^{omr}) f(\mathbf{r}_i|\mathbf{b}_i^r, \mathbf{b}_i^{or}, \mathbf{b}_i^{mr}, \mathbf{b}_i^{omr}) f(\mathbf{b}_i) d\mathbf{b}_i}{\int f(\mathbf{y}_i^o|\mathbf{b}_i^o, \mathbf{b}_i^{om}, \mathbf{b}_i^{or}, \mathbf{b}_i^{omr}) f(\mathbf{r}_i|\mathbf{b}_i^r, \mathbf{b}_i^{or}, \mathbf{b}_i^{mr}, \mathbf{b}_i^{omr}) f(\mathbf{b}_i) d\mathbf{b}_i}$$
$$= \frac{\int f(\mathbf{y}_i^o|\mathbf{b}_i^o, \mathbf{b}_i^{om}, \mathbf{b}_i^{or}, \mathbf{b}_i^{omr}) f(\mathbf{y}_i^m|\mathbf{y}_i^o, \mathbf{b}_i^m, \mathbf{b}_i^{om}, \mathbf{b}_i^{mr}, \mathbf{b}_i^{omr}) f(\mathbf{b}_i) d\mathbf{b}_i}{\int f(\mathbf{y}_i^o|\mathbf{b}_i^o, \mathbf{b}_i^{om}, \mathbf{b}_i^{or}, \mathbf{b}_i^{omr}) f(\mathbf{b}_i) d\mathbf{b}_i}. \quad (2.30)$$

Intuitivamente, um caso especial que satisfaz (2.30) é obtido ao se omitir os efeitos aleatórios $\mathbf{b}_i^{om}$, $\mathbf{b}_i^{mr}$ e $\mathbf{b}_i^{omr}$, responsáveis pela dependência entre os dados observados e faltantes; especificamente,

$$f(\mathbf{y}_i, \mathbf{r}_i|\mathbf{b}_i) = f(\mathbf{y}_i^o|\mathbf{b}_i^o, \mathbf{b}_i^{or}) f(\mathbf{y}_i^m|\mathbf{y}_i^o, \mathbf{b}_i^m) f(\mathbf{r}_i|\mathbf{b}_i^r, \mathbf{b}_i^{or}). \quad (2.31)$$

Creemers *et al.* (2009) exemplificaram um modelo não incluso em (2.31) que satisfaz (2.30). Além disso, notando que (2.19) pode ser encarada como uma definição de um mecanismo MAR (em estudos longitudinais com padrões de omissão monótonos) com a sequência (em j) de y_{ij}'s deslocada, esses autores estabeleceram que um membro da família (2.29) é MNFD se, e somente se,

$$\frac{\int f(\mathbf{y}_i^p|\mathbf{b}_i^p, \mathbf{b}_i^{pf}, \mathbf{b}_i^{pt}, \mathbf{b}_i^{pft}) f(\mathbf{y}_i^f|\mathbf{y}_i^p, \mathbf{b}_i^f, \mathbf{b}_i^{pf}, \mathbf{b}_i^{ft}, \mathbf{b}_i^{pft}) f(t_i = s|\mathbf{b}_i^t, \mathbf{b}_i^{pt}, \mathbf{b}_i^{ft}, \mathbf{b}_i^{pft}) f(\mathbf{b}_i) d\mathbf{b}_i}{\int f(\mathbf{y}_i^p|\mathbf{b}_i^p, \mathbf{b}_i^{pf}, \mathbf{b}_i^{pt}, \mathbf{b}_i^{pft}) f(t_i = s|\mathbf{b}_i^t, \mathbf{b}_i^{pt}, \mathbf{b}_i^{ft}, \mathbf{b}_i^{pft}) f(\mathbf{b}_i) d\mathbf{b}_i}$$
$$= \frac{\int f(\mathbf{y}_i^p|\mathbf{b}_i^p, \mathbf{b}_i^{pf}, \mathbf{b}_i^{pt}, \mathbf{b}_i^{pft}) f(\mathbf{y}_i^f|\mathbf{y}_i^p, \mathbf{b}_i^f, \mathbf{b}_i^{pf}, \mathbf{b}_i^{ft}, \mathbf{b}_i^{pft}) f(\mathbf{b}_i) d\mathbf{b}_i}{\int f(\mathbf{y}_i^p|\mathbf{b}_i^p, \mathbf{b}_i^{pf}, \mathbf{b}_i^{pt}, \mathbf{b}_i^{pft}) f(\mathbf{b}_i) d\mathbf{b}_i}, \quad (2.32)$$

em que o índice r foi substituído por t para representar os efeitos aleatórios associados à distribuição de t_i, $\mathbf{y}_i^p = (y_{i1}, y_{i2}, \ldots, y_{i,s+1})'$ e $\mathbf{y}_i^f = (y_{i,s+2}, y_{i,s+3}, \ldots, y_{iJ})'$ denotam, respectivamente, as variáveis do "passado e presente" e do "futuro" e os índices o e m atribuídos aos efeitos aleatórios foram substituídos por p e f seguindo a mesma lógica. Logo, todos os efeitos aleatórios com índices o e m foram substituídos por p e f, respectivamente, com exceção dos efeitos aleatórios associados à primeira variável faltante na sequência (o "presente"), $y_{i,s+1}$, que estavam agrupados nos $\mathbf{b}_i$'s com índice m e agora estão incluídos nos $\mathbf{b}_i$'s com índice p. Realizando o mesmo tipo de deslocamento em (2.31), obtém-se a subclasse

$$f(\mathbf{y}_i, t_i|\mathbf{b}_i) = f(\mathbf{y}_i^p|\mathbf{b}_i^p, \mathbf{b}_i^{pt}) f(\mathbf{y}_i^f|\mathbf{y}_i^p, \mathbf{b}_i^f) f(t_i|\mathbf{b}_i^t, \mathbf{b}_i^{pt}), \quad (2.33)$$

que satisfaz (2.32) e possui uma forma mais intuitiva. Por fim, o mecanismo MCAR é obviamente obtido ao se omitirem todos os efeitos aleatórios que seriam compartilhados por duas ou mais distri-

buições, *i.e.*, $\mathbf{b}_i^{om}$, $\mathbf{b}_i^{or}$, $\mathbf{b}_i^{mr}$ e $\mathbf{b}_i^{omr}$, resultando em

$$f(\mathbf{y}_i, \mathbf{r}_i|\mathbf{b}_i) = f(\mathbf{y}_i^o|\mathbf{b}_i^o) f(\mathbf{y}_i^m|\mathbf{y}_i^o, \mathbf{b}_i^m) f(\mathbf{r}_i|\mathbf{b}_i^r). \tag{2.34}$$

2.3.4 Resumo e outras abordagens

Analogamente ao que foi apresentado por Creemers *et al.* (2009), ilustra-se na Figura 2.1 a miscelânea de definições de mecanismos de omissão e acrônimos descritos nesta seção, bem como as relações entre famílias de modelos.

$$\begin{array}{ccccccccc}
 & & & & \text{NCMV (2.22)}^* & & \text{CCMV (2.21)}^*/(2.20) & & \\
 & & & & & \subset & \cap & & \\
 & & & & \text{ACMV (2.23)}^* & \subset & \text{Interior (2.24)}^* & & \\
 & & & & \updownarrow^* & & & \subset & \\
\text{PMM (2.3)} & : & \text{MCAR (2.26)} & \subset & \text{MAR (2.25)} & \subset & \text{NFMV (2.27)}^* & \subset & \text{MNAR geral} \\
 & & \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow \\
\text{SeM (2.2)} & : & \text{MCAR (2.12)} & \subset & \text{MAR (2.10)} & \subset & \text{MNFD (2.19)}^* & \subset & \text{MNAR geral} \\
 & & \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow \\
\text{SPM (2.29)} & : & \text{MCAR (2.34)} & \subset & \text{MAR (2.30)} & \subset & \text{MNFD (2.32)}^* & \subset & \text{MNAR geral} \\
 & & & & \cup & & \cup & & \\
 & & & & (2.31) & \subset & (2.33)^* & &
\end{array}$$

Figura 2.1: Relações entre famílias de modelos nas fatorações de modelos de seleção (SeM), modelos de mistura de padrões (PMM) e modelos de parâmetros compartilhados (SPM); MCAR: *missing completely at random*, MAR: *missing at random*, MNAR: *missing not at random*, ACMV: *available case missing values*, NCMV: *neighboring case missing values*, CCMV: *complete case missing values*, NFMV: *non-future missing values*, MNFD: *missing non-future dependence*; $\subset$ e suas versões rotacionadas indicam quais modelos estão aninhados entre si em cada fatoração, $\updownarrow$ denota a equivalência entre os modelos de diferentes fatorações e * enfatiza a utilização do padrão de omissão monótono.

O investimento nas definições de mecanismos de omissão para as fatorações PMM e SPM equivalentes às da fatoração SeM se deve ao fato de que a terminologia de Rubin (1976) é largamente empregada na literatura. Contudo, cabe ressaltar que essa preferência não é unânime. Por exemplo, Lindsey (1999, pp.363-372) defende que as distinções feitas por Rubin são inadequadas para estudos com medidas repetidas e Vansteelandt & Robins (2007) argumentam que a suposição MAR raramente é razoável para padrões de omissão não-monótonos em estudos longitudinais. Além disso, Mouchart (2007) desenvolve um arcabouço para o estudo de dados faltantes e defende que as definições de ignorabilidade, conforme apresentadas por Rubin, requerem cuidado particular que depende da especificidade da aplicação.

2.4 Análises de sensibilidade

Apesar das simplificações convenientes obtidas sob um mecanismo MAR, os estudos não costumam fornecer justificativas para a sua completa adoção, uma vez que os motivos para as omissões são variados e, em geral, difíceis ou impossíveis de serem especificados. Por essa razão, há interesse de se considerar afastamentos em relação à suposição MAR. Esta tarefa torna-se obtusa devido à falta de informação amostral característica dos dados com omissão, que não permite, por exemplo, que se distinga um mecanismo MAR saturado de outro MNAR saturado por meio de critérios estatísticos, pois ambos podem ser perfeitamente ajustados aos dados observados. Além disso, as diferentes suposições que fundamentam esses mecanismos de omissão podem trazer resultados discrepantes para o processo de mensuração/categorização. Esses comentários ficam patentes no trabalho de Molenberghs, Beunckens, Sotto & Kenward (2008), que estabelecem uma estratégia de como obter ajustes de modelos MAR a partir de ajustes de modelos MNAR, em que ambos terão o mesmo grau de ajustamento aos dados observados. Por esses motivos, as análises de sensibilidade são úteis para contornar os problemas associados à falta de informação.

A análise de sensibilidade clássica mais comum, aqui apelidada de informal, costuma ser a de propor diferentes estruturas identificáveis plausíveis (embora subjetivas), MAR e MNAR, para o mecanismo de omissão e, em seguida, verificar a estabilidade das inferências de interesse e a razoabilidade dos valores esperados estimados para os dados totais. Kenward, Goetghebeur & Molenberghs (2001) mostram que a análise de sensibilidade informal pode levar a conclusões equivocadas. Tanto nesse trabalho quanto em Molenberghs, Kenward & Goetghebeur (2001) e Vansteelandt, Goetghebeur, Kenward & Molenberghs (2006), os autores sugerem uma alternativa, aqui chamada de formal, que possibilita a utilização de modelos sobreparametrizados por meio da realização de repetidas análises de sensibilidade.

Por outro lado, no paradigma bayesiano, a eliciação dos hiperparâmetros e a tradicional análise de sensibilidade das distribuições *a priori* merecem uma maior cautela para mecanismos de omissão que não tenham qualquer restrição imposta, devido às funções paramétricas que continuam, *a posteriori*, dependentes apenas da distribuição proposta *a priori*. Contudo, a informação *a priori* não atualizada é ainda menos restritiva do que as alternativas que admitem estruturas identificáveis para o mecanismo de omissão, uma vez que a primeira desempenha um papel estocástico, enquanto que as últimas têm uma natureza determinística.

Kenward *et al.* (2001) e Molenberghs *et al.* (2001) também ressaltaram que o intervalo para o melhor-pior caso (*best-worst case interval*), segundo o qual se alocam as unidades omissas a categorias que produzam casos extremos para as inferências de interesse, apesar de ter grande amplitude, é um método relativamente simples, bastante informativo e um ponto de partida honesto para uma modelagem cautelosa. No entanto, a análise de sensibilidade formal clássica e os métodos bayesianos que não impõem restrições no mecanismo de omissão devem gerar intervalos com menores amplitudes.

No Capítulo 3, investigam-se e comparam-se as análises de sensibilidade bayesiana e formal clássica em modelos sobreparametrizados. Poleto (2006) ilustra a construção do intervalo para o melhor-pior caso e a realização de análises de sensibilidade clássicas informal e formal com diversos exemplos.

Van Steen, Molenberghs, Verbeke & Thijs (2001) e Jansen, Molenberghs, Aerts, Thijs & Van Steen (2003) seguem outra vertente de análise de sensibilidade, utilizando o método de diagnóstico de influência local (Cook, 1986) incluindo perturbações na verossimilhança de um mecanismo MAR na direção de mecanismos MNAR em dados ordinais e binários, respectivamente. Pequenas perturbações em alguma resposta que provoquem variações desproporcionais nos resultados sugerem que ela provavelmente contribui de uma maneira particular para as conclusões. Contudo, pesquisas em torno dessa ferramenta continuam a ser realizadas em análises de dados com omissão com respostas contínuas e categorizadas (para uma revisão, veja Molenberghs & Kenward, 2007, Cap. 22 e 23). As respostas que se destacam por dirigir as conclusões para um mecanismo MNAR podem ser responsáveis não apenas devido ao verdadeiro mecanismo de omissão, mas também por causas associadas ao processo de mensuração/categorização. Similarmente, é possível que unidades experimentais que se desviem da massa de dados por serem geradas por um mecanismo MNAR não sejam detectadas por meio desse método. Por isso, é necessário tomar cuidado para se avaliar a maneira exata com que as unidades experimentais ressaltadas nessa abordagem afetam as conclusões.

Todas as análises de sensibilidade discutidas até aqui utilizam as fatorações SeM e PMM. Creemers *et al.* (2010) propõem uma análise de sensibilidade sob a fatoração SPM.

Capítulo 3

Modelos sobreparametrizados para a análise de dados categorizados com respostas omissas e suas implicações inferenciais

Modelos que levam em consideração o processo gerador dos dados incompletos são, em sua forma mais geral, sobreparametrizados e inidentificáveis. Paulino & Pereira (1994) apresentam uma excelente revisão de inidentificabilidade em estatística e discutem suas consequências em análises clássicas e bayesianas. Neath & Samaniego (1997) e Gustafson (2005) também apresentam discussões interessantes sob o ponto de vista bayesiano, enquanto Daniels & Hogan (2007) especializam-se em dados longitudinais com omissão. Para levar a cabo inferências clássicas, a estratégia mais comum para superar a inidentificabilidade é considerar restrições identificadoras que permitam que o modelo reflita algum mecanismo gerador dos dados faltantes. Em virtude de as suposições subjacentes serem em geral inverificáveis, estatísticos usualmente realizam uma análise de sensibilidade "informal" baseada em um conjunto de modelos identificáveis plausíveis (mas subjetivos), mesmo que tais modelos ainda não sejam capazes de refletir algum mecanismo de omissão mais complexo de interesse. Uma alternativa mais formal, que se chama aqui de **análise de sensibilidade clássica**, envolve a realização de análises de sensibilidades repetidas com diferentes modelos sobreparametrizados para os quais valores especificados são fixados para os parâmetros inidentificáveis (veja *e.g.*, Nordheim, 1984; Copas & Li, 1997; Scharfstein *et al.*, 1999; Vansteelandt *et al.*, 2006). Esses caminhos podem ser desnecessários sob a ótica bayesiana porque o uso de distribuições *a priori* próprias desbloqueia o processo inferencial (Paulino & Pereira, 1995). Entretanto, uma vez que os dados não contêm informação para atualizar a distribuição *a priori* para todos os parâmetros, deve-se ter cuidado adicional na sua eliciação. De fato, a inidentificabilidade pode estar escondida sob as distribuições *a priori* propostas e isso pode criar um falso senso de precisão se não houver cautela em suas escolhas.

Molenberghs *et al.* (2001) afirmam que uma maior sobreparametrização produz uma maior incerteza, enquanto que modelos muito parcimoniosos podem errar o modelo verdadeiro. Na prática, entretanto, estes e outros autores (*e.g.*, Daniels & Hogan, 2007, Cap. 10) usualmente propõem

reduções na dimensão dos modelos adotados e consideram apenas alguns poucos parâmetros em suas análises de sensibilidade. Mantendo isso em mente, o primeiro objetivo deste capítulo é ilustrar que tais reduções de dimensão precisam ser levadas a cabo com grande prudência para evitar conclusões enganosas. Como a utilização de modelos sobreparametrizados reduzidos é uma prática muito comum nas análises de sensibilidade clássicas, escrutina-se essa questão apenas sob esta abordagem, embora consequências similares também sejam esperadas sob o ponto de vista bayesiano.

Ao considerar análises bayesianas de dados categorizados incompletos, Paulino & Pereira (1992, 1995) e Forster & Smith (1998) claramente indicam que resumos *a posteriori* dos parâmetros de interesse são dependentes da distribuição *a priori*. Soares & Paulino (2001) também observam isto, mas erroneamente acreditam que as implicações desta dependência devem ser moderadas para grandes amostras, dado que os dados atualizam algumas funções dos parâmetros de interesse. De fato, Neath & Samaniego (1997) concluem:

> *Bayesian analysis cannot be used with impunity in estimating non-identifiable parameters. ... Because posterior estimates of non-identifiable parameters are strongly influenced by prior modelling, even as the sample size grows without bound, it is important to use the utmost care in applying and interpreting Bayesian analysis in such settings.*

Recentemente, Tian *et al.* (2003) e Jiang & Dickey (2008) consideraram níveis maiores de sobreparametrização em dados categorizados (possibilitando erros de classificação, além de omissão) sem esclarecer que os parâmetros de interesse são inidentificáveis e sem contabilizar adequadamente a dependência das distribuições *a priori* correspondentes. Assim, o segundo objetivo deste capítulo é mostrar que uma dependência considerável da distribuição *a priori* permanece para parâmetros inidentificáveis mesmo em análises com tamanhos de amostra imensos e com distribuições *a priori* consideradas não-informativas ou levemente informativas. Concretiza-se esse objetivo analisando exemplos simples com respostas categorizadas com omissão, embora a essência das conclusões seja extensível tanto para respostas contínuas com omissão quanto para outras áreas da estatística em que a inidentificabilidade é uma preocupação, como aquelas de erros de classificação e de mensuração.

Revisam-se as estratégias bayesiana e clássica para a análise de dados categorizados incompletos na Seção 3.1. Exploram-se e comparam-se as duas abordagens usando dados simulados na Seção 3.2 e reanalisando na Seção 3.3 os dados do *Collaborative Perinatal Project* exibidos na Tabela 1.1. No primeiro caso, examinam-se os efeitos de um padrão de omissão monótono sobre parâmetros identificáveis e inidentificáveis com dados de amostras com tamanhos variando de pequeno a grande (*e.g.*, 40 a 4 000 000). No segundo caso, mostra-se que as conclusões das análises podem mudar dependendo das escolhas dos componentes subjetivos das abordagens bayesiana e clássica. Na Seção 3.4, conclui-se o capítulo com uma breve discussão.

3.1 Abordagens inferenciais

Paulino & Pereira (1992) desenvolveram uma solução bayesiana baseada em modelos de omissão irrestritos para análise de dados categorizados incompletos quando os padrões de omissão podem ser estruturados em partições do conjunto de categorias. Exemplos desses padrões são aqueles gerados por classificações incompletas em subtabelas marginais como na Tabela 1.1. Tal requisito foi relaxado por Paulino & Pereira (1995), que consideraram padrões de censura mais gerais. A abordagem desses autores serve de base para a Seção 3.1.1, onde se esboça o problema, se define a notação e se especifica a verossimilhança; além disso, as distribuições *a priori* de Paulino & Pereira (1995) também servem como ponto de partida para a identificação de distribuições *a priori* e *a posteriori* discutidas na Seção 3.1.2. Na Seção 3.1.3, revisa-se a análise de sensibilidade clássica de Vansteelandt *et al.* (2006). Finalmente, na Seção 3.1.4, discutem-se as diferenças entre as duas abordagens e algumas possíveis extensões.

3.1.1 Descrição do problema, notação e verossimilhança

Considere-se uma amostra aleatória de tamanho n, em que cada uma das unidades é classificada em uma categoria de resposta r com probabilidade θ_r, $r = 1, \ldots, R$, e R corresponde ao número de combinações dos níveis das variáveis respostas. Por várias razões atribuíveis a mecanismos de censura ou omissão, pode ser possível observar apenas as frequências de unidades em subconjuntos não-vazios $\mathcal{C}$ de $\{1, \ldots, R\}$, que se denominam classes de respostas. Em particular, a resposta para uma unidade é completamente categorizada (na categoria r) ou completamente omissa se $\mathcal{C} = \{r\}$ ou $\mathcal{C} = \{1, \ldots, R\}$, respectivamente. Supõe-se que unidades com resposta na categoria r são observadas na classe $\mathcal{C}$ com probabilidade $\lambda_{\mathcal{C}(r)}$. Supõe-se também que não há erro de classificação, *i.e.*, $\lambda_{\mathcal{C}(r)} = 0$ sempre que $r \notin \mathcal{C}$. Tomando $\mathcal{P}_o$ como a união das classes de respostas que não contêm omissão e $\mathcal{P}_m$ como a união das que possuem algum grau de omissão, segue que $\mathcal{P} = \mathcal{P}_o \cup \mathcal{P}_m$ abrange todos os padrões de resposta possíveis. Da mesma forma, os dados podem ser resumidos no vetor $\mathbf{N} = (\mathbf{N}_o', \mathbf{N}_m')'$, em que $\mathbf{N}_o = (n_{\mathcal{C}}, \mathcal{C} \in \mathcal{P}_o)' = (n_r, r = 1, \ldots, R)'$ empilha as frequências das observações completamente categorizadas e $\mathbf{N}_m = (n_{\mathcal{C}}, \mathcal{C} \in \mathcal{P}_m)'$ inclui as frequências das observações com omissão parcial ou completa. Sejam $\boldsymbol{\theta} = (\theta_r, r = 1, \ldots, R)'$, o vetor de probabilidades marginais de categorização, o parâmetro de interesse e $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_r', r = 1, \ldots, R)'$, o vetor de probabilidades condicionais de omissão, em que $\boldsymbol{\lambda}_r = (\lambda_{\mathcal{C}(r)}, \mathcal{C} \in \mathcal{P}_r)'$ e $\mathcal{P}_r = \{\mathcal{C} \in \mathcal{P} : r \in \mathcal{C}\}$ contém as classes de respostas que incluem a categoria r. Note-se que as restrições naturais são $\sum_{r=1}^{R} \theta_r = 1$ e $\sum_{\mathcal{C} \in \mathcal{P}_r} \lambda_{\mathcal{C}(r)} = 1$, $r = 1, \ldots, R$. Consequentemente, segundo o modelo multinomial, a função de verossimilhança para $(\boldsymbol{\theta}', \boldsymbol{\lambda}')'$ é

$$L(\boldsymbol{\theta}, \boldsymbol{\lambda} | \mathbf{N}) \propto \prod_{\mathcal{C} \in \mathcal{P}} \left(\sum_{r \in \mathcal{C}} \theta_r \lambda_{\mathcal{C}(r)} \right)^{n_{\mathcal{C}}} = \prod_{r=1}^{R} \left(\theta_r \lambda_{r(r)} \right)^{n_r} \times \prod_{\mathcal{C} \in \mathcal{P}_m} \left(\sum_{r \in \mathcal{C}} \theta_r \lambda_{\mathcal{C}(r)} \right)^{n_{\mathcal{C}}}.$$

Esta parametrização é conhecida como modelo de seleção (Glynn *et al.*, 1986; Little & Rubin, 2002). Alternativamente, podem-se considerar as probabilidades conjuntas $\boldsymbol{\mu} = \{\mu_{\mathcal{C}r}\}$, em que $\mu_{\mathcal{C}r} = \theta_r \lambda_{\mathcal{C}(r)}$,

ou ainda uma versão da parametrização do modelo de mistura de padrões (Paulino & Pereira, 1995) que é conveniente para identificar funções paramétricas para as quais a amostra não contém informação. Sejam $\gamma_o = \sum_{r=1}^{R} \mu_{rr}$ a probabilidade de uma unidade ser completamente observada e $\gamma_{\mathcal{C}} = \sum_{r \in \mathcal{C}} \mu_{\mathcal{C}r}$ a probabilidade de observar a classe de resposta $\mathcal{C}, \mathcal{C} \in \mathcal{P}_m$. Do mesmo modo, sejam $\boldsymbol{\alpha}_o = (\alpha_{r(o)}, r = 1, \ldots, R)'$, com $\alpha_{r(o)} = \mu_{rr}/\gamma_o$, as probabilidades condicionais das categorias de resposta $r = 1, \ldots, R$ dada uma observação completa, e $\boldsymbol{\alpha}_{\mathcal{C}} = (\alpha_{r(\mathcal{C})}, r \in \mathcal{C})'$, com $\alpha_{r(\mathcal{C})} = \mu_{\mathcal{C}r}/\gamma_{\mathcal{C}}$, denotam as probabilidades condicionais de pertencer a cada uma das categorias incluídas na classe de resposta $\mathcal{C}$ dada essa categorização parcial, $\mathcal{C} \in \mathcal{P}_m$. Então, a função de verossimilhança de $(\boldsymbol{\gamma}, \boldsymbol{\alpha})$, em que $\boldsymbol{\gamma} = (\gamma_o, \gamma_{\mathcal{C}}, \mathcal{C} \in \mathcal{P}_m)'$ e $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_o', \boldsymbol{\alpha}_{\mathcal{C}}', \mathcal{C} \in \mathcal{P}_m)'$, é expressa como

$$L(\boldsymbol{\gamma}, \boldsymbol{\alpha}|\mathbf{N}) \propto \gamma_o^{n_o} \prod_{\mathcal{C} \in \mathcal{P}_m} \gamma_{\mathcal{C}}^{n_{\mathcal{C}}} \times \prod_{r=1}^{R} \alpha_{r(o)}^{n_r} \times \prod_{\mathcal{C} \in \mathcal{P}_m} \left(\sum_{r \in \mathcal{C}} \alpha_{r(\mathcal{C})} \right)^{n_{\mathcal{C}}}, \tag{3.1}$$

em que $n_o = \sum_{r=1}^{R} n_r$. Uma vez que as restrições naturais são $\gamma_o + \sum_{\mathcal{C} \in \mathcal{P}_m} \gamma_{\mathcal{C}} = 1$, $\sum_{r=1}^{R} \alpha_{r(o)} = 1$, e $\sum_{r \in \mathcal{C}} \alpha_{r(\mathcal{C})} = 1$, $\mathcal{C} \in \mathcal{P}_m$, fica claro que os parâmetros $\{\alpha_{r(\mathcal{C})}, \mathcal{C} \in \mathcal{P}_m\}$ na verdade não aparecem em (3.1) e, por esta razão, não há informação amostral para eles. Na notação das probabilidade condicionais ($\lambda_{\mathcal{C}(r)}$, $\alpha_{r(o)}$ e $\alpha_{r(\mathcal{C})}$), indica-se explicitamente dentro de parênteses os índices dos termos sobre os quais se está condicionando.

Para ilustrar os conceitos e a notação introduzidos até o momento, volta-se para os dados apresentados na Tabela 1.1, em que se substitui o índice r por dois índices, nomeadamente, $i = 1, 2$, para indicar as mães fumantes e não-fumantes, respectivamente, e $j = 1, 2$, para indicar o peso do recém-nascido < 2.5 e ≥ 2.5 kg, respectivamente. Então, as classes de respostas $\mathcal{C}$ associadas ao dados da Tabela 1.1 são dispostos na Tabela 3.1 e, consequentemente, $\mathbf{N}_o = (4\,512, 21\,009, 3\,394, 24\,132)'$ e $\mathbf{N}_m = (1\,049, 1\,135, 142, 464, 1\,224)'$ são as frequências correspondentes às classes de respostas em $\mathcal{P}_o = \big\{\{11\}, \{12\}, \{21\}, \{22\}\big\}$ e $\mathcal{P}_m = \big\{\{11, 12\}, \{21, 22\}, \{11, 21\}, \{12, 22\}, \{11, 12, 21, 22\}\big\}$, respectivamente. Tomando, por exemplo, $ij = 11$, segue que θ_{11} denota a probabilidade marginal de categorização na categoria de resposta 11, $\boldsymbol{\lambda}_{11} = (\lambda_{\{11\}(11)}, \lambda_{\{11,12\}(11)}, \lambda_{\{11,21\}(11)}, \lambda_{\{11,12,21,22\}(11)})'$ é o vetor de probabilidades condicionais de se observar as classes de respostas $\mathcal{C} = \{11\}, \{11, 12\}, \{11, 21\}$ e $\{11, 12, 21, 22\}$ dado que a unidade pertence à categoria de resposta 11, $\mu_{\{11,12\}11}$ denota a probabilidade conjunta de se observar uma unidade na categoria de resposta 11 e classificada na classe de resposta $\{11, 12\}$, $\alpha_{11(o)}$ denota a probabilidade condicional de se observar uma unidade na catego-

Tabela 3.1: Classes de respostas associadas aos dados da Tabela 1.1.

Mãe	Peso do recém-nascido (kg)		
fumante	< 2.5	≥ 2.5	omisso
sim	$\{11\}$	$\{12\}$	$\{11, 12\}$
não	$\{21\}$	$\{22\}$	$\{21, 22\}$
omisso	$\{11, 21\}$	$\{12, 22\}$	$\{11, 12, 21, 22\}$

ria de resposta 11 dada uma observação completa e $\alpha_{11(\{11,12\})}$ denota a probabilidade condicional de selecionar uma unidade da categoria de resposta 11 dado que se observou a classe de resposta $\{11, 12\}$.

3.1.2 Abordagem bayesiana

O uso de distribuições *a priori* próprias no paradigma bayesiano prontamente permite que se obtenham inferências válidas até mesmo para modelos inidentificáveis. A questão aqui recai na escolha de tais distribuições *a priori*. Discute-se algumas alternativas baseadas no trabalho de Paulino & Pereira (1995). Estes autores enfatizaram que admitir uma distribuição *a priori* (DPri) de Dirichlet com hiperparâmetros $\{a_{\mathcal{C}r}\}$ para $\boldsymbol{\mu}$, *i.e.*,

$$\text{DPri 1} : \boldsymbol{\mu} \sim D(\{a_{\mathcal{C}r}\}),$$

é equivalente a supor-se

$$\text{DPri 1} : \begin{cases} \boldsymbol{\theta} \sim D(a_{+r}, r = 1, \ldots, R), \text{ em que } a_{+r} = \sum_{\mathcal{C} \in \mathcal{P}_r} a_{\mathcal{C}r}, \\ \boldsymbol{\lambda}_r \sim D(a_{\mathcal{C}r}, \mathcal{C} \in \mathcal{P}_r), \ r = 1, \ldots, R, \end{cases}$$

com todas as variáveis aleatórias mutuamente independentes, e também a admitir-se

$$\text{DPri 1} : \begin{cases} \boldsymbol{\gamma} \sim D(a_o, a_{\mathcal{C}+}, \mathcal{C} \in \mathcal{P}_m), \text{ em que } a_o = \sum_{r=1}^{R} a_{rr} \text{ e } a_{\mathcal{C}+} = \sum_{r \in \mathcal{C}} a_{\mathcal{C}r}, \\ \boldsymbol{\alpha}_o \sim D(a_{rr}, r = 1, \ldots, R), \\ \boldsymbol{\alpha}_{\mathcal{C}} \sim D(a_{\mathcal{C}r}, r \in \mathcal{C}), \ \mathcal{C} \in \mathcal{P}_m, \end{cases}$$

com todas as variáveis aleatórias mutuamente independentes. A distribuição *a posteriori* de $\boldsymbol{\mu}$, e também a de $\boldsymbol{\theta}$, podem ser expressas como distribuições de Dirichlet generalizadas (*i.e.*, uma mistura finita de distribuições Dirichlet), definidas em Dickey (1983) e sem forma fechada simples; Tian *et al.* (2003) seguem este caminho. Por outro lado, a parametrização do modelo de mistura de padrões leva a $\boldsymbol{\gamma}|\mathbf{N} \sim D(a_o + n_o, a_{\mathcal{C}+} + n_{\mathcal{C}}, \mathcal{C} \in \mathcal{P}_m)$, $\boldsymbol{\alpha}_o|\mathbf{N} \sim D(a_{rr} + n_r, r = 1, \ldots, R)$, $\boldsymbol{\alpha}_{\mathcal{C}}|\mathbf{N} \sim D(a_{\mathcal{C}r}, r \in \mathcal{C})$, $\mathcal{C} \in \mathcal{P}_m$, com todas as variáveis aleatórias mutuamente independentes. A partir daí, Paulino & Pereira (1992, 1995) usam a relação $\theta_r = \gamma_o \, \alpha_{r(o)} + \sum_{\mathcal{C} \in \mathcal{P}_r \cap \mathcal{P}_m} \gamma_{\mathcal{C}} \, \alpha_{r(\mathcal{C})}$ para obter a média e a matriz de covariância de $\boldsymbol{\theta}|\mathbf{N}$. Usando também $\lambda_{r(r)} = \gamma_o \alpha_{r(o)}/\theta_r$ e $\lambda_{\mathcal{C}(r)} = \gamma_{\mathcal{C}} \alpha_{r(\mathcal{C})}/\theta_r$, $\mathcal{C} \in \mathcal{P}_r$, Soares & Paulino (2001) empregam métodos de simulação de Monte Carlo que facilmente permitem que se realizem inferências sobre funções de $(\boldsymbol{\theta}, \boldsymbol{\lambda})$. Mesmo sabendo que a distribuição *a priori* de $\{\boldsymbol{\alpha}_{\mathcal{C}}, \mathcal{C} \in \mathcal{P}_m\}$ não é atualizada pelos dados, eles (equivocadamente) acreditam que as possíveis implicações de dependência da distribuição *a priori* na distribuição *a posteriori* de $(\boldsymbol{\theta}, \boldsymbol{\lambda})$ deveriam ser brandas para grandes amostras, uma vez que $\boldsymbol{\gamma}$ e $\boldsymbol{\alpha}_o$ são atualizados.

O especialista não é livre para fazer julgamentos individuais acerca dos parâmetros que possuem um maior significado ($\boldsymbol{\theta}$ e $\boldsymbol{\lambda}$) com a DPri 1, porque uma vez que os hiperparâmetros $\{a_{\mathcal{C}r}\}$ tenham

sido escolhidos para a distribuição *a priori* de $\boldsymbol{\lambda}$, os hiperparâmetros da distribuição *a priori* de $\boldsymbol{\theta}$ são automaticamente definidos como $\{a_{+r} = \sum_{\mathcal{C}\in\mathcal{P}_r} a_{\mathcal{C}r}\}$. Como essa conexão entre os hiperparâmetros não é geralmente justificável na prática, Soares & Paulino (2007) e Jiang & Dickey (2008) adotam a

$$\text{DPri 2}: \begin{cases} \boldsymbol{\theta} \sim D(b_r, r = 1, \ldots, R), \\ \boldsymbol{\lambda}_r \sim D(a_{\mathcal{C}r}, \mathcal{C} \in \mathcal{P}_r),\ r = 1, \ldots, R, \end{cases}$$

com todas as variáveis aleatórias mutuamente independentes; tanto esses autores quanto Tian *et al.* (2003) também consideram erros de classificação na sua estratégia de modelagem. Jiang & Dickey (2008) trabalham com a resultante distribuição *a posteriori* de Dirichlet generalizada, enquanto Soares & Paulino (2007) usam um algoritmo de ampliação de dados em cadeia (Tanner & Wong, 1987) para obter uma amostra da distribuição *a posteriori*, em que os dados ampliados são as frequências hipotéticas $\mathbf{m} = \{m_{\mathcal{C}r}\}$ de unidades com a categoria de resposta r e classificadas na classe de resposta $\mathcal{C}$. Com exceção do caso em que não há dados com omissão ($m_{rr} = n_r$), essas frequências são não-observáveis, sabendo-se apenas que $\sum_{r\in\mathcal{C}} m_{\mathcal{C}r} = n_{\mathcal{C}}$, $\mathcal{C} \in \mathcal{P}_m$. Dado $(\boldsymbol{\theta}, \boldsymbol{\lambda})$ e $\mathbf{N}$, as frequências $\{m_{\mathcal{C}r}, r \in \mathcal{C}\}$ são distribuídas multinomialmente com parâmetros $n_{\mathcal{C}}$ e $\boldsymbol{\alpha}_{\mathcal{C}}(\boldsymbol{\theta}, \boldsymbol{\lambda})$, $\mathcal{C} \in \mathcal{P}_m$; além disso, $\boldsymbol{\theta}|\mathbf{m} \sim D(b_r + m_{+r}, r = 1, \ldots, R)$, em que $m_{+r} = \sum_{\mathcal{C}\in\mathcal{P}_r} m_{\mathcal{C}r}$, e $\boldsymbol{\lambda}_r|\mathbf{m} \sim D(a_{\mathcal{C}r} + m_{\mathcal{C}r}, \mathcal{C} \in \mathcal{P}_r)$, $r = 1, \ldots, R$, com todas as variáveis aleatórias mutuamente independentes. Consequentemente, o algoritmo proposto por Soares & Paulino (2007) consiste em se gerar amostras alternadamente dessas distribuições. Apesar de a amostra não conter informação alguma sobre $\{\boldsymbol{\alpha}_{\mathcal{C}}, \mathcal{C} \in \mathcal{P}_m\}$, suas distribuições *a priori* e *a posteriori* não são mais as mesmas nesse caso, porque $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}_o$ e $\{\boldsymbol{\alpha}_{\mathcal{C}}, \mathcal{C} \in \mathcal{P}_m\}$ já não são mutuamente independentes *a priori* e esta dependência leva a uma aprendizagem indireta entre esses parâmetros, conforme Scharfstein *et al.* (2003) apontam.

A DPri 2 parece ser suficientemente flexível para representar informação eliciada por especialistas ou obtida em dados históricos. Contudo, quando tal informação não pode ser obtida, como em muitos casos aplicados, a questão é: que distribuição *a priori* pode ser considerada não-informativa? Escolhas naturais correspondem a fixar todos $\{b_r\}$ e $\{a_{\mathcal{C}r}\}$ iguais a 1, 0.5 ou algum valor pequeno, como 0.1. A primeira opção leva a distribuição *a priori* de Bayes-Laplace, *i.e.*, a distribuições uniformes nos simplex correspondentes. A segunda escolha assemelha-se à distribuição *a priori* de Jeffreys no caso de dados completos. A terceira é ainda mais difusa, na direção da distribuição *a priori* de Haldane, que não pode ser utilizada porque a inidentificabilidade não permite que se usem distribuições *a priori* impróprias para todos os parâmetros, fixando todos $\{b_r\}$ e $\{a_{\mathcal{C}r}\}$ iguais a 0. Se a DPri 1 for considerada, as escolhas correspondentes são ainda menos claras porque as conexões entre os hiperparâmetros não permitem, por exemplo, que se empreguem os mesmos valores para os hiperparâmetros de $\boldsymbol{\theta}$ e $\{\boldsymbol{\lambda}_r\}$.

O problema é que a distribuição *a priori* de Jeffreys não pode ser calculada para modelos inidentificáveis, porque a matriz de informação de Fisher é singular. Num contexto diferente, Wang & Ghosh (2000) contornaram o problema com a seguinte estratégia:

i. escolhem-se alguns parâmetros inidentificáveis que tornam os demais condicionalmente identificáveis,

ii. especifica-se uma versão condicional da distribuição *a priori* de Bayes-Laplace para os parâmetros inidentificáveis selecionados dados os parâmetros condicionalmente identificáveis,

iii. calcula-se a função de verossimilhança marginal correspondente aos parâmetros condicionalmente identificáveis e

iv. obtém-se a distribuição *a priori* marginal de Jeffreys para os parâmetros condicionalmente identificáveis.

Apesar de não existir uma escolha única, uma opção é considerar $\boldsymbol{\theta}$ e, para cada classe $\mathcal{C} \in \mathcal{P}_m$, um $\lambda_{\mathcal{C}(r)}$, $r \in \mathcal{C}$, no conjunto de parâmetros condicionalmente identificáveis, e os outros $\lambda_{\mathcal{C}(r)}$, $\mathcal{C} \in \mathcal{P}_m$, no conjunto de parâmetros inidentificáveis; os parâmetros $\{\lambda_{r(r)}\}$ podem ser obtidos por meio das restrições naturais, e por isto são também considerados condicionalmente identificáveis. Dados os parâmetros condicionalmente identificáveis, a distribuição *a priori* de Bayes-Laplace para os parâmetros inidentificáveis selecionados é uma distribuição uniforme no espaço obtido pelo produto direto dos simplex "reduzidos" de $\boldsymbol{\lambda}_r$, $r = 1, \ldots, R$, gerados ao se fixarem os parâmetros condicionalmente identificáveis. Apesar de esta rota poder ser levada adiante, é mais simples mudar para a parametrização do modelo de mistura de padrões desde que se creia que a suposição de independência entre $(\boldsymbol{\gamma}, \boldsymbol{\alpha}_o)$ e $\{\boldsymbol{\alpha}_{\mathcal{C}}, \mathcal{C} \in \mathcal{P}_m\}$ seja razoável. Como os espaços paramétricos de $(\boldsymbol{\gamma}, \boldsymbol{\alpha}_o)$ e $\{\boldsymbol{\alpha}_{\mathcal{C}}, \mathcal{C} \in \mathcal{P}_m\}$ não são relacionados, a suposição de independência entre eles leva à mesma função de verossimilhança marginal (3.1) sob qualquer distribuição *a priori* para $\{\boldsymbol{\alpha}_{\mathcal{C}}, \mathcal{C} \in \mathcal{P}_m\}$. A distribuição *a priori* marginal de Jeffreys é então $\boldsymbol{\gamma} \sim D(R/2, \{0.5, \mathcal{C} \in \mathcal{P}_m\})$ e $\boldsymbol{\alpha}_o \sim D(0.5, r = 1, \ldots, R)$, com as variáveis aleatórias mutuamente independentes. Ela encaixa-se no figurino da DPri 1 com $a_{rr} = 0.5$, $r = 1, \ldots, R$, e $\sum_{r \in \mathcal{C}} a_{\mathcal{C}r} = 0.5$, $\mathcal{C} \in \mathcal{P}_m$. Para completar a especificação da DPri 1, uma escolha natural é $a_{\mathcal{C}r} = 0.5/\#\mathcal{C}$, $\mathcal{C} \in \mathcal{P}_m$, em que $\#\mathcal{C}$ é a cardinalidade da classe $\mathcal{C}$. Pode-se também quebrar as conexões entre os hiperparâmetros da DPri 1 e considerar a

$$\text{DPri 3}: \begin{cases} \boldsymbol{\gamma} \sim D(R/2, \{0.5, \mathcal{C} \in \mathcal{P}_m\}), \\ \boldsymbol{\alpha}_o \sim D(0.5, r = 1, \ldots, R), \\ \boldsymbol{\alpha}_{\mathcal{C}} \sim D(c_{\mathcal{C}r}, r \in \mathcal{C}),\ \mathcal{C} \in \mathcal{P}_m, \end{cases}$$

com todas as variáveis aleatórias mutuamente independentes. Com a DPri 3 há mais flexibilidade para se usar a distribuição *a priori* de Bayes-Laplace ($c_{\mathcal{C}r} = 1$), uma análoga à distribuição *a priori* de Jeffreys ($c_{\mathcal{C}r} = 0.5$) ou distribuições *a priori* ainda mais difusas (*e.g.*, $c_{\mathcal{C}r} = 0.1$) para esses parâmetros inidentificáveis. Scharfstein *et al.* (2003) apresentam razões pelas quais a suposição de independência supracitada pode não ser adequada na prática, mas isso não parece ter impacto em análises com distribuições *a priori* não-informativas. Da mesma forma que na DPri 1 com essa parametrização, obtêm-se as distribuições *a posteriori* facilmente, levando a $\boldsymbol{\gamma}|\mathbf{N} \sim D(R/2 + n_o, 0.5 + n_{\mathcal{C}}, \mathcal{C} \in \mathcal{P}_m)$, $\boldsymbol{\alpha}_o|\mathbf{N} \sim D(0.5 + n_r, r = 1, \ldots, R)$, $\boldsymbol{\alpha}_{\mathcal{C}}|\mathbf{N} \sim D(c_{\mathcal{C}r}, r \in \mathcal{C})$, $\mathcal{C} \in \mathcal{P}_m$, todas mutuamente independentes.

3.1.3 Análise de sensibilidade clássica

Considere-se novamente os dados da Tabela 1.1 e a notação apresentada no último parágrafo da Seção 3.1.1, mas, por simplicidade, admita que o hábito de fumo da mãe foi sempre observado, ou seja, que a última linha da Tabela 1.1 foi excluída. Sob qualquer das parametrizações consideradas naquela seção, há 7 parâmetros não funcionalmente relacionados, mas apenas 5 frequências observadas dado que o total, n, está fixado. Por essa razão, qualquer modelo com mais do que 5 parâmetros é sobreparametrizado. Algumas alternativas para abordar esta questão são discutidas a seguir.

Kenward *et al.* (2001) e Molenberghs *et al.* (2001) diferenciam dois tipos de incertezas estatísticas: imprecisão estatística e ignorância estatística. A primeira é causada por não se observar a população inteira, e é usualmente quantificada por erros padrões e regiões de confiança. A última é devida a deficiências do processo observacional; *e.g.*, quando algumas respostas estão em falta e/ou podem ser classificadas ou medidas com erro. À medida que o tamanho da amostra tende para infinito, a magnitude da imprecisão estatística decresce para zero, mas a da ignorância estatística pode permanecer inalterada. No caso em foco, a ignorância estatística está relacionada com a distribuição das classes de resposta censuradas em relação às categorias de resposta apropriadas. Uma vez que não é possível realizar inferências frequentistas para modelos inidentificáveis, uma forma usual de se contornar o problema é a de primeiro fixar valores para alguns dos parâmetros inidentificáveis de tal forma que os demais sejam condicionalmente identificáveis e, posteriormente, repetir a análise para diversos conjuntos desses valores avaliando a sensibilidade das inferências. No exemplo do parágrafo anterior, dado que se atribua um valor fixado para $\boldsymbol{\omega} = (\lambda_{\{11,12\}(11)}, \lambda_{\{21,22\}(21)})'$, pode-se estimar $\boldsymbol{\theta}$ e os outros $\{\lambda_{\mathcal{C}(ij)}\}$ como uma função de $\boldsymbol{\omega}$ e também obter regiões de confiança a $100(1-\alpha)\%$ para eles. Depois de se repetir a análise para um conjunto $\boldsymbol{\Omega}$ de valores para os parâmetros de sensibilidade, $\boldsymbol{\omega}$, os autores propõem que se use o intervalo de estimativas pontuais dos parâmetros estimáveis como uma estimativa para a região de ignorância (que quantifica a ignorância estatística) e, por sua vez, que se use a união das regiões de confiança a $100(1-\alpha)\%$ como uma estimativa para a região de incerteza a $100(1-\alpha)\%$ (que mede a incerteza estatística).

Vansteelandt *et al.* (2006) apresentam definições apropriadas de consistência para a região de ignorância e de cobertura para a região de incerteza. As regiões estimadas são chamadas de Região de Ignorância Estimada (RIgE) e Região de Incerteza Estimada (RInE). Para um parâmetro estimável escalar β de interesse, sejam $\boldsymbol{\omega}_l$ e $\boldsymbol{\omega}_u$ os valores em $\boldsymbol{\Omega}$ que correspondem, respectivamente, aos limites inferior e superior do intervalo de ignorância, de modo que a região de ignorância é denotada por $\mathrm{ri}(\beta, \boldsymbol{\Omega}) = [\beta(\boldsymbol{\omega}_l), \beta(\boldsymbol{\omega}_u)] = [\beta_l, \beta_u]$. Diz-se que a RIgE, $\widehat{\mathrm{ri}}(\beta, \boldsymbol{\Omega}) = [\hat{\beta}_l, \hat{\beta}_u]$, é fracamente consistente para o verdadeiro $\beta_0 = \beta(\boldsymbol{\omega}_0)$, se a convergência em probabilidade de $\hat{\beta} = \hat{\beta}(\boldsymbol{\omega})$ para $\beta = \beta(\boldsymbol{\omega})$ ocorre para todos os $\boldsymbol{\omega} \in \boldsymbol{\Omega}$. É claro que sempre se faz a suposição que $\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}$. Regiões de incerteza para β_0 com nível de incerteza $100(1-\alpha)\%$ são definidas de três formas diferentes:

i. RInE fortes cobrem $\beta(\boldsymbol{\omega})$ simultaneamente para todos $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ com uma probabilidade de pelo menos $100(1-\alpha)\%$,

ii. RInE ponto a ponto cobrem $\beta(\boldsymbol{\omega})$ uniformemente sobre $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ com uma probabilidade de pelo menos $100(1-\alpha)\%$,

iii. RInE fracas têm uma sobreposição esperada com a região de ignorância com uma probabilidade de pelo menos $100(1-\alpha)\%$.

Os autores também fornecem algoritmos para construir essas RInE, todos com a forma usual $[\hat{\beta}_l - c_{\alpha^*/2}\,\widehat{\mathrm{ep}}(\hat{\beta}_l), \hat{\beta}_u + c_{\alpha^*/2}\,\widehat{\mathrm{ep}}(\hat{\beta}_u)]$, em que $\hat{\beta}_l$ e $\hat{\beta}_u$ são obtidos de estimadores consistentes e assintoticamente normais de β_l e β_u e $\widehat{\mathrm{ep}}(\hat{\beta}_l)$ e $\widehat{\mathrm{ep}}(\hat{\beta}_u)$ são obtidos de estimadores consistentes dos erros padrões de $\hat{\beta}_l$ e $\hat{\beta}_u$. Para RInE fortes, o valor crítico $c_{\alpha^*/2}$ é o percentil $100(1-\alpha/2)\%$ da distribuição normal padrão. Este tipo de intervalo é usado por Kenward *et al.* (2001) e Molenberghs *et al.* (2001), entre outros. Para RInE ponto a ponto, $c_{\alpha^*/2}$ é a solução de

$$\min\left[\Phi(c_{\alpha^*/2}) - \Phi\left(-c_{\alpha^*/2} - \frac{\hat{\beta}_u - \hat{\beta}_l}{\widehat{\mathrm{ep}}(\hat{\beta}_u)}\right), \Phi\left(c_{\alpha^*/2} + \frac{\hat{\beta}_u - \hat{\beta}_l}{\widehat{\mathrm{ep}}(\hat{\beta}_l)}\right) - \Phi(-c_{\alpha^*/2})\right] = 1-\alpha,$$

em que Φ denota a função de distribuição acumulada da normal padrão. Para RInE fracas, $c_{\alpha^*/2}$ é a solução de

$$\alpha = \frac{\widehat{\mathrm{ep}}(\hat{\beta}_l) + \widehat{\mathrm{ep}}(\hat{\beta}_u)}{\hat{\beta}_u - \hat{\beta}_l} \int_0^{+\infty} z\,\varphi(z + c_{\alpha^*/2})\,dz + \varepsilon,$$

em que φ é a função de densidade da normal padrão e ε é o termo de correção

$$\begin{aligned}\varepsilon = &\int_{(\hat{\beta}_u-\hat{\beta}_l)/\widehat{\mathrm{ep}}(\hat{\beta}_u)}^{+\infty} \varphi(z + c_{\alpha^*/2})\,dz - \frac{\widehat{\mathrm{ep}}(\hat{\beta}_u)}{\hat{\beta}_u - \hat{\beta}_l}\int_{(\hat{\beta}_u-\hat{\beta}_l)/\widehat{\mathrm{ep}}(\hat{\beta}_u)}^{+\infty} z\,\varphi(z + c_{\alpha^*/2})\,dz + \\ &\int_{(\hat{\beta}_u-\hat{\beta}_l)/\widehat{\mathrm{ep}}(\hat{\beta}_l)}^{+\infty} \varphi(z + c_{\alpha^*/2})\,dz - \frac{\widehat{\mathrm{ep}}(\hat{\beta}_l)}{\hat{\beta}_u - \hat{\beta}_l}\int_{(\hat{\beta}_u-\hat{\beta}_l)/\widehat{\mathrm{ep}}(\hat{\beta}_l)}^{+\infty} z\,\varphi(z + c_{\alpha^*/2})\,dz\end{aligned}$$

que pode ser fixado igual a zero a não ser que o tamanho da amostra seja pequeno e/ou houver pouca ignorância sobre β. RInE fortes são RInE ponto a ponto conservadoras, que por sua vez, são RInE fracas conservadoras. Quando há muita ignorância sobre β e o tamanho da amostra é grande, RInE ponto a ponto aproximam-se de RInE fortes. A escolha entre as três versões de RInE depende de qual é a definição mais apropriada para a região de incerteza e de qual é o grau de conservadorismo desejado. As três RInE serão aplicadas nas análises das Seções 3.2 e 3.3.

A dimensão do parâmetro de sensibilidade pode crescer substancialmente com a dimensão da tabela. Em tais casos, devido à carga computacional e à complexidade da escolha de $\boldsymbol{\Omega}$, é comum consider-se modelos sobreparametrizados parcimoniosos para realizar análises de sensibilidade. Nesses casos, a inclusão de restrições é usualmente executada depois de uma reparametrização de $\boldsymbol{\lambda}$, conforme ilustrado nas próximas seções.

3.1.4 Diferenças entre as abordagens bayesiana e clássica e extensões

Muitas críticas aos métodos bayesianos baseiam-se essencialmente na escolha subjetiva da distribuição *a priori*. Tais críticas podem também ser aplicadas às análises de sensibilidade clássicas, devido às escolhas subjetivas dos valores utilizados para o parâmetro de sensibilidade e do modelo sobreparametrizado.

O conjunto $\boldsymbol{\Omega}$ dos valores possíveis para os parâmetros de sensibilidade pode cobrir uma grade detalhada de: (i) todo o espaço paramétrico de $\boldsymbol{\omega}$ ou (ii) apenas uma "extensão plausível" eliciada por especialistas, se a ignorância com relação ao processo gerador dos dados faltantes não for completa. A última alternativa, considerada por Vansteelandt *et al.* (2006) num cenário simples, pode soar como uma distribuição *a priori* para o parâmetro de sensibilidade, mas na verdade ela apenas restringe o espaço paramétrico e não faz qualquer juízo probabilístico com respeito a valores mais ou menos prováveis. Esse esforço para eliminar cautelosamente ignorância desnecessária é bem-vindo, mas não será explorado extensivamente aqui. Isso também pode ser realizado por meio das distribuições *a priori* na abordagem bayesiana e, da mesma forma, pode apenas ser aplicado a parâmetros que tenham significado prático.

Modelos sobreparametrizados parcimoniosos simplificam o problema sob um ponto de vista computacional, mas têm a desvantagem de restringir a atenção a essas famílias particulares de modelos arbitrários. Dependendo das funções paramétricas que se investigam, a análise de sensibilidade resultante pode ser afetada de formas imprevisíveis como se mostra nas próximas seções. Por essa razão, a menos que esses modelos também traduzam crenças de especialistas, pode ser mais razoável usar os modelos de omissão irrestritos (ou não-paramétricos) da Seção 3.1.1. Para isso, pode-se usar a estratégia descrita no último parágrafo da Seção 3.1.2, *i.e.*, dividir $(\boldsymbol{\theta}, \boldsymbol{\lambda})$ em parâmetros condicionalmente identificáveis e inidentificáveis selecionados para serem respectivamente empregados como parâmetros estimáveis e de sensibilidade; adicionalmente, pode-se usar $\{\log(\lambda_{\mathcal{C}(r)}/\lambda_{r(r)}), \mathcal{C} \in \mathcal{P}_r \cap \mathcal{P}_m\}$ em vez de $\boldsymbol{\lambda}_r$ para incorporar mais facilmente as restrições naturais e também simplificar a tarefa de trabalhar com $\boldsymbol{\Omega}$ no espaço euclidiano em vez de em simplex. Uma alternativa é usar $(\boldsymbol{\gamma}, \boldsymbol{\alpha}_o)$ como parâmetros estimáveis e $\{\boldsymbol{\alpha}_{\mathcal{C}}, \mathcal{C} \in \mathcal{P}_m\}$ como parâmetros de sensibilidade. A região de ignorância correspondente, também chamada de não-paramétrica ou de limites de melhor-pior caso em tal contexto, é previsivelmente muito mais extensa, mas um ponto de partida sensato para uma modelagem cautelosa, conforme defendem Kenward *et al.* (2001).

Modelos reduzidos, sobreparametrizados ou não, podem também ser considerados sob a abordagem bayesiana, mas as distribuições *a priori* e/ou algoritmos descritos na Seção 3.1.2 necessitam modificação. Por exemplo, para ajustar modelos log-lineares para $\boldsymbol{\theta}$, seguindo a proposta de Bedrick, Christensen & Johnson (1996), Soares & Paulino (2007) usaram as distribuições *a priori* induzidas pela DPri 2 para os parâmetros do modelo log-linear e então condicionaram-nas relativamente aos parâmetros supostos iguais a zero. As ideias de amostragem em fatia de Neal (2003) são empregadas em um dos passos do algoritmo de ampliação de dados em cadeia para se obter uma amostra da

distribuição *a posteriori* dos parâmetros do modelo log-linear. A amostragem em fatia também é usada por Soares (2004) para se obter uma amostra da distribuição *a posteriori* dos parâmetros de modelos de omissão reduzidos (*e.g.*, modelo de omissão aleatória: $\lambda_{\mathcal{C}(r)} = \tau_{\mathcal{C}(\mathcal{C})}, \forall r \in \mathcal{C}$). Para superar a difícil tarefa de se trabalhar com distribuições *a priori* que levam em conta ambas as restrições naturais e do modelo, ele usa ideias de Walker (1996) e penaliza a distribuição *a priori* do modelo irrestrito na direção do modelo postulado.

Resumindo, ambas as abordagens têm algum grau de subjetividade, mas esta é uma característica inevitável nas análises de dados incompletos que se deve controlar cuidadosamente, conforme apontam Molenberghs *et al.* (2001). A escolha de uma rota particular de análise de sensibilidade é uma questão de facilidade de implementação ou adequação ao contexto do problema, ou ainda de alguma preferência pessoal. Por exemplo, Scharfstein *et al.* (2003) insatisfazem-se por não obter um único resumo com a região de ignorância da estratégia clássica, enquanto Vansteelandt *et al.* (2006) não utilizam a maquinaria bayesiana por considerarem que as distribuições *a priori* acabam eliminando os valores extremos que são obtidos em intervalos de ignorância e de incerteza. Nas seções subsequentes, mostra-se que as médias *a posteriori* podem alterar-se facilmente com pequenas mudanças das distribuições *a priori*, mesmo para tamanhos de amostra grandes. Por essa razão, crê-se que a região de ignorância clássica é uma alternativa interessante. Apesar disso, as conclusões da abordagem bayesiana não devem ser obtidas de médias, mas de regiões ou intervalos de credibilidade, que refletem tanto a ignorância quanto a imprecisão estatística e, consequentemente, evitam o dilema de se ter que escolher entre as versões de RInE da abordagem clássica. Usar um modelo sobreparametrizado reduzido e/ou limitar a extensão de $\boldsymbol{\Omega}$ pode ofuscar extremos ainda mais do que a ponderação bayesiana, mas todos esses efeitos podem ser contornados, seja usando um modelo sobreparametrizado irrestrito, seja tentando cobrir todo o espaço paramétrico de $\boldsymbol{\omega}$ ou ainda usando distribuições *a priori* difusas.

3.2 Estudo de simulação

3.2.1 Descrição das análises

Para explorar e comparar as abordagens bayesiana e clássica, consideram-se conjuntos de dados simulados com tamanhos $n = 40$, 400, 4 000, 40 000 e 4 000 000, todos com as mesmas frequências relativas exibidas na Tabela 3.2.

Tabela 3.2: Frequências relativas dos dados simulados.

$Y_1 \setminus Y_2$	1	2	omisso
1	0.05	0.35	0.05
2	0.05	0.40	0.10

Novamente, substitui-se o índice r pelos índices, i, j, para indicar, respectivamente, os valores possíveis de Y_1 e Y_2, que podem ser iguais a 1 ou 2. Nessa configuração de padrão de omissão

monótono, como Y_1 é sempre observado, suas probabilidades marginais (*e.g.*, θ_{2+}) são identificáveis. Entretanto, a diferença entre as probabilidades marginais de Y_1 e Y_2 (*e.g.*, $\theta_{+2} - \theta_{2+}$), a razão de chances (RC $= \theta_{11}\theta_{22}/[\theta_{12}\theta_{21}]$), ou seu logaritmo, são inidentificáveis, porque eles não podem ser expressos apenas em função dos parâmetros identificáveis $\boldsymbol{\gamma}$ e $\boldsymbol{\alpha}_o$. Uma comparação dos resultados para esses dois conjuntos de funções paramétricas para tamanhos de amostra crescentes fornece uma nítida distinção entre casos em que as análises são afetadas apenas por imprecisão estatística ou, adicionalmente, por ignorância estatística.

Sob uma perspectiva bayesiana, consideram-se as distribuições *a priori* não-informativas mencionadas na Seção 3.1.2, nomeadamente, tanto aquelas com todos os hiperparâmetros fixados em 1.0, 0.5 e 0.1 nas DPri 2 e 3, quanto a sugestão feita para se completar a especificação da distribuição *a priori* de Jeffreys encaixada na DPri 1, que aqui, induz a DPri 3 com $\{c_{\mathcal{C}ij} = 0.25\}$. Também se empregou a DPri 3 com $\{c_{\mathcal{C}ij} = 0.001\}$. Para contemplar distribuições *a priori* levemente informativas, considera-se a DPri 2 com os hiperparâmetros

$$
\begin{aligned}
(b_{11}, b_{12}, b_{21}, b_{22}) &= (0.1, 0.3, 0.1, 0.5), \\
(a_{\{11\}11}, a_{\{11,12\}11}) = (0.6, 0.4), &\qquad (a_{\{12\}12}, a_{\{11,12\}12}) = (0.8, 0.2), \\
(a_{\{21\}21}, a_{\{21,22\}21}) = (0.7, 0.3), &\qquad (a_{\{22\}22}, a_{\{21,22\}22}) = (0.8, 0.2),
\end{aligned}
$$

e também outra versão com todos os valores multiplicados por 10. Como os hiperparâmetros de cada distribuição ($\boldsymbol{\theta}$ e $\{\boldsymbol{\lambda}_{ij}\}$) somam 1 e 10, rotula-se essas distribuições *a priori* de Inf1 e Inf10, respectivamente. Além disso, chama-se essas distribuições *a priori* de *levemente informativas*, porque elas não assumem probabilidades iguais e também porque os parâmetros de precisão das correspondentes distribuições de Dirichlet são pequenos quando comparados com tamanhos de amostra $n \geq 400$. Em primeiro lugar, combina-se esses hiperparâmetros para $\{b_{ij}\}$ com $\{a_{\mathcal{C}ij} = 0.5\}$, e depois, inversamente, usam-se os valores de $\{a_{\mathcal{C}ij}\}$ com $\{b_{ij} = 0.5\}$, de forma que se podem comparar ambos os casos com $\{b_{ij} = a_{\mathcal{C}ij} = 0.5\}$ e avaliar a mudança das distribuições *a priori* de $\boldsymbol{\theta}$ e $\boldsymbol{\lambda}$ separadamente. As médias *a priori* resultantes para θ_{2+}, $\theta_{+2} - \theta_{2+}$ $(= \theta_{12} - \theta_{21})$ e $\log \mathrm{RC}$ são, respectivamente, 0.5, 0.0 e 0.0, para as distribuições *a priori* não-informativas, e 0.6, 0.2 e > 0.5, para as distribuições *a priori* informativas. Os resumos *a posteriori* para esses parâmetros são exibidos nas Tabelas 3.3, 3.5 e 3.7. Métodos usuais de diagnóstico foram empregados para se avaliar a convergência das cadeias nas análises baseadas na DPri 2 (Heidelberger & Welch, 1983; Geweke, 1992; Raftery & Lewis, 1992). Para todas as distribuições *a priori* e tamanhos de amostra, as magnitudes dos erros de Monte Carlo são menores do que a precisão dos valores apresentados nas tabelas. Concordando com Agresti & Min (2005) relativamente a uma possível mudança de conclusão ao se analisar 1/RC em vez de RC com intervalos de maior densidade *a posteriori*, utilizam-se intervalos de credibilidade (IC) centrais.

Reparametrizando $\boldsymbol{\lambda}$ como

$$
\text{logito}\left(\lambda_{\{ij\}(ij)}\right) = \phi_0 + \phi_1(i-1) + \phi_2(j-1) + \phi_3(i-1)(j-1), \quad i, j = 1, 2,
$$

consideram-se três modelos sobreparametrizados para a análise de sensibilidade clássica. Os Modelos A e B são, respectivamente, obtidos com as restrições $\phi_3 = 0$ e $\phi_2 = 0$; os parâmetros de sensibilidade correspondentes são $\boldsymbol{\omega} = \phi_2$ e $\boldsymbol{\omega} = \phi_3$. O modelo irrestrito, com parâmetros de sensibilidade $\boldsymbol{\omega} = (\phi_2, \phi_3)$, é rotulado de Modelo C. Embora ϕ_2 e ϕ_3 admitam qualquer valor real, usa-se pragmaticamente $\boldsymbol{\Omega} = [-20; 20]$ para os Modelos A e B e $\boldsymbol{\Omega} = [-20; 20]^2$ para o Modelo C. As grades são construídas com incrementos de 0.1 e 1, respectivamente. Aumentar a extensão de $\boldsymbol{\Omega}$ não altera os resultados, que são dispostos nas Tabelas 3.4, 3.6 e 3.8. RInE fracas não foram calculadas para θ_{2+} uma vez que não há ignorância sobre este parâmetro. Para este parâmetro identificável, é interessante notar que as RInE ponto a ponto coincidem com as RInE fortes, correspondendo aos intervalos obtidos por meio da usual distribuição normal assintótica, e também que a RIgE fica reduzida a uma estimativa pontual. Apresentam-se os resultados usando a versão exata das RInE fracas (*i.e.*, calculando ε), mas comparam-se esses resultados aos obtidos considerando $\varepsilon = 0$ e notam-se diferenças desprezáveis (< 0.01 para $\log \mathrm{RC}$ e < 0.001 para $\theta_{+2} - \theta_{2+}$) para as RInE resultantes sob todos os modelos com $n > 40$ ou sob os Modelos A (para $\theta_{+2} - \theta_{2+}$) e C (para ambas funções paramétricas) com $n = 40$.

3.2.2 Resultados

O efeito das distribuições *a priori* nos resumos *a posteriori* de θ_{2+} praticamente desaparece para $n \geq 400$ (Tabela 3.3). Ademais, as médias e os IC coincidem, respectivamente, com as RIgE (pontuais) e as RInE para tais tamanhos de amostra (Tabela 3.4). Esses resultados usualmente encontrados em análises de parâmetros identificáveis não são compartilhados pelos obtidos para parâmetros inidentificáveis ($\theta_{+2} - \theta_{2+}$ e $\log \mathrm{RC}$).

As médias, desvios padrões (DP) e IC *a posteriori* para $\theta_{+2} - \theta_{2+}$ e $\log \mathrm{RC}$ variam substancialmente para as diferentes distribuições *a priori* (Tabelas 3.5 e 3.7). A dependência relativamente à distribuição *a priori* foi também exibida por Forster & Smith (1998) e Soares & Paulino (2001), mas o fato de o nível de dependência permanecer praticamente inalterado para $n \geq 4\,000$ é alarmante. Apesar de as flutuações das médias *a posteriori* não serem muito grandes quando comparadas com os correspondentes DP, as análises de sensibilidade das distribuições *a priori* sugerem que não se deve confiar muito neste único resumo. Ao lidar com ignorância estatística, há sempre um conjunto de distribuições (ampliadas) não-observáveis compatível com a distribuição que se pode observar. Distribuições *a posteriori* apontam na direção de um ou outro subconjunto dependendo da distribuição *a priori* escolhida. Sintetizar demais neste cenário de falta de informação pode ser inadequado. Acredita-se que é mais apropriado considerar IC.

Para esses parâmetros inidentificáveis, os DP *a posteriori* não tendem mais para zero conforme o tamanho de amostra cresce. Pelo contrário, eles repousam em patamares que são drasticamente diferentes dependendo da distribuição *a priori*. Essas diferenças são mais impressionantes para as distribuições *a priori* comumente consideradas como não-informativas. Obviamente, não há nada

Tabela 3.3: Médias, desvios padrões (DP) e intervalos de credibilidade (IC) centrais a 95% *a posteriori* para θ_{2+}.

DPri 3	$\{c_{\mathcal{C}ij} = 0.1\}$			$\{c_{\mathcal{C}ij} = 0.25\}$			$\{c_{\mathcal{C}ij} = 0.5\}$			$\{c_{\mathcal{C}ij} = 1.0\}$		
n	Média	DP	IC 95%	Média	DP	IC 95%	Média	DP	IC 95%	Média	DP	IC 95%
40	0.546	0.075	[0.398; 0.691]	0.546	0.075	[0.398; 0.691]	0.546	0.075	[0.399; 0.690]	0.546	0.075	[0.398; 0.691]
400	0.550	0.025	[0.501; 0.598]	0.550	0.025	[0.501; 0.598]	0.550	0.025	[0.501; 0.598]	0.550	0.025	[0.501; 0.598]
4 000	0.550	0.008	[0.534; 0.565]	0.550	0.008	[0.535; 0.565]	0.550	0.008	[0.535; 0.565]	0.550	0.008	[0.535; 0.565]
40 000	0.550	0.002	[0.545; 0.555]	0.550	0.002	[0.545; 0.555]	0.550	0.002	[0.545; 0.555]	0.550	0.002	[0.545; 0.555]
4 000 000	0.550	0.000	[0.550; 0.550]	0.550	0.000	[0.550; 0.550]	0.550	0.000	[0.550; 0.550]	0.550	0.000	[0.550; 0.550]
DPri 3/2	$\{c_{\mathcal{C}ij} = 0.001\}$			$\{b_{ij} = a_{\mathcal{C}ij} = 0.1\}$			$\{b_{ij} = a_{\mathcal{C}ij} = 0.5\}$			$\{b_{ij} = a_{\mathcal{C}ij} = 1.0\}$		
n	Média	DP	IC 95%	Média	DP	IC 95%	Média	DP	IC 95%	Média	DP	IC 95%
40	0.547	0.075	[0.398; 0.691]	0.550	0.077	[0.397; 0.698]	0.547	0.075	[0.398; 0.693]	0.546	0.074	[0.399; 0.689]
400	0.550	0.025	[0.501; 0.598]	0.550	0.025	[0.501; 0.599]	0.550	0.025	[0.501; 0.598]	0.549	0.025	[0.500; 0.597]
4 000	0.550	0.008	[0.535; 0.565]	0.550	0.008	[0.535; 0.565]	0.550	0.008	[0.535; 0.565]	0.550	0.008	[0.534; 0.565]
40 000	0.550	0.002	[0.545; 0.555]	0.550	0.002	[0.545; 0.555]	0.550	0.002	[0.545; 0.555]	0.550	0.002	[0.545; 0.555]
4 000 000	0.550	0.000	[0.550; 0.550]	0.550	0.000	[0.550; 0.550]	0.550	0.000	[0.550; 0.550]	0.550	0.000	[0.550; 0.550]
DPri 2	$\{b_{ij}\} = \text{Inf1}$, $\{a_{\mathcal{C}ij} = 0.5\}$			$\{b_{ij}\} = \text{Inf10}$, $\{a_{\mathcal{C}ij} = 0.5\}$			$\{b_{ij} = 0.5\}$, $\{a_{\mathcal{C}ij}\} = \text{Inf1}$			$\{b_{ij} = 0.5\}$, $\{a_{\mathcal{C}ij}\} = \text{Inf10}$		
n	Média	DP	IC 95%	Média	DP	IC 95%	Média	DP	IC 95%	Média	DP	IC 95%
40	0.551	0.077	[0.399; 0.698]	0.560	0.070	[0.422; 0.694]	0.548	0.076	[0.397; 0.693]	0.547	0.076	[0.397; 0.693]
400	0.550	0.025	[0.502; 0.598]	0.551	0.025	[0.503; 0.599]	0.550	0.025	[0.501; 0.597]	0.550	0.025	[0.500; 0.598]
4 000	0.550	0.008	[0.535; 0.565]	0.550	0.008	[0.535; 0.565]	0.550	0.008	[0.535; 0.565]	0.550	0.008	[0.535; 0.566]
40 000	0.550	0.002	[0.545; 0.555]	0.550	0.002	[0.545; 0.555]	0.550	0.002	[0.545; 0.555]	0.550	0.002	[0.545; 0.555]
4 000 000	0.550	0.000	[0.550; 0.550]	0.550	0.000	[0.550; 0.550]	0.550	0.000	[0.550; 0.550]	0.550	0.000	[0.550; 0.550]

DPri 3: $\boldsymbol{\gamma} \sim D(2, 0.5, 0.5)$, $\boldsymbol{\alpha}_o \sim D(0.5, 0.5, 0.5, 0.5)$, $\boldsymbol{\alpha}_{\mathcal{C}} \sim D(c_{\mathcal{C}ij}, ij \in \mathcal{C})$, $\mathcal{C} \in \mathcal{P}_m$.
DPri 2: $\boldsymbol{\theta} \sim D(b_{11}, b_{12}, b_{21}, b_{22})$, $\boldsymbol{\lambda}_{ij} \sim D(a_{\mathcal{C}ij}, \mathcal{C} \in \mathcal{P}_{ij})$, $ij = 11, \ldots, 22$.
Inf1: $\{b_{ij}\} = (0.1, 0.3, 0.1, 0.5)$, $\{a_{\mathcal{C}11}\} = (0.6, 0.4)$, $\{a_{\mathcal{C}12}\} = (0.8, 0.2)$, $\{a_{\mathcal{C}21}\} = (0.7, 0.3)$, $\{a_{\mathcal{C}22}\} = (0.8, 0.2)$. Inf10: Inf1 $\times 10$.

Tabela 3.4: RIgE e RInE fortes e ponto a ponto a 95% para θ_{2+} usando estimativas de máxima verossimilhança.

	Modelos A, B e C - RIgE: [0.550; 0.550]
n	Forte e Ponto a ponto
40	[0.396; 0.704]
400	[0.501; 0.599]
4 000	[0.535; 0.565]
40 000	[0.545; 0.555]
4 000 000	[0.550; 0.550]

Modelos A ($\phi_3 = 0$, $\boldsymbol{\omega} = \phi_2$) e B ($\phi_2 = 0$, $\boldsymbol{\omega} = \phi_3$) são encaixados no Modelo C ($\boldsymbol{\omega} = (\phi_2, \phi_3)$).

Tabela 3.5: Médias, desvios padrões (DP) e intervalos de credibilidade (IC) centrais a 95% *a posteriori* para $\theta_{+2} - \theta_{2+}$.

DPri 3	$\{c_{\mathcal{C}ij} = 0.1\}$			$\{c_{\mathcal{C}ij} = 0.25\}$			$\{c_{\mathcal{C}ij} = 0.5\}$			$\{c_{\mathcal{C}ij} = 1.0\}$		
n	Média	DP	IC 95%	Média	DP	IC 95%	Média	DP	IC 95%	Média	DP	IC 95%
40	0.256	0.110	[0.029; 0.461]	0.256	0.107	[0.037; 0.455]	0.256	0.103	[0.046; 0.450]	0.256	0.099	[0.055; 0.445]
400	0.273	0.060	[0.159; 0.379]	0.273	0.055	[0.166; 0.373]	0.273	0.050	[0.174; 0.366]	0.273	0.045	[0.185; 0.357]
4 000	0.275	0.052	[0.190; 0.358]	0.275	0.047	[0.194; 0.354]	0.275	0.041	[0.202; 0.347]	0.275	0.034	[0.212; 0.337]
40 000	0.275	0.051	[0.197; 0.352]	0.275	0.046	[0.200; 0.350]	0.275	0.040	[0.205; 0.345]	0.275	0.032	[0.215; 0.335]
4 000 000	0.275	0.051	[0.200; 0.350]	0.275	0.046	[0.201; 0.349]	0.275	0.039	[0.205; 0.344]	0.275	0.032	[0.216; 0.334]
DPri 3/2	$\{c_{\mathcal{C}ij} = 0.001\}$			$\{b_{ij} = a_{\mathcal{C}ij} = 0.1\}$			$\{b_{ij} = a_{\mathcal{C}ij} = 0.5\}$			$\{b_{ij} = a_{\mathcal{C}ij} = 1.0\}$		
n	Média	DP	IC 95%	Média	DP	IC 95%	Média	DP	IC 95%	Média	DP	IC 95%
40	0.256	0.113	[0.021; 0.465]	0.275	0.111	[0.045; 0.480]	0.270	0.105	[0.054; 0.467]	0.260	0.102	[0.053; 0.452]
400	0.273	0.064	[0.153; 0.385]	0.277	0.060	[0.162; 0.382]	0.281	0.051	[0.179; 0.372]	0.282	0.045	[0.190; 0.366]
4 000	0.275	0.057	[0.186; 0.361]	0.277	0.052	[0.190; 0.359]	0.282	0.041	[0.204; 0.350]	0.284	0.034	[0.217; 0.342]
40 000	0.275	0.056	[0.196; 0.354]	0.277	0.051	[0.198; 0.352]	0.282	0.040	[0.207; 0.346]	0.284	0.032	[0.220; 0.339]
4 000 000	0.275	0.056	[0.200; 0.350]	0.277	0.051	[0.200; 0.350]	0.282	0.040	[0.207; 0.346]	0.284	0.032	[0.220; 0.339]
DPri 2	$\{b_{ij}\} = \text{Inf1}$, $\{a_{\mathcal{C}ij} = 0.5\}$			$\{b_{ij}\} = \text{Inf10}$, $\{a_{\mathcal{C}ij} = 0.5\}$			$\{b_{ij} = 0.5\}$, $\{a_{\mathcal{C}ij}\} = \text{Inf1}$			$\{b_{ij} = 0.5\}$, $\{a_{\mathcal{C}ij}\} = \text{Inf10}$		
n	Média	DP	IC 95%	Média	DP	IC 95%	Média	DP	IC 95%	Média	DP	IC 95%
40	0.291	0.104	[0.075; 0.486]	0.271	0.093	[0.082; 0.448]	0.253	0.108	[0.031; 0.455]	0.289	0.099	[0.089; 0.475]
400	0.289	0.050	[0.185; 0.377]	0.286	0.049	[0.185; 0.372]	0.265	0.055	[0.160; 0.367]	0.299	0.037	[0.222; 0.368]
4 000	0.289	0.040	[0.208; 0.352]	0.287	0.040	[0.207; 0.350]	0.265	0.046	[0.191; 0.349]	0.299	0.022	[0.248; 0.337]
40 000	0.289	0.039	[0.210; 0.348]	0.287	0.039	[0.209; 0.347]	0.265	0.046	[0.198; 0.347]	0.299	0.020	[0.253; 0.331]
4 000 000	0.289	0.039	[0.210; 0.347]	0.287	0.039	[0.209; 0.347]	0.265	0.045	[0.200; 0.347]	0.299	0.020	[0.254; 0.330]

DPri 3: $\boldsymbol{\gamma} \sim D(2, 0.5, 0.5)$, $\boldsymbol{\alpha}_o \sim D(0.5, 0.5, 0.5, 0.5)$, $\boldsymbol{\alpha}_{\mathcal{C}} \sim D(c_{\mathcal{C}ij}, ij \in \mathcal{C})$, $\mathcal{C} \in \mathcal{P}_m$.
DPri 2: $\boldsymbol{\theta} \sim D(b_{11}, b_{12}, b_{21}, b_{22})$, $\boldsymbol{\lambda}_{ij} \sim D(a_{\mathcal{C}ij}, \mathcal{C} \in \mathcal{P}_{ij})$, $ij = 11, \ldots, 22$.
Inf1: $\{b_{ij}\} = (0.1, 0.3, 0.1, 0.5)$, $\{a_{\mathcal{C}11}\} = (0.6, 0.4)$, $\{a_{\mathcal{C}12}\} = (0.8, 0.2)$, $\{a_{\mathcal{C}21}\} = (0.7, 0.3)$, $\{a_{\mathcal{C}22}\} = (0.8, 0.2)$. Inf10: Inf1 $\times 10$.

Tabela 3.6: RIgE e RInE fortes, ponto a ponto e fracas a 95% para $\theta_{+2} - \theta_{2+}$ usando estimativas de máxima verossimilhança.

	Modelo B - RIgE: [0.244; 0.344]			Modelos A e C - RIgE: [0.200; 0.350]		
n	Forte	Ponto a ponto	Fraca	Forte	Ponto a ponto	Fraca
40	[0.027; 0.522]	[0.056; 0.498]	[0.073; 0.484]	[−0.010; 0.527]	[0.022; 0.500]	[0.052; 0.475]
400	[0.175; 0.400]	[0.186; 0.391]	[0.208; 0.373]	[0.134; 0.406]	[0.144; 0.397]	[0.173; 0.373]
4 000	[0.222; 0.362]	[0.226; 0.359]	[0.240; 0.347]	[0.179; 0.368]	[0.182; 0.365]	[0.200; 0.350]
40 000	[0.237; 0.349]	[0.238; 0.348]	[0.246; 0.342]	[0.193; 0.356]	[0.194; 0.355]	[0.204; 0.347]
4 000 000	[0.243; 0.344]	[0.243; 0.344]	[0.246; 0.341]	[0.199; 0.351]	[0.199; 0.350]	[0.204; 0.347]

Modelos A ($\phi_3 = 0$, $\boldsymbol{\omega} = \phi_2$) e B ($\phi_2 = 0$, $\boldsymbol{\omega} = \phi_3$) são encaixados no Modelo C ($\boldsymbol{\omega} = (\phi_2, \phi_3)$).

Tabela 3.7: Médias, desvios padrões (DP) e intervalos de credibilidade (IC) centrais a 95% *a posteriori* para log RC.

DPri 3	$\{c_{\mathcal{C}ij} = 0.1\}$			$\{c_{\mathcal{C}ij} = 0.25\}$			$\{c_{\mathcal{C}ij} = 0.5\}$			$\{c_{\mathcal{C}ij} = 1.0\}$		
n	Média	DP	IC 95%	Média	DP	IC 95%	Média	DP	IC 95%	Média	DP	IC 95%
40	−0.02	1.15	[−2.25; 2.29]	−0.04	1.07	[−2.10; 2.11]	−0.06	0.97	[−1.95; 1.90]	−0.06	0.88	[−1.78; 1.68]
400	−0.04	0.77	[−1.39; 1.36]	−0.05	0.69	[−1.30; 1.25]	−0.07	0.61	[−1.18; 1.12]	−0.08	0.51	[−1.02; 0.93]
4 000	−0.04	0.72	[−1.18; 1.13]	−0.05	0.64	[−1.12; 1.07]	−0.07	0.55	[−1.03; 0.97]	−0.08	0.45	[−0.89; 0.80]
40 000	−0.04	0.71	[−1.12; 1.07]	−0.05	0.63	[−1.09; 1.04]	−0.07	0.55	[−1.02; 0.96]	−0.08	0.44	[−0.87; 0.79]
4 000 000	−0.04	0.71	[−1.10; 1.05]	−0.05	0.63	[−1.09; 1.04]	−0.07	0.54	[−1.02; 0.96]	−0.08	0.44	[−0.87; 0.79]
DPri 3/2	$\{c_{\mathcal{C}ij} = 0.001\}$			$\{b_{ij} = a_{\mathcal{C}ij} = 0.1\}$			$\{b_{ij} = a_{\mathcal{C}ij} = 0.5\}$			$\{b_{ij} = a_{\mathcal{C}ij} = 1.0\}$		
n	Média	DP	IC 95%	Média	DP	IC 95%	Média	DP	IC 95%	Média	DP	IC 95%
40	−0.01	1.24	[−2.40; 2.46]	−0.04	1.27	[−2.50; 2.53]	−0.01	1.10	[−2.15; 2.21]	0.00	0.98	[−1.92; 1.95]
400	−0.03	0.84	[−1.47; 1.44]	−0.02	0.78	[−1.41; 1.37]	−0.02	0.63	[−1.19; 1.16]	−0.02	0.54	[−1.04; 1.02]
4 000	−0.03	0.79	[−1.21; 1.17]	−0.02	0.72	[−1.18; 1.13]	−0.02	0.56	[−1.03; 0.98]	−0.02	0.46	[−0.88; 0.83]
40 000	−0.03	0.78	[−1.13; 1.09]	−0.02	0.71	[−1.12; 1.07]	−0.02	0.55	[−1.02; 0.97]	−0.03	0.45	[−0.87; 0.82]
4 000 000	−0.03	0.78	[−1.10; 1.05]	−0.03	0.71	[−1.10; 1.05]	−0.02	0.55	[−1.02; 0.97]	−0.03	0.45	[−0.87; 0.82]
DPri 2	$\{b_{ij}\} = \text{Inf1}$, $\{a_{\mathcal{C}ij} = 0.5\}$			$\{b_{ij}\} = \text{Inf10}$, $\{a_{\mathcal{C}ij} = 0.5\}$			$\{b_{ij} = 0.5\}$, $\{a_{\mathcal{C}ij}\} = \text{Inf1}$			$\{b_{ij} = 0.5\}$, $\{a_{\mathcal{C}ij}\} = \text{Inf10}$		
n	Média	DP	IC 95%	Média	DP	IC 95%	Média	DP	IC 95%	Média	DP	IC 95%
40	0.03	1.22	[−2.36; 2.47]	0.14	0.98	[−1.77; 2.08]	0.01	1.10	[−2.11; 2.24]	0.14	1.06	[−1.96; 2.28]
400	0.02	0.63	[−1.17; 1.19]	0.06	0.61	[−1.13; 1.18]	0.01	0.68	[−1.24; 1.30]	0.19	0.43	[−0.68; 1.02]
4 000	0.02	0.56	[−1.03; 1.00]	0.05	0.55	[−1.00; 0.99]	0.01	0.63	[−1.08; 1.08]	0.21	0.31	[−0.46; 0.76]
40 000	0.02	0.55	[−1.01; 0.97]	0.05	0.54	[−0.99; 0.98]	0.01	0.62	[−1.07; 1.05]	0.21	0.30	[−0.43; 0.72]
4 000 000	0.02	0.55	[−1.01; 0.97]	0.05	0.54	[−0.99; 0.98]	0.01	0.62	[−1.07; 1.04]	0.21	0.29	[−0.42; 0.72]

DPri 3: $\boldsymbol{\gamma} \sim D(2, 0.5, 0.5)$, $\boldsymbol{\alpha}_o \sim D(0.5, 0.5, 0.5, 0.5)$, $\boldsymbol{\alpha}_{\mathcal{C}} \sim D(c_{\mathcal{C}ij}, ij \in \mathcal{C})$, $\mathcal{C} \in \mathcal{P}_m$.
DPri 2: $\boldsymbol{\theta} \sim D(b_{11}, b_{12}, b_{21}, b_{22})$, $\boldsymbol{\lambda}_{ij} \sim D(a_{\mathcal{C}ij}, \mathcal{C} \in \mathcal{P}_{ij})$, $ij = 11, \ldots, 22$.
Inf1: $\{b_{ij}\} = (0.1, 0.3, 0.1, 0.5)$, $\{a_{\mathcal{C}11}\} = (0.6, 0.4)$, $\{a_{\mathcal{C}12}\} = (0.8, 0.2)$, $\{a_{\mathcal{C}21}\} = (0.7, 0.3)$, $\{a_{\mathcal{C}22}\} = (0.8, 0.2)$. Inf10: Inf1 $\times 10$.

Tabela 3.8: RIgE e RInE fortes, ponto a ponto e fracas a 95% para log RC usando estimativas de máxima verossimilhança.

	Modelo A - RIgE: [−0.27; 0.22]			Modelo B - RIgE: [−0.97; 0.36]			Modelo C - RIgE: [−1.10; 1.05]		
n	Forte	Ponto a ponto	Fraca	Forte	Ponto a ponto	Fraca	Forte	Ponto a ponto	Fraca
40	[−1.73; 2.29]	[−1.59; 2.10]	[−1.54; 2.02]	[−2.72; 2.43]	[−2.45; 2.12]	[−2.24; 1.87]	[−2.84; 2.88]	[−2.56; 2.59]	[−2.14; 2.15]
400	[−0.73; 0.88]	[−0.66; 0.77]	[−0.58; 0.66]	[−1.52; 1.01]	[−1.43; 0.91]	[−1.21; 0.65]	[−1.65; 1.63]	[−1.56; 1.54]	[−1.25; 1.21]
4 000	[−0.42; 0.43]	[−0.39; 0.40]	[−0.33; 0.30]	[−1.14; 0.56]	[−1.11; 0.53]	[−0.98; 0.37]	[−1.27; 1.23]	[−1.24; 1.20]	[−1.07; 1.02]
40 000	[−0.32; 0.29]	[−0.31; 0.28]	[−0.27; 0.22]	[−1.02; 0.42]	[−1.01; 0.41]	[−0.94; 0.32]	[−1.15; 1.11]	[−1.14; 1.10]	[−1.05; 1.00]
4 000 000	[−0.28; 0.23]	[−0.28; 0.23]	[−0.26; 0.21]	[−0.97; 0.36]	[−0.97; 0.36]	[−0.93; 0.32]	[−1.10; 1.06]	[−1.10; 1.05]	[−1.05; 0.99]

Modelos A ($\phi_3 = 0$, $\boldsymbol{\omega} = \phi_2$) e B ($\phi_2 = 0$, $\boldsymbol{\omega} = \phi_3$) são encaixados no Modelo C ($\boldsymbol{\omega} = (\phi_2, \phi_3)$).

de errado em se usar distribuições *a priori* sugeridas por um especialista. Não obstante, quando se busca uma distribuição *a priori* realmente não-informativa, não há escolha óbvia, mas acredita-se que uma escolha adequada certamente deve ter os hiperparâmetros muito menores do que os valores 0.5 ou 1.0 usualmente escolhidos para a família Dirichlet. Repetiram-se as análises baseadas na DPri 3 com distribuições *a priori* cada vez mais difusas e os resultados estabilizaram-se para $n = 4\,000\,000$ apenas para valores dos hiperparâmetros menores do que 0.001. Com o mesmo espírito, deve-se ter em mente que as distribuições *a priori* usualmente consideradas levemente informativas podem na verdade modificar distribuições *a posteriori* para qualquer tamanho de amostra.

Distribuições *a priori* com todos os hiperparâmetros menores do que 1.0 conduzem a densidades *a posteriori* multimodais para $\theta_{+2} - \theta_{2+}$ e $\log \mathrm{RC}$ quando $n \geq 4\,000$ (veja na Figura 3.1 algumas densidades *a posteriori* obtidas a partir da DPri 3). O mesmo comportamento é observado para $\boldsymbol{\theta}$, quando $n \geq 400$, e para $\boldsymbol{\lambda}$, independentemente do tamanho de amostra (resultados não apresentados). Essa multimodalidade pode não ser surpreendente, levando em consideração as densidades beta em forma de U adotadas como distribuições *a priori* para os parâmetros inidentificáveis. Contudo, no caso de dados contínuos, Scharfstein *et al.* (2003) apontam que isso também pode ocorrer quando distribuições *a priori* informativas para $\boldsymbol{\theta}$ e $\boldsymbol{\lambda}$ não são simultaneamente 'compatíveis' com os dados observados; eles sugerem que, nesses casos, os especialistas reavaliem suas crenças reduzindo a quantidade de informação de uma dessas distribuições *a priori*. Crê-se, aqui, que a unimodalidade não deve ser esperada para densidades *a posteriori* de parâmetros inidentificáveis. Pelo menos para as análises consideradas neste capítulo, a multimodalidade pode ser mais comum que o usual, dado que a distribuição *a posteriori* de $\boldsymbol{\theta}$ pode ser escrita como uma mistura finita de distribuições Dirichlet. Por conseguinte, essa é uma característica inerente que deve ser tomada em consideração e não simplesmente evitada.

Os resultados da inferência clássica usando o (parcimonioso) Modelo A coincidem com aqueles baseados no (irrestrito) Modelo C para $\theta_{+2} - \theta_{2+}$, mas não para $\log \mathrm{RC}$ (Tabelas 3.6 e 3.8). Isto mostra que para funções paramétricas diferentes, modelos de omissão sobreparametrizados podem afetar as inferências de várias formas. Por exemplo, no caso da RC, obviamente a presença da interação entre Y_1 e Y_2 no modelo de omissão é mais importante do que a presença do efeito principal de Y_2. Isso leva ao (*ad hoc*) Modelo B não-hierárquico. A RIgE para $\log \mathrm{RC}$ baseada no Modelo B é consideravelmente mais extensa do que aquela resultante do Modelo A, mas seu limite superior ainda está muito distante daquele obtido com o Modelo C. Por outro lado, o limite inferior da RIgE para $\theta_{+2} - \theta_{2+}$ baseado no Modelo B também está distante dos correspondentes obtidos com os Modelos A e C. Os intervalos mais estreitos obtidos sob os Modelos A e B são bem-vindos, mas apenas se estes modelos fizerem sentido para os especialistas. Em caso contrário, eles podem gerar conclusões equivocadas.

Comparações entre as abordagens bayesiana e clássica fazem sentido apenas quando restringidas às distribuições *a priori* menos informativas e ao Modelo C. Os IC são, em geral, próximos das RInE

ponto a ponto nos casos em que os hiperparâmetros se aproximam de zero. A diferença entre os dois intervalos decresce à medida que se aumenta o tamanho de amostra.

É natural buscarem-se modelos de omissão mais parcimoniosos, uma vez que levam a inferências

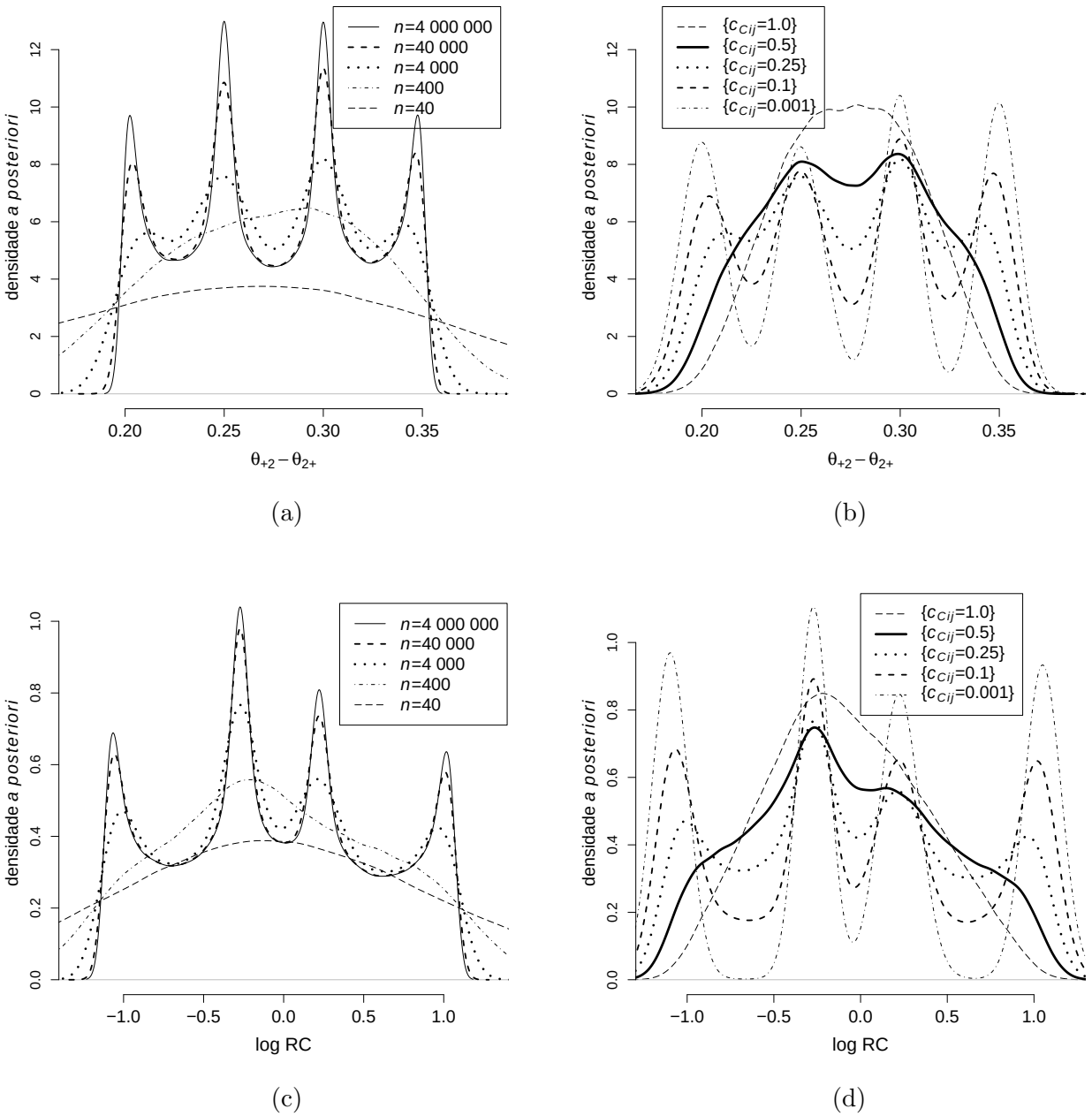


Figura 3.1: Densidades *a posteriori* de $\theta_{+2} - \theta_{2+}$, (a) e (b), e de $\log \mathrm{RC}$, (c) e (d), usando a DPri 3, $\boldsymbol{\gamma} \sim D(2, 0.5, 0.5)$, $\boldsymbol{\alpha}_o \sim D(0.5, 0.5, 0.5, 0.5)$, $\boldsymbol{\alpha}_{\mathcal{C}} \sim D(c_{\mathcal{C}ij}, ij \in \mathcal{C})$, $\mathcal{C} \in \mathcal{P}_m$, com $\{c_{\mathcal{C}ij} = 0.25\}$ em (a) e (c), e $n = 4\,000$ em (b) e (d).

mais precisas. O chamado modelo de omissão aleatória (MAR, de *missing at random*) é usualmente o foco de atenção, não apenas porque em muitos problemas faz sentido supor que a ocorrência de dados faltantes depende apenas dos dados observados, mas também pela ignorabilidade de $\boldsymbol{\lambda}$ do ponto de vista de inferências verossimilhancistas ou bayesianas sobre $\boldsymbol{\theta}$, quando ambos os tipos de parâmetros não são relacionados funcionalmente ou *a priori* (Rubin, 1976). No caso desta seção, o modelo MAR pode ser obtido sob as restrições $\lambda_{\{11,12\}(11)} = \lambda_{\{11,12\}(12)}$ e $\lambda_{\{21,22\}(21)} = \lambda_{\{21,22\}(22)}$. Comparar as distribuições *a posteriori* destes parâmetros individuais como em Soares & Paulino (2001) pode erroneamente apontar para uma rejeição da hipótese MAR, ao passo que as diferenças entre esses parâmetros indicam um retrato mais justo. Isso fica claro na Figura 3.2, onde se dispõem diagramas em caixa (*box plots*) dos valores sorteados *a posteriori* dos parâmetros, e diferenças entre eles, com base na DPri 2 com os hiperparâmetros correspondentes às distribuições *a priori* mais informativas e com $n = 4\,000\,000$. O modelo MAR não é rejeitado *a posteriori* para quaisquer das distribuições *a priori* e tamanhos de amostra. A avaliação da hipótese MAR também pode ser conduzida com a abordagem clássica se $\boldsymbol{\lambda}$ for transferido para o conjunto de parâmetros estimáveis. Contudo, não é possível distinguir entre modelos MAR e não-MAR apenas com base na informação presente nos dados, conforme mostram, por exemplo, Molenberghs *et al.* (2008). Não obstante, distinções podem ocorrer se distribuições *a priori* mais informativas ou faixas mais estreitas para $\boldsymbol{\Omega}$ forem utilizadas, e/ou modelos reduzidos forem adotados.

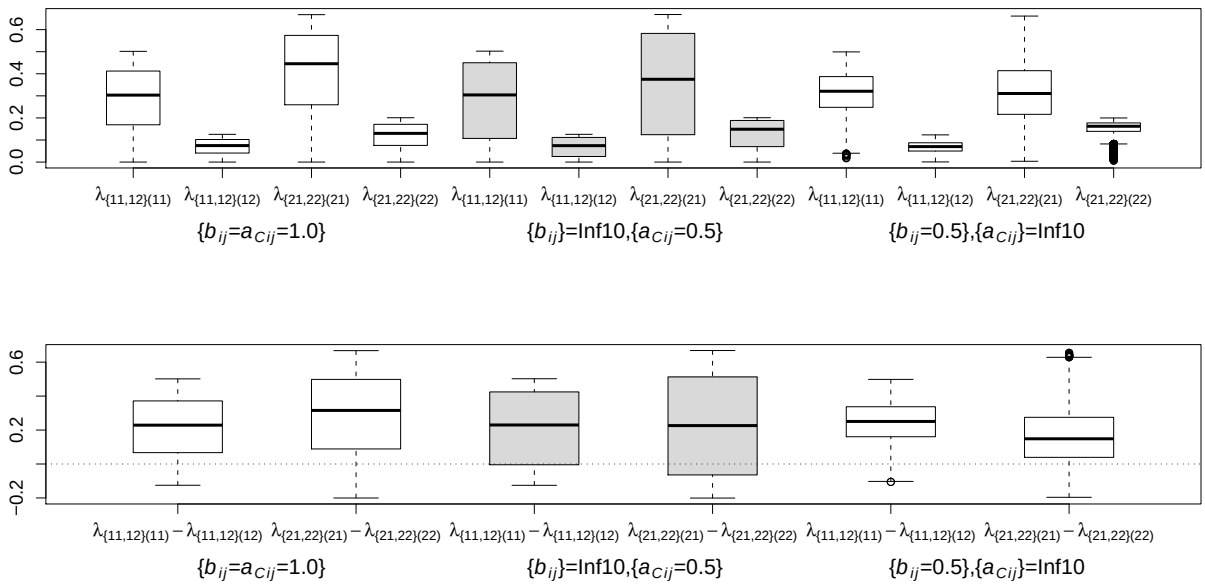


Figura 3.2: Diagramas em caixa de alguns parâmetros de $\boldsymbol{\lambda}$ *a posteriori*, e diferenças deles, para a DPri 2, $\boldsymbol{\theta} \sim D(b_{11}, b_{12}, b_{21}, b_{22})$, $\boldsymbol{\lambda}_{ij} \sim D(a_{\mathcal{C}ij}, \mathcal{C} \in \mathcal{P}_{ij})$, $ij = 11, \ldots, 22$, com $n = 4\,000\,000$. Inf10: $\{b_{ij}\} = (1, 3, 1, 5)$, $\{a_{\mathcal{C}11}\} = (6, 4)$, $\{a_{\mathcal{C}12}\} = (8, 2)$, $\{a_{\mathcal{C}21}\} = (7, 3)$, $\{a_{\mathcal{C}22}\} = (8, 2)$.

Pelas mesmas razões porque a hipótese MAR não pode ser aceita ou rejeitada com base nos resultados presentes, não se pode afirmar se há uma associação entre Y_1 e Y_2 ou não. Na prática, não-rejeições são usualmente tomadas como evidência a favor da hipótese nula. O uso de testes de hipóteses sem análise do poder para diferentes alternativas pode não ser considerado um procedimento condenável em análises de modelos identificáveis regulares quando se espera que o poder do teste seja alto. Todavia, isto é mais perigoso quando a hipótese envolve parâmetros inidentificáveis. Por definição, testes obtidos via RInE têm poder muito baixo para grande parte das alternativas que pertencem à região de ignorância, independentemente do tamanho da amostra. Da mesma forma, IC baseados em distribuições *a priori* para os quais hiperparâmetros têm os valores próximos de zero também são geralmente inúteis para rejeitar a hipótese se a alternativa está dentro do intervalo de ignorância não-paramétrico. Esses intervalos mais amplos oferecem proteção contra suposições inverificáveis, mas necessitam de uma avaliação cuidadosa.

Apesar de se ter tratado de uma série de questões importantes para análises de modelos inidentificáveis, a rejeição da homogeneidade marginal (ou simetria) e demais conclusões inferenciais permanecem inalteradas para todas as distribuições *a priori* e modelos considerados. Na próxima seção, apresenta-se um diferente cenário motivador.

3.3 Reanálise dos dados do *Collaborative Perinatal Project*

Reanalisam-se os dados da Tabela 1.1 nos moldes da seção anterior. Para considerar modelos sobreparametrizados parcimoniosos na análise de sensibilidade clássica, constroem-se modelos condicionais considerados por Fay (1986) em que, como sugerido por Molenberghs *et al.* (1999), os elementos de $\boldsymbol{\lambda}$ são primeiro reparametrizados por $\psi_{1(ij)}$, a probabilidade de observar o hábito de fumo materno, $\psi_{21(ij)}$, a probabilidade condicional de observar o peso do recém-nascido dado que o hábito de fumo materno é observado, e $\psi_{20(ij)}$, a probabilidade condicional de observar o peso do recém-nascido dado que o hábito de fumo materno está em falta. Estas probabilidades são também condicionadas ao hábito de fumo materno e ao peso do recém-nascido. Em segundo lugar, reparametrizam-se os parâmetros ψ como

$$\begin{aligned}
\text{logito}\left(\psi_{1(ij)}\right) &= \phi_{10} + \phi_{11}(i-1) + \phi_{12}(j-1) + \phi_{13}(i-1)(j-1),\\
\text{logito}\left(\psi_{21(ij)}\right) &= \phi_{20} + \phi_{21}(i-1) + \phi_{22}(j-1) + \phi_{23}(i-1)(j-1),\\
\text{logito}\left(\psi_{20(ij)}\right) &= \phi_{30} + \phi_{31}(i-1) + \phi_{32}(j-1) + \phi_{23}(i-1)(j-1).
\end{aligned}$$

Finalmente, consideram-se modelos reduzidos com a imposição das seguintes restrições

$$\begin{aligned}
&\text{Modelo 1}: \phi_{1l} = \phi_{2l} = \phi_{3l} = \phi_l, \quad l = 1,2,3,\\
&\text{Modelo 2}: \phi_{21} = \phi_{31}, \quad \phi_{22} = \phi_{32}, \quad \phi_{13} = \phi_{23} = \phi_{33} = 0.
\end{aligned}$$

Esses modelos possuem, respectivamente, 9 e 10 parâmetros não redundantes quando combinados com $\boldsymbol{\theta}$. Como há apenas 8 frequências observadas (dado que o total, $n = 57,061$, é fixado), os Modelos 1 e 2 possuem, respectivamente, 1 (ϕ_3) e 2 (ϕ_{11} e ϕ_{22}) parâmetros de sensibilidade, enquanto o modelo irrestrito tem 7 parâmetros deste tipo. Conduz-se a análise de sensibilidade clássica com o modelo de omissão irrestrito conforme descrito na Seção 3.1.4. Os resultados obtidos por meio das abordagens bayesiana e clássica são dispostos, respectivamente, nas Tabelas 3.9 e 3.10.

Tabela 3.9: Médias, desvios padrões (DP) e intervalos de credibilidade (IC) centrais a 95% *a posteriori* para RC.

DPri	Hiperparâmeters	Média	DP	IC 95%
3	$\{c_{\mathcal{C}ij} = 0.001\}$	1.49	0.40	[0.85; 2.40]
3	$\{c_{\mathcal{C}ij} = 0.1\}$	1.48	0.35	[0.90; 2.27]
3	$\{c_{\mathcal{C}ij}\} = {}^*$	1.47	0.33	[0.93; 2.19]
3	$\{c_{\mathcal{C}ij} = 0.5\}$	1.45	0.26	[1.02; 2.01]
3	$\{c_{\mathcal{C}ij} = 1.0\}$	1.44	0.21	[1.09; 1.89]
2	$\{b_{ij} = a_{\mathcal{C}ij} = 0.1\}$	1.48	0.35	[0.90; 2.28]
2	$\{b_{ij} = a_{\mathcal{C}ij} = 0.5\}$	1.46	0.26	[1.03; 2.02]
2	$\{b_{ij} = a_{\mathcal{C}ij} = 1.0\}$	1.46	0.21	[1.10; 1.90]

* $c_{\mathcal{C}ij} = 0.125$, se $\mathcal{C} = \{11, 12, 21, 22\}$, e $c_{\mathcal{C}ij} = 0.25$, em caso contrário.
DPri 3: $\boldsymbol{\gamma} \sim D(2, 0.5, 0.5, 0.5, 0.5, 0.5)$, $\boldsymbol{\alpha}_o \sim D(0.5, 0.5, 0.5, 0.5)$,
$\boldsymbol{\alpha}_{\mathcal{C}} \sim D(c_{\mathcal{C}ij}, ij \in \mathcal{C})$, $\mathcal{C} \in \mathcal{P}_m$.
DPri 2: $\boldsymbol{\theta} \sim D(b_{11}, b_{12}, b_{21}, b_{22})$, $\boldsymbol{\lambda}_{ij} \sim D(a_{\mathcal{C}ij}, \mathcal{C} \in \mathcal{P}_{ij})$, $ij = 11, \ldots, 22$.

Tabela 3.10: RIgE e RInE fortes, ponto a ponto e fracas a 95% para RC usando estimativas de máxima verossimilhança.

Modelo	$\boldsymbol{\Omega}$	RIgE	Forte	Ponto a ponto	Fraca
1	$[-3; 3]$	[1.06; 2.04]	[1.01; 2.14]	[1.02; 2.12]	[1.07; 2.02]
1	$[-5; 5]$	[0.94; 2.23]	[0.90; 2.33]	[0.91; 2.32]	[0.96; 2.19]
1	$[-10; 10]$	[0.89; 2.32]	[0.84; 2.45]	[0.85; 2.43]	[0.90; 2.28]
2	$[-10; 10]^2$	[1.10; 1.73]	[1.06; 1.81]	[1.07; 1.80]	[1.11; 1.73]
Irrestrito	$[-10; 10]^7$	[0.82; 2.50]	[0.79; 2.61]	[0.79; 2.59]	[0.84; 2.44]

Modelos 1, 2 e irrestrito possuem, respectivamente, 1, 2 e 7 parâmetros de sensibilidade.

Além da discussão anterior, nota-se que mesmo o IC baseado na distribuição *a priori* menos informativa (DPri 3 com $\{c_{\mathcal{C}ij} = 0.001\}$) está agora mais próximo da RInE fraca do que a RInE ponto a ponto para os resultados obtidos com o modelo de omissão irrestrito. O problema aqui é a possibilidade de obter diferentes conclusões dependendo das escolhas subjetivas em cada uma das abordagens. A menos que alguma informação *a priori* sugira que é razoável restringir inferências para o Modelo 2 ou para o Modelo 1 com $\boldsymbol{\Omega} = [-3; 3]$, ou usar uma distribuição *a priori* informativa (ou aquelas "não-informativas" para as quais os hiperparâmetros tenham valores ≥ 0.5), não se pode concluir se o hábito de fumo da mãe e o peso do recém-nascido estão associados ou não. Essas

conclusões são diferentes das de Baker *et al.* (1992), que conduzem uma análise de sensibilidade informal baseada em 9 modelos identificáveis que reforçam suas crenças numa associação significativa. Kenward *et al.* (2001) apresentam outro exemplo em que uma análise de sensibilidade informal que leva a conclusões consistentes é enganadora, mas é interessante notar que isto pode ocorrer quando o tamanho de amostra é tão grande quanto $n = 57\,061$ e a proporção de dados com omissão é tão pequena quanto 7%.

3.4 Discussão

No esforço de se incluir aspectos relevantes de um problema em sua análise estatística, pode-se frequentemente chegar a modelos inidentificáveis. Exploraram-se abordagens bayesiana e clássica de análise desses modelos especializados no caso de respostas categorizadas com omissão. Ambas as metodologias fornecem respostas sensatas se uma atenção meticulosa for dedicada às escolhas subjetivas e se houver uma compreensão clara dos pontos fortes e das limitações das análises. Em caso contrário, pode-se obter conclusões equivocadas.

A questão primordial para compreender análises de modelos sobreparametrizados relaciona-se com a distinção entre dados observados e não-observados, ou com o que se pode ou não aprender do estudo. À primeira vista, simplificar o modelo tornando-o identificável parece ser uma boa decisão que direciona o analista para o que se pode aprender do estudo e o previne de ter que lidar com a parte obscura não-observável. De fato, até certo grau, inferências obtidas destes modelos são muito menos incertas, mas se baseiam em suposições subjacentes inverificáveis. Assim, os modelos identificáveis acabam por ser uma forma extrema e subjetiva de se lidar com a ignorância sobre os dados não-observáveis.

No processo de se lidar com a ignorância estatística, realizando análises de modelos inidentificáveis, novas questões apareceram. Particularmente, concluiu-se que tamanhos de amostra maiores não aumentam o conhecimento sobre parâmetros inidentificáveis quando a imprecisão estatística já é suficiente pequena. Assintoticamente, intervalos de credibilidade e de incerteza cobrem o intervalo de ignorância. Por essa razão, os primeiros intervalos são inúteis se houver interesse em se distinguir valores dos parâmetros que pertencem ao intervalo de ignorância. Como consequência, se a alternativa pertence ao intervalo de ignorância, o poder dos testes já não tendem mais para um à medida que o tamanho da amostra cresce. Esta é a razão pela qual é inadmissível considerar uma não-rejeição de uma hipótese como sua aceitação, em geral. Isto deve ser sempre esclarecido em qualquer relatório que envolva esse tipo de análise.

Uma estratégia objetiva é a de se usar modelos de omissão irrestritos, distribuições *a priori* difusas nas análises bayesianas ou tentar cobrir todo o espaço paramétrico do parâmetro de sensibilidade sob a abordagem clássica. Apesar de este ser um ponto de partida razoável que, em muitos casos, pode ser a única alternativa, acredita-se que uma análise mais apropriada deve incorporar toda a informação disponível. Isto pode ser atingido seguindo a rota oposta, *i.e.*, considerando distribuições

a priori informativas ou faixas de valores plausíveis para os parâmetros de sensibilidade e/ou modelos de omissão reduzidos. O objetivo deste capítulo é apenas ilustrar que tais informações adicionadas devem ser avaliadas com cautela, porque escolhas arbitrárias podem ter um impacto significativo nos resultados.

Sob a perspectiva bayesiana, mostrou-se que distribuições *a priori* usualmente consideradas como não-informativas ou levemente informativas (*e.g.*, fixando os valores dos hiperparâmetros das distribuições Dirichlet iguais a 0.5 ou 1.0) podem na verdade adicionar informação substancial, independentemente do tamanho de amostra. Além disso, modelos sobreparametrizados reduzidos podem também afetar os resultados consideravelmente. Embora tais modelos também possam ser utilizados em análises bayesianas, eles são mais frequentemente adotados na abordagem clássica, principalmente pela dificuldade de se lidar com muitos parâmetros de sensibilidade. Na reanálise do *Collaborative Perinatal Project*, por exemplo, consideraram-se apenas 3 valores, $\{-10, 0, 10\}$, para cada um dos 7 parâmetros de sensibilidade, resultando em $3^7 = 2\,187$ combinações. Usar uma grade tão grosseira poderia ter gerado um intervalo de ignorância ainda menor do que os obtidos com modelos de omissão reduzidos, mas não foi difícil, para essa tabela de contingência particular, identificar as configurações que levam a inferências extremas para a razão de chances e, como consequência, pode-se confirmar o intervalo de ignorância do modelo irrestrito.

Como os intervalos de credibilidade obtidos com distribuições *a priori* difusas ficaram, de certa forma, próximos dos intervalos de incerteza obtidos com modelos de omissão irrestritos, a abordagem bayesiana pode oferecer vantagens sobre a análise de sensibilidade clássica quando (a) há informação *a priori* para ser incorporada ou (b) nenhuma informação *a priori* está disponível, mas há interesse de se obter resultados para um modelo de omissão irrestrito com uma alta dimensão do parâmetro de sensibilidade. Neste caso, contudo, as distribuições *a priori* difusas (próprias) são convenientes para se evitar a obtenção de intervalos de credibilidade muito estreitos, mas estes intervalos podem ainda ser um pouco menos amplos do que os intervalos de incerteza fraca que seriam obtidos com a metodologia clássica.

Embora a DPri 2 seja a mais apropriada para incorporar informação *a priori*, as cadeias geradas pelo método de amostragem de Gibbs tornam-se altamente autocorrelacionadas conforme o tamanho de amostra cresce e os valores dos hiperparâmetros aproximam-se de zero. O tempo computacional que se leva para executar algumas das análises varia de horas a semanas, dependendo se $n \leq 4\,000$, $n = 40\,000$ ou $4\,000\,000$. Entretanto, pode-se usar o método ordinário de Monte Carlo com a DPri 3 e conduzir análises para quaisquer valores de hiperparâmetros em alguns minutos. Por essa razão, quando não há informação *a priori*, sugere-se usar a distribuição *a priori* marginal de Jeffreys para os parâmetros identificáveis com os valores dos hiperparâmetros da distribuição *a priori* dos parâmetros inidentificáveis próximos de zero.

Quando se tenta incorporar informação disponível sobre o mecanismo gerador dos dados com omissão, não só covariáveis podem ser úteis, mas também as justificativas para os dados faltantes

podem facilitar a tarefa de se incluir certas suposições, como MAR ou não-MAR, para algumas unidades amostrais. Em estudos longitudinais, o tempo geralmente é um fator adicional importante. Quando o padrão de omissão é monótono, pode ser razoável assegurar que as respostas das unidades que abandonaram o estudo não dependem dos valores futuros não-observados (Diggle & Kenward, 1994; Kenward *et al.*, 2003), levando a um modelo sobreparametrizado mais parcimonioso do que o irrestrito.

Capítulo 4

Análise bayesiana semiparamétrica de respostas binárias com uma covariável contínua sujeita a omissão informativa

Em diversos estudos, dados omissos ocorrem em pelo menos algumas variáveis explicativas. Nesses casos, para não ter que excluir as unidades amostrais ou as variáveis que contêm dados faltantes da análise, além de se modelar o mecanismo de omissão, se este não puder ser ignorado, é necessário especificar um modelo para a distribuição marginal das variáveis explicativas[1] ($\mathbf{X}$) mesmo se o interesse recair somente na distribuição condicional das variáveis respostas ($\mathbf{Y}$) dado $\mathbf{X}$. Quando todas as variáveis de $\mathbf{X}$ são categorizadas e o número de combinações dos níveis dessas variáveis for bem menor que o número de unidades amostrais, pode ser razoável supor que $\mathbf{X}$ siga uma distribuição multinomial[2]; esta e outras estratégias de modelagem semelhantes foram estudadas por Ibrahim (1990), Gibbons & Hosmer (1991), Vach & Schumacher (1993), Lipsitz & Ibrahim (1996), Lipsitz, Parzen & Ewell (1998), Horton & Laird (1999), Satten & Carroll (2000) e Horton & Laird (2001) sob suposições de omissão ignorável e análises admitindo omissão não-ignorável foram propostas por Vach & Blettner (1995), Vach (1997), Ibrahim, Lipsitz & Chen (1999), Lipsitz, Ibrahim, Chen & Peterson (1999) e Paik (2004). Nos casos em que alguma variável explicativa é contínua, pode não existir informação *a priori* sobre algum modelo paramétrico plausível. Por esta razão, alguns autores adotaram modelos semiparamétricos/não-paramétricos para $\mathbf{X}$ após identificar um modelo com a suposição de omissão ignorável (Chen & Little, 1999; Chen, 2002, 2004, 2009; Zhang & Rockette, 2005, 2006, 2007; Zhao, 2009). Outra vertente seguiu um caminho oposto, considerando modelos paramétricos para $\mathbf{X}$ e mecanismos de omissão informativa (Lipsitz, Ibrahim, Chen & Peterson, 1999; Huang, Chen & Ibrahim, 2005; Stubbendick & Ibrahim, 2003, 2006; Miranda & Rabe-Hesketh, 2010). Uma excelente revisão sobre análise de modelos lineares generalizados com covariáveis omissas

[1] Ou pelo menos especificar a distribuição condicional das variáveis explicativas que possam estar em falta dadas as variáveis explicativas que são sempre observadas.

[2] Supondo também que a distribuição condicional de $\mathbf{Y}$ dado $\mathbf{X}$ siga um produto de distribuições multinomiais, a análise da distribuição multinomial resultante para a distribuição conjunta de $(\mathbf{X}, \mathbf{Y})$ pode ser realizada utilizando a teoria descrita em Poleto (2006) com o auxílio do conjunto de subrotinas computacionais Catdata implementado para o ambiente estatístico R (disponível em http://www.poleto.com/missing.html).

é apresentada por Ibrahim, Chen, Lipsitz & Herring (2005).

Suposições incorretas, quer para o mecanismo de omissão quer para a distribuição das covariáveis, podem gerar inferências enviesadas para a distribuição condicional das respostas dadas as covariáveis. Neste capítulo pretende-se utilizar uma distribuição flexível para $\mathbf{X}$ juntamente com análises de sensibilidade para o mecanismo de omissão, informativa ou não. Chen (2004) comenta que sua metodologia pode ser estendida para casos de omissão não-ignorável. Contudo, não se segue essa via aqui porque como sua abordagem já é computacionalmente intensiva no caso de omissão ignorável, talvez seja inviável realizá-la repetidamente, como é necessário na alternativa clássica de análise de sensibilidade qualificada de "formal" nos últimos capítulos. Então, por pragmatismo, adota-se a abordagem bayesiana para a análise de sensibilidade do mecanismo de omissão; esse enfoque lida com a falta de identificabilidade do modelo por meio de distribuições *a priori* próprias. Além disso, utiliza-se um modelo não-paramétrico para $\mathbf{X}$ baseado numa mistura por processo de Dirichlet. Uma abordagem bayesiana não-paramétrica similar foi utilizada por Scharfstein *et al.* (2003) para o caso de uma variável resposta contínua com omissão. Por simplicidade, restringe-se aqui ao caso de apenas uma covariável contínua omissa, embora na última seção se comente sobre possíveis extensões para o caso multivariado.

Na Seção 4.1, apresenta-se um conjunto de dados reais que será utilizado para ilustrar os métodos descritos nas seções seguintes. Na Seção 4.2, introduz-se a abordagem bayesiana não-paramétrica com processo de Dirichlet para o caso de análise de dados completos. Na Seção 4.3, estende-se o modelo para contemplar também o mecanismo gerador dos dados omissos. Na Seção 4.4, realiza-se um estudo de simulação para avaliar o desempenho do método proposto. Na Seção 4.5, analisa-se o conjunto de dados da Seção 4.1. Na Seção 4.6, conclui-se o capítulo com uma breve discussão.

4.1 Dados de embolia pulmonar

Wicki *et al.* (2001) analisaram dados de 1 090 pacientes que se apresentaram consecutivamente ao pronto-socorro do Hospital Universitário de Genebra com suspeita clínica de embolia pulmonar, *i.e.*, bloqueio da artéria pulmonar ou de um de seus ramos. O objetivo do estudo foi criar um escore que indicasse a probabilidade de pacientes terem essa doença cardiovascular com base em exames e outras informações facilmente obtidas. Por simplicidade, considera-se aqui apenas algumas das variáveis explicativas do modelo final apresentado por esses autores.

O indicador de presença de embolia pulmonar (variável resposta) bem como quatro variáveis explicativas (idade, ocorrência anterior de embolia pulmonar ou trombose venosa profunda, realização de cirurgia recente e pulsação arterial) foram observadas para todos os pacientes, enquanto as duas variáveis que indicam presenças de determinadas características no raio-X do tórax (atelectasia laminar e elevação da hemicúpula diafragmática) estavam em falta para apenas um paciente. Por outro lado, a pressão parcial de gás carbônico ($PaCO_2$), obtida da gasometria arterial realizada a partir de um exame de sangue, estava omissa para 103 (9%) pacientes.

Na Figura 4.1, apresentam-se o histograma de $PaCO_2$ e estimativas de densidades pelo método do núcleo gaussiano baseadas nos dados observados e em valores sorteados das distribuições preditivas

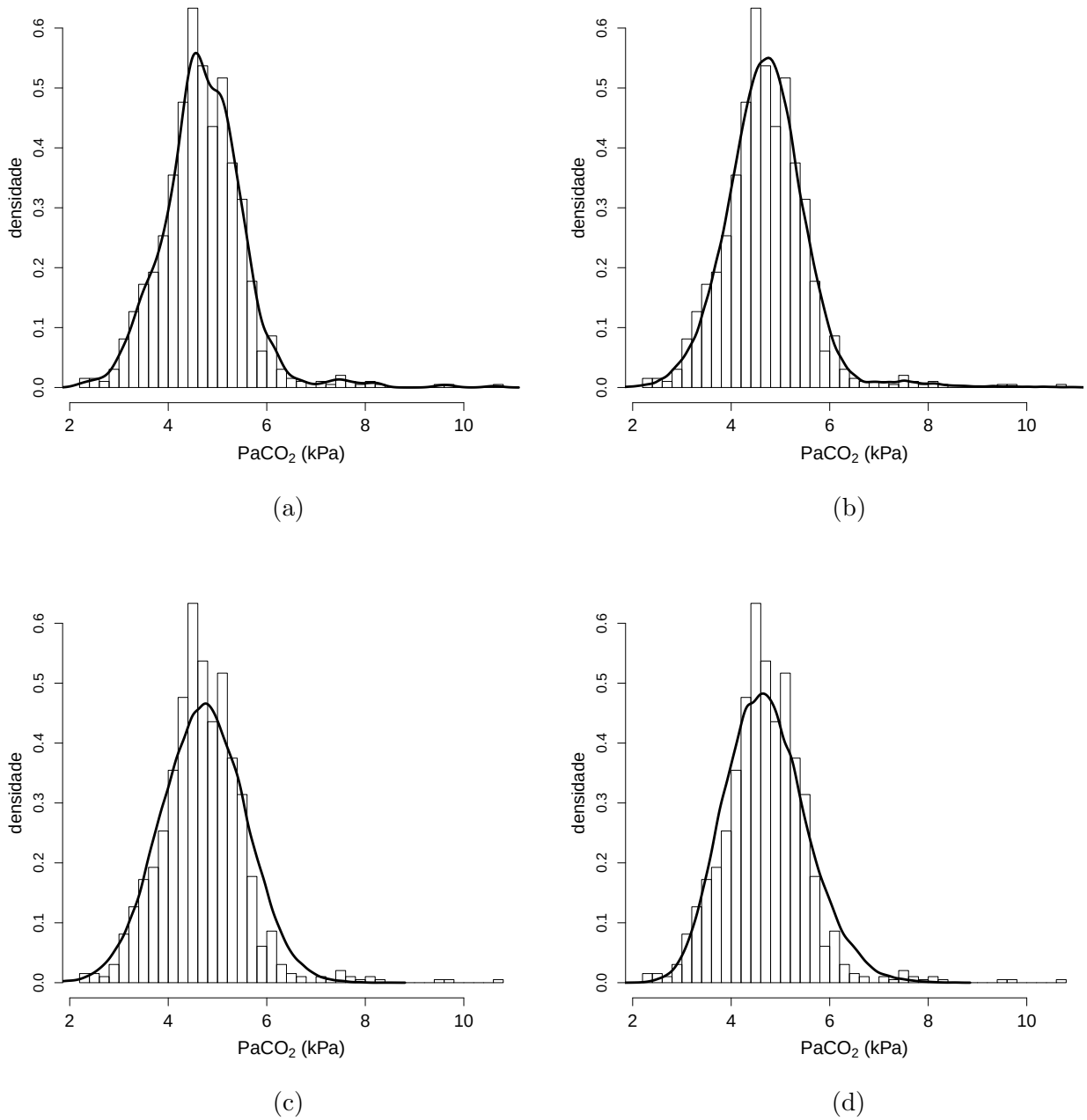


Figura 4.1: Histograma da pressão parcial de gás carbônico ($PaCO_2$), em kPa, e estimativas de densidades pelo método do núcleo gaussiano baseadas (a) nos dados observados e (b)-(d) em 50 000 valores amostrados das distribuições preditivas *a posteriori* obtidas a partir de ajustes dos modelos (b) não-paramétrico, (c) normal e (d) gama.

a posteriori obtidas a partir dos ajustes do modelo não-paramétrico da Seção 4.2 e dos dois modelos paramétricos que obtiveram os melhores ajustes dentre as famílias normal, log-normal e gama. Nota-se que os dados observados parecem ser mais bem acomodados pela densidade preditiva *a posteriori* do modelo não-paramétrico do que pelas densidades correspondentes dos modelos paramétricos. O mesmo resultado é evidenciado a partir da estatística de Kolmogorov-Smirnov que, na comparação da distribuição empírica dos dados observados com as distribuições *a posteriori* de uma nova observação dos modelos não-paramétrico e paramétricos normal, log-normal e gama, produzem, respectivamente, os valores 3.4%, 6.1%, 6.8% e 5.7% (valores-p 0.207, 0.002, <0.001 e 0.004).

4.2 Modelos não-paramétricos para variáveis contínuas completas

Seja X_i, $i = 1, \ldots, n$, uma amostra aleatória de tamanho n com função de distribuição F. Nos métodos paramétricos usuais, supõe-se uma forma conhecida para F, indexada por um parâmetro de dimensão finita especificado *a priori*, embora geralmente desconhecido. Com o objetivo de enfraquecer as suposições, permitindo maior flexibilidade na modelagem e robustez contra má-especificação de F, consideram-se modelos não-paramétricos. Apesar da ideia que o termo incute, o seu emprego não significa que os modelos são completamente desprovidos de parâmetros; em vez disso, indica que o número e a natureza dos parâmetros são variáveis e determinados de alguma forma pelos dados, podendo chegar até a pertencer a um espaço paramétrico de dimensão infinita. Revisões de alguns métodos não-paramétricos sob os paradigmas bayesiano e clássico são apresentadas por, respectivamente, Müller & Quintana (2004) e Scott (1992).

Uma maneira de flexibilizar a forma de F é empregar medidas de probabilidade aleatórias (MPA), que são distribuições de probabilidade sobre o espaço de medidas de probabilidade. Ferguson (1973) introduziu o processo de Dirichlet (*Dirichlet process*, DP) como uma MPA. Admitir que F segue um DP, simbolicamente, $F \sim DP(\alpha, F_0)$, significa que para qualquer partição mensurável $A_1, \ldots, A_M$ do espaço amostral, o vetor de probabilidades $[F(A_1), \ldots, F(A_M)]$ segue uma distribuição de Dirichlet com vetor de parâmetros $[\alpha F_0(A_1), \ldots, \alpha F_0(A_M)]$, em que α é um parâmetro de precisão e F_0 é uma distribuição de referência sobre o espaço amostral mensurável. Tendo em conta a parametrização indicada, F_0 é a esperança *a priori* de F e à medida que α cresce há uma maior concentração *a priori* de F em torno de F_0, até o caso extremo de que $\alpha \longrightarrow \infty$ indica que se supõe F igual a F_0; por outro lado, valores pequenos de α (*e.g.*, < 5) permitem, em geral, que F se afaste consideravelmente de F_0 (Congdon, 2006, p.201). Dadas n observações independentes e identicamente distribuídas, a distribuição *a posteriori* é $F|(x_1, \ldots, x_n) \sim DP(\alpha + n, F_1)$, em que $F_1 = (\alpha F_0 + nF_n)/(\alpha + n)$ e F_n é a função de distribuição empírica das observações.

A simplicidade das propriedades do DP e a facilidade de obtenção da distribuição *a posteriori* tornam o uso deste modelo atrativo. Contudo, o DP gera quase certamente uma distribuição discreta, o que é uma propriedade indesejável em muitos casos. Uma maneira simples de remover essa restrição para gerar uma MPA compatível com distribuições absolutamente contínuas é a de admitir-se que

X_i segue uma distribuição absolutamente contínua dado o valor de um parâmetro específico e, por sua vez, que todos esses parâmetros sejam uma amostra aleatória de um DP, ou seja,

$$X_i|\theta_i \overset{\text{ind.}}{\sim} F_{\theta_i}, \quad i = 1, \ldots, n, \tag{4.1}$$

$$(\theta_1, \ldots, \theta_n)|G \overset{\text{i.i.d.}}{\sim} G, \quad G|(\alpha, G_0) \sim DP(\alpha, G_0). \tag{4.2}$$

Com esta abordagem, uma maior flexibilidade na modelagem foi introduzida ao se supor que os parâmetros $\{\theta_i\}$ seguem uma distribuição *a priori* do tipo DP centrada em G_0, ao invés da abordagem usual de supor diretamente que estes parâmetros seguem uma distribuição paramétrica G_0 (Walker, Damien, Laud & Smith, 1999). A nomenclatura de "mistura **por** processo de Dirichlet" (*Dirichlet process mixture*, DPM) se deve ao fato de a formulação hierárquica (4.1)-(4.2) implicar que se está admitindo que a distribuição marginal para X_i é uma mistura, *i.e.*,

$$f(x_i) = \int f(x_i|\theta_i) dG(\theta_i), \quad G|(\alpha, G_0) \sim DP(\alpha, G_0). \tag{4.3}$$

Não se deve confundir esse modelo com o de "mistura **de** processos de Dirichlet" (*mixture of Dirichlet processes*, MDP), sugerido por Antoniak (1974), pelo fato de esse ser uma mistura paramétrica induzida por distribuições impostas sobre parâmetros do DP, assim, encarados como hiperparâmetros. Da mesma forma que o DP, a MDP gera quase certamente uma distribuição discreta, enquanto que a DPM, considerando ou não distribuições *a priori* para parâmetros do DP, produz distribuições absolutamente contínuas.

A definição construtiva do DP, apresentada por Sethuraman (1994), mostra que $G|(\alpha, G_0) \sim DP(\alpha, G_0)$ pode ser representado por

$$G(A) = \sum_{j=1}^{\infty} p_j \delta_{\theta_j}(A) \tag{4.4}$$

para qualquer subconjunto mensurável A do espaço dos valores dos $\{\theta_j\}$, em que

$$p_1 = V_1, \quad p_j = V_j \prod_{k=1}^{j-1} (1 - V_k), j > 1, \quad V_j \overset{\text{i.i.d.}}{\sim} Beta(1, \alpha), \ j = 1, 2, \ldots, \tag{4.5}$$

$\delta_{\theta_j}(A)$ é a medida de Dirac, isto é, tem massa igual a um se $\theta_j \in A$ e tem massa nula em caso contrário, e

$$\theta_j \overset{\text{i.i.d.}}{\sim} G_0, \ j = 1, 2, \ldots \tag{4.6}$$

(Walker *et al.*, 1999). Esse resultado é muito útil, pois permite delinear algoritmos eficientes para o

ajuste do modelo DPM, uma vez que possibilita reescrever (4.3) como

$$f(x_i) = \sum_{j=1}^{\infty} p_j f(x_i|\theta_j), \quad i = 1, \ldots, n. \tag{4.7}$$

O procedimento (4.5) de construção dos pesos aleatórios $\{p_j\}$ da mistura é chamado de quebra-vara (*stick-breaking*). Note-se que $\sum_{j=1}^{\infty} p_j = 1$.

Na prática, contudo, por simplicidade, costuma-se truncar a mistura (4.7) em M componentes (veja, *e.g.*, Ishwaran & James, 2002), o que equivale a aproximar o $DP(\alpha, G_0)$ por um processo de Dirichlet truncado (TDP), denotado por $TDP(\alpha, G_0, M)$. Neste caso, para obter os pesos $p_1, \ldots, p_M$, geram-se variáveis $V_j \sim Beta(1, \alpha)$, $j = 1, \ldots, M-1$, e fixa-se $V_M = 1$.

A escolha de M é uma questão primordial na abordagem por distribuição *a priori* de TPD. Primeiramente, pode-se recorrer ao limite $M = n$, pois no máximo ter-se-ia o caso em que cada unidade x_i da amostra poderia estar associada a um θ_j diferente em (4.7). Em segundo lugar, como muitas vezes o valor exato da variável contínua acaba arredondado pelo instrumento de medição ou pelo processo observacional, o número de valores distintos presentes na amostra costuma ser muito inferior a n e, então, por raciocínio análogo ao anterior, não faz sentido supor que M exceda esse número. No estudo de embolia pulmonar, por exemplo, dos 987 valores observados de $PaCO_2$ há 243 valores distintos. Por fim, mesmo que todos os valores sejam distintos, Antoniak (1974) mostra que (i) o DP naturalmente propicia um agrupamento dos $\{\theta_i\}$, supondo que observações muito próximas correspondam a um mesmo valor de θ_j, (ii) uma distribuição *a priori* para α induz uma distribuição *a priori* para o número de valores distintos de $\{\theta_i\}$, denotado por M^*, e (iii) para n grande,

$$\mathrm{E}(M^*|\alpha) \cong \alpha \ln\left(1 + \frac{n}{\alpha}\right), \tag{4.8}$$

Na Tabela 4.1, apresentam-se alguns valores para $\mathrm{E}(M^*|\alpha)$, variando α e n na expressão (4.8). A dependência entre M^* e α é patente em (4.5) uma vez que à medida que o valor de α decresce, as distribuições $\{V_j\}$ concentram-se mais em valores distantes de zero e, por conseguinte, tende-se a ter uma menor quantidade de pesos que não são tão próximos de zero.

Tabela 4.1: Número médio de grupos distintos, $\mathrm{E}(M^*|\alpha)$, obtido com (4.8), variando α e n.

$\alpha \backslash n$	1 000	10 000	30 000	50 000
2	12.4	17.0	19.2	20.3
3	17.4	24.3	27.6	29.2
5	26.5	38.0	43.5	46.1
8	38.7	57.1	65.8	69.9
10	46.2	69.1	80.1	85.2

West (1992) e Escobar & West (1995) apresentam análises fixando $\alpha = 1$, mas este último

trabalho bem como os de Escobar & West (1998) e Ishwaran & James (2002), dentre outros, sugerem a utilização da distribuição *a priori*

$$\alpha|(\lambda_1, \lambda_2) \sim Gama(\lambda_1, \lambda_2), \tag{4.9}$$

em que $Gama(\lambda_1, \lambda_2)$ denota uma distribuição gama com parâmetro de forma λ_1 e parâmetro de escala λ_2, de tal maneira que a média é λ_1/λ_2. Seguindo a sugestão de Ishwaran & James (2002), em todas as análises deste capítulo utilizam-se os hiperparâmetros $\lambda_1 = \lambda_2 = 2$, que concentra a larga maioria (da ordem de 98%) dos valores de α entre 0 a 3, permitindo assim que G se afaste consideravelmente de G_0.

A distribuição *a posteriori* de M^* pode ser utilizada para avaliar o truncamento do DP, *i.e.*, se foi realizado com um valor muito baixo para M. Assim, se o quantil 97.5% *a posteriori* de M^* estiver muito próximo de M, é razoável que se aumente o valor de M; por outro lado, se o quantil de M^* estiver muito distante de M, pode-se diminuir o valor de M sem prejudicar a aproximação do DP pelo TDP e ainda obter resultados com velocidade maior, uma vez que, quanto maior o valor de M, maior é o esforço computacional. A distribuição *a posteriori* de α e a expressão (4.8) podem auxiliar a escolha do novo valor. Na análise da variável $PaCO_2$ do estudo de embolia pulmonar, por exemplo, o quantil 97.5% da distribuição *a posteriori* de M^* foi igual a 11, um valor muito menor do que o $M = 20$ empregado; como o quantil 97.5% da distribuição *a posteriori* de α foi um pouco menor do que 2, utilizando os resultados da Tabela 4.1 pode-se reduzir M para um valor próximo de 13. Não é por acaso que a distribuição *a posteriori* de α não se afasta muito da distribuição *a priori*, mesmo com esses tamanhos de amostra consideráveis; Leonard (1996) discute possíveis problemas de identificabilidade de α, principalmente se os valores observados não contiverem empates.

O caso mais frequente na literatura de DPM é o de se supor uma distribuição normal[3] para F. West (1992) explora conexões entre o caso normal via DPM e técnicas de estimação de densidades pelo método do núcleo gaussiano (Silverman, 1986). Escobar & West (1995) apresentam um dos primeiros trabalhos que possibilitam a implementação computacional de um amostrador de Gibbs para DPM e, por isso, popularizaram a abordagem. Contudo, nesses trabalhos iniciais, os esquemas de amostragem via método de Monte Carlo baseado em cadeia de Markov (*Markov chain Monte Carlo*, MCMC) integram o DP, utilizando a sua representação na forma de urna de Pólya (Blackwell & MacQueen, 1973), uma vez que não é possível sortear valores exatos do DP. A estratégia seguida neste trabalho não integra o DP, aproximando-o pelo TDP, conforme discutido por Ishwaran & James (2002) e Congdon (2006, pp.201-207). Essa é uma opção pragmática, pois a versão de (4.7) truncada em M componentes pode ser facilmente implementada nos pacotes computacionais do projeto BUGS (*Bayesian inference Using Gibbs Sampling*), cujas principais versões são o WinBUGS e o OpenBUGS (Lunn *et al.*, 2000, 2009), este último, de código aberto, possibilitando que se obtenham amostras das

[3]Um modelo mais simples, com estrutura Poisson-gama, pode ser encontrado em Escobar & West (1998) e Congdon (2006, p.205), em que F_{θ_i} em (4.1) é $Poisson(\theta_i)$ e G_0 em (4.2) é $Gama(\lambda_1, \lambda_2)$.

distribuições *a posteriori* de interesse. Nessas abordagens computacionais, geralmente introduzem-se variáveis latentes $(s_1, \ldots, s_n)$ que indicam quais observações correspondem a um mesmo valor de θ_j, $j = 1, 2, \ldots$. Isso possibilita reescrever a versão truncada de (4.7) como o modelo hierárquico

$$\begin{aligned}
X_i|\theta_{s_i} &\overset{\text{ind.}}{\sim} F_{\theta_{s_i}}, \quad i = 1, \ldots, n, \\
P(s_i = j) &= p_j, \quad j = 1, \ldots, M, \quad i = 1, \ldots, n, \\
p_1 = V_1, &\quad p_j = V_j \prod_{k=1}^{j-1} (1 - V_k), \; j = 2, \ldots, M, \\
V_j &\overset{\text{i.i.d.}}{\sim} Beta(1, \alpha), \; j = 1, \ldots, M-1, \quad V_M = 1, \\
\theta_j &\overset{\text{i.i.d.}}{\sim} G_0, \quad j = 1, \ldots, M.
\end{aligned}$$

Ishwaran & James (2001) mostram que o amostrador de Gibbs tem as seguintes características.

1. Os elementos de $(\theta_1, \ldots, \theta_M)$ são condicionalmente independentes dadas as demais variáveis e a distribuição condicional completa de θ_j é proporcional a $g_0(\theta_j) \prod_{\{i:s_i=j\}} f(x_i|\theta_j)$, em que g_0 é a função de densidade da distribuição G_0.
2. Os elementos de $(V_1, \ldots, V_{M-1})$ são condicionalmente independentes dadas as demais variáveis e a distribuição condicional completa de V_j é $Beta(1 + a_j, \alpha + b_j)$, em que $a_j = \sum_{i=1}^n I(s_i = j)$, $b_j = \sum_{i=1}^n I(s_i > j)$ e $I()$ denota a função indicadora, assumindo o valor 1 quando a condição explicitada entre parênteses for satisfeita e resultando o valor 0 em caso contrário.
3. Os elementos de $(s_1, \ldots, s_n)$ são condicionalmente independentes dadas as demais variáveis e cada um dos s_i é atualizado a partir da distribuição $P(s_i = j) \propto p_j f(x_i|\theta_j)$, $j = 1, \ldots, M$.

Uma revisão dos avanços computacionais descrevendo esses e outros algoritmos para se fazer inferência em DPM é apresentada por Griffin & Holmes (2010).

Seguindo West (1992), considere-se a versão mais simples do modelo normal univariado

$$X_i|\mu_i, V \overset{\text{ind.}}{\sim} Normal(\mu_i, V), \quad i = 1, \ldots, n, \tag{4.10}$$

$$(\mu_1, \ldots, \mu_n)|G \overset{\text{i.i.d.}}{\sim} G, \quad G|(\alpha, G_0) \sim DP(\alpha, G_0), \tag{4.11}$$

em que $Normal(\mu_i, V)$ denota a distribuição normal com média μ_i e variância V. West (1992) admite que G_0 é $Normal(\mu_0, \tau V)$, o que se denota aqui por

$$G_0|(\mu_0, \tau, V) = Normal(\mu_0, \tau V), \tag{4.12}$$

fixando um valor para μ_0 e adotando as distribuições *a priori*

$$V^{-1}|(s_0, S_0) \sim Gama(s_0/2, S_0/2), \tag{4.13}$$

$$\tau^{-1}|(w, W) \sim Gama(w/2, W/2), \tag{4.14}$$

em que S_0/s_0 é o palpite *a priori* para V e s_0 mede a crença *a priori* neste palpite. West (1992) não menciona como escolher w e W, mas numa análise utiliza $w = 2$ e $W = 10$, dizendo resultar numa distribuição *a priori* aproximadamente difusa. Além disso, alerta que em geral não há muita informação sobre τ na amostra, mas que a distribuição *a priori* especificada para τ é necessária da mesma forma que a escolha subjetiva do parâmetro de suavização nos métodos tradicionais de núcleo. Por fim, diz que valores grandes de τ favorecem um número maior de modas para a distribuição preditiva *a posteriori* de X. Escobar & West (1995) consideram o caso heterocedástico, *i.e.*, em que cada X_i em (4.10) pode ter uma variância diferente, V_i, e adotam a distribuição *a priori*

$$\mu_0|(a, A) \sim Normal(a, A), \tag{4.15}$$

com $A^{-1} \longrightarrow 0$. No exemplo que analisam, primeiro avaliam a distribuição do número de modas induzida sob diferentes valores de τ e, em seguida, utilizam os hiperparâmetros $w = 1$ e $W = 100$ que consideram compatíveis com suas crenças.

Ishwaran & James (2002) também contemplam os casos de variâncias heterogêneas e homogêneas. No último, ao invés de (4.12), supõem

$$G_0|(\mu_0, \tau) = Normal(\mu_0, \tau), \tag{4.16}$$

ou seja, que a variância da distribuição de referência de μ_i independe, *a priori*, da variância de X_i. Ao invés de utilizar a distribuição *a priori* (4.14), sugerem que se fixe o valor de τ de forma que (4.16) cubra valores que se acredita que $\{\mu_i\}$ podem ter, sugerindo que uma boa escolha é igualar $\sqrt{\tau}$ a quatro vezes o desvio padrão dos dados. Com relação à distribuição *a priori* para V, após comparar os resultados de (4.13), utilizando valores pequenos para s_0 e S_0 (*e.g.*, $s_0 = S_0 = 0.02$), com os obtidos sob

$$V|T \sim Unif[0, T], \tag{4.17}$$

em que $Unif[0, T]$ denota uma distribuição uniforme contínua no intervalo $[0, T]$, tomando T igual à variância dos dados, concluem, para o caso heterocedástico, que a distribuição *a priori* gama pode ser informativa em certas situações, mesmo com hiperparâmetros considerados não-informativos, o que acaba suavizando os dados de forma inadequada (pouco ou muito), e, assim, sugerem que a distribuição *a priori* uniforme pode ser mais interessante.

Devido à dificuldade de se escolherem os hiperparâmetros de (4.14) ao utilizar a distribuição de referência (4.12) e pela possibilidade de (4.13) ser mais informativa do que se imagina para a distribuição de referência (4.16), utilizam-se neste trabalho as sugestões de Ishwaran & James (2002) com a distribuição *a priori* uniforme para V. Não se consideram aqui versões heterocedásticas para evitar uma sobreparametrização ainda maior, uma vez que além de a amostra não conter muita

informação sobre α, os modelos informativos mais gerais para o mecanismo de omissão já sofrem de problemas de identificabilidade.

4.3 Modelo semiparamétrico para respostas binárias com uma variável explicativa sujeita a omissão informativa

Sejam Y_i uma resposta binária sempre observada, X_i uma covariável contínua potencialmente omissa e R_i o indicador de observação de X_i, ou seja, $R_i = 1$ se X_i é observada e $R_i = 0$ se X_i é omissa, $i = 1, \ldots, n$. Apesar de o interesse recair apenas na distribuição condicional de Y_i dado X_i, é necessário considerar um modelo para X_i, uma vez que não se quer desprezar a parcela da amostra para a qual X_i está omissa. Ao admitir que o mecanismo que gera os dados faltantes pode depender dos valores não observados, deve-se também propor um modelo para R_i.

Utilizando a fatoração de modelos de seleção, propõe-se o modelo

$$R_i|(Y_i, X_i, \delta_0, \delta_1, \delta_2, \delta_3) \overset{\text{ind.}}{\sim} Bern(\theta_i),\ \text{logito}(\theta_i) = \delta_0 + \delta_1 X_i + \delta_2 Y_i + \delta_3 X_i Y_i, \quad i = 1, \ldots, n, \tag{4.18}$$

$$Y_i|(X_i, \beta_0, \beta_1) \overset{\text{ind.}}{\sim} Bern(\pi_i),\ \text{logito}(\pi_i) = \beta_0 + \beta_1 X_i, \quad i = 1, \ldots, n, \tag{4.19}$$

$$X_i|(\mu_i, V) \overset{\text{ind.}}{\sim} Normal(\mu_i, V), \quad i = 1, \ldots, n, \tag{4.20}$$

em que $Bern(\theta_i)$ denota a distribuição de Bernoulli com probabilidade de sucesso θ_i, e sugerem-se as distribuições *a priori*

$$\delta_j|(\mu_{\delta_j}, \sigma_{\delta_j}) \overset{\text{ind.}}{\sim} Normal(\mu_{\delta_j}, \sigma_{\delta_j}), \quad j = 0, 1, 2, 3, \tag{4.21}$$

$$\beta_j|(\mu_{\beta_j}, \sigma_{\beta_j}) \overset{\text{ind.}}{\sim} Normal(\mu_{\beta_j}, \sigma_{\beta_j}), \quad j = 0, 1, \tag{4.22}$$

$$(\mu_1, \ldots, \mu_n)|G \overset{\text{i.i.d.}}{\sim} G, \quad G|\alpha, G_0, M \sim TDP(\alpha, G_0, M), \tag{4.23}$$

$$V|T \sim Unif[0, T], \tag{4.24}$$

$$\alpha|(a, b) \sim Gama(\lambda_1, \lambda_2), \tag{4.25}$$

$$G_0|(\mu_0, \tau) = Normal(\mu_0, \tau), \tag{4.26}$$

$$\mu_0|(a, A) \sim Normal(a, A), \tag{4.27}$$

todas mutuamente independentes.

Os valores dos hiperparâmetros fixos dessas distribuições *a priori* serão indicados nas aplicações das próximas seções. Note que o modelo (4.18)-(4.20) não permite por si só obter a verossimilhança dos dados observados, mas apenas a verossimilhança dos dados completos, que é justamente o que os pacotes computacionais do projeto BUGS necessitam. Nesses pacotes, os dados em falta são indicados como "NA" e são tratados como variáveis latentes. Assim, em uma das etapas do esquema de amostragem via método MCMC, o algoritmo sorteará um valor para o dado faltante da sua distribuição condicional dadas as demais variáveis (observadas e não-observadas). Com relação ao

DP, em cada iteração do MCMC o algoritmo sorteará (1) μ_0, (2) μ_j, $j = 1, \ldots, M$, *i.e.*, os M valores distintos de $\{\mu_i\}$, (3) α, (4) V, (5) V_j, $j = 1, \ldots, M-1$ (para obter $p_1, \ldots, p_M$) e (6) quais dos μ_j, $j = 1, \ldots, M$ serão alocados para cada um dos μ_i, $i = 1, \ldots, n$.

Diz-se que o modelo é semiparamétrico, uma vez que emprega a abordagem não-paramétrica da seção anterior para o modelo marginal de X_i e modelos paramétricos convencionais para as distribuições condicionais de Y_i dado X_i e de R_i dado Y_i e X_i.

O mecanismo de omissão (4.18) é informativo, pois considera que a probabilidade de se ter covariáveis em falta pode depender de seus valores não observados. Por outro lado, se for acrescentada a restrição

$$\text{MAR} : \delta_1 = \delta_3 = 0, \tag{4.28}$$

o mecanismo de omissão passa a ser não-informativo e ignorável sob o ponto de vista de inferências bayesianas para β_0 e β_1, uma vez que se supõe independência *a priori* entre (δ_0, δ_2) e os demais parâmetros. Uma subclasse do modelo MAR é o mecanismo de omissão completamente aleatória que pode ser imposto pela restrição

$$\text{MCAR} : \delta_1 = \delta_2 = \delta_3 = 0. \tag{4.29}$$

É importante destacar que no caso em foco, de omissão em variáveis explicativas, não é só sob o mecanismo MCAR que, em geral, a análise de casos completos (ACC) produz inferências não-enviesadas para β_0 e β_1, mas também sob qualquer mecanismo que não dependa da resposta Y_i, como

$$\text{MNAR}_{\text{red}} : \delta_2 = \delta_3 = 0. \tag{4.30}$$

A ACC de dados gerados sob o mecanismo de omissão informativa (4.30) resulta em inferências enviesadas para a distribuição marginal de X_i, mas não para a distribuição condicional de Y_i dado X_i. Contudo, na ACC não é necessário propor um modelo para X_i se o interesse for apenas na distribuição condicional de Y_i dado X_i.

4.4 Estudo de simulação

Consideram-se as seguintes distribuições para a variável explicativa

$$X^N \sim Normal(12, 3^2), \tag{4.31}$$

$$X^L \sim Log{-}normal(2.45, 0.246^2), \tag{4.32}$$

$$X^C = 0.8 \times X^{C1} + 0.2 \times X^{C2},\ X^{C1} \sim Unif[8, 12],\ X^{C2} \sim Log{-}normal(2.79, 0.642^2), \tag{4.33}$$

em que $Log{-}normal(\mu, \sigma^2)$ denota a distribuição log-normal e μ e σ são, respectivamente, a média e o desvio padrão da variável na escala logarítmica. A média e o desvio padrão de X^L e X^C coincidem com os parâmetros correspondentes de X^N, embora suas densidades sejam bem diferentes, conforme pode-se verificar na Figura 4.2.

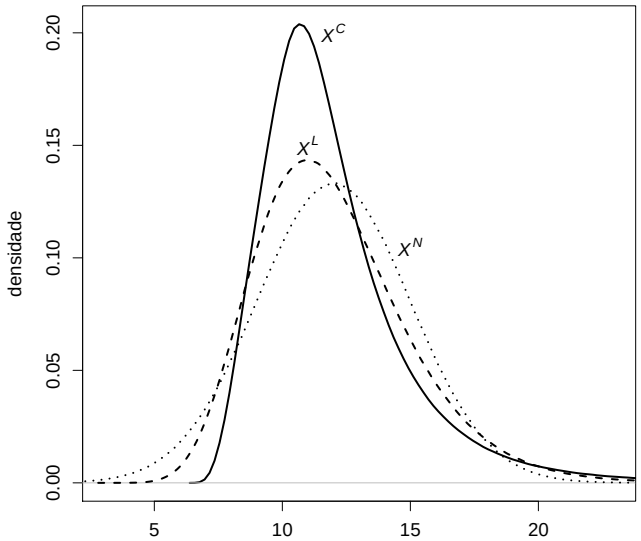


Figura 4.2: Densidades das distribuições normal (X^N), log-normal (X^L) e combinação linear (X^C) de uma uniforme e uma log-normal.

Com o objetivo de avaliar o impacto de resultados obtidos sob diferentes suposições distribucionais para a covariável, gera-se uma amostra de X de tamanho $n = 10\,000$ para cada uma das 3 distribuições (4.31), (4.32) e (4.33); em seguida, para cada valor gerado sob cada uma das covariáveis, gera-se Y sob (4.19) com $\beta_0 = 6$ e $\beta_1 = -0.5$ e, por fim, gera-se R sob (4.18) com $\delta_0 = -3$, $\delta_1 = 0.5$ e $\delta_2 = \delta_3 = 0$; posteriormente, ajusta-se o modelo semiparamétrico da seção anterior e os modelos paramétricos normal e log-normal para os conjuntos de dados gerados a partir de cada uma das covariáveis X^N, X^L e X^C. Nos casos paramétricos normal e log-normal, o modelo não-paramétrico (4.20) é substituído, respectivamente, por

$$X_i|\mu_0, \tau \overset{\text{i.i.d.}}{\sim} Normal(\mu_0, \tau), \quad i = 1, \ldots, n, \tag{4.34}$$

$$X_i|\mu_0, \tau \overset{\text{i.i.d.}}{\sim} Log{-}normal(\mu_0, \tau), \quad i = 1, \ldots, n, \tag{4.35}$$

e, em ambos, as distribuições *a priori* (4.23)-(4.26) são substituídas por (4.14), com escolhas para os hiperparâmetros associadas a distribuições consideradas vagas, *i.e.*, A grande (10^6 e 10^3 para, respectivamente, os casos normal e log-normal) e w pequeno (2 para ambos os casos). No caso do modelo semiparamétrico, os hiperparâmetros de (4.23)-(4.27) foram escolhidos conforme descrito na Seção 4.2. Em todos os modelos, adotaram-se distribuições *a priori* vagas para δ_j e β_j, utilizando os hiperparâmetros $\mu_{\delta_j} = \mu_{\beta_j} = 0$ e $\sigma_{\delta_j} = \sigma_{\beta_j} = 10^3$, $j = 0, 1$. Além disso, supôs-se sempre a estrutura correta para o mecanismo de omissão, *i.e.*, $\delta_2 = \delta_3 = 0$, para que as únicas variantes sejam

a distribuição utilizada para gerar a covariável e a distribuição adotada para a covariável na análise. Na Tabela 4.2, apresentam-se resumos das distribuições *a posteriori* de β_0 e β_1. Em todas as análises, as magnitudes dos erros de Monte Carlo foram menores do que a precisão dos números apresentados nas tabelas. Métodos usuais de diagnóstico foram utilizados para avaliar a convergência das cadeias de Markov geradas relativas a β_0 e β_1 (Heidelberger & Welch, 1983; Gelman & Rubin, 1992; Geweke, 1992; Raftery & Lewis, 1992). Não houve razões para suspeitar da convergência das cadeias obtidas nas análises desta seção e da seguinte. Contudo, algumas dessas análises chegaram a durar vários dias porque foi necessário espaçar as cadeias em até 500 valores e optou-se por gerar 10 cadeias, cada uma com até 5 000 valores após o espaçamento.

Na Tabela 4.2, nota-se que os resumos das amostras obtidas das distribuições *a posteriori* de ambos os parâmetros são muito próximos quando se comparam os resultados obtidos com o modelo não-paramétrico para a covariável e os correspondentes modelos paramétricos verdadeiros, quer com a distribuição normal, quer com a log-normal; nestes casos, os intervalos de credibilidade contêm os verdadeiros valores, o que não ocorre nas análises sob modelos paramétricos incorretos para X^N e X^L. Por outro lado, no caso de X^C, apenas os intervalos de credibilidade da análise com o modelo não-paramétrico para a covariável contiveram os verdadeiros valores de β_0 e β_1.

Um dos motivos para se considerarem modelos MNAR sob a restrição (4.30) é que, neste caso, como a ACC não leva a inferências enviesadas para β_0 e β_1, pode-se utilizar os correspondentes resultados para comparar com os obtidos sob os modelos propostos, que adicionalmente incluem um modelo para a omissão e outro para a covariável. Os resultados obtidos com a ACC (Tabela 4.3) são muito próximos dos obtidos na Tabela 4.2 com os modelos paramétricos corretos e com o modelo não-paramétrico para a covariável. Isso é um indício, por exemplo, de que o fato de os valores verdadeiros de β_0 e β_1 estarem próximos, respectivamente, dos limites superior e inferior das análises de X^C com o modelo não-paramétrico para a covariável deve ser uma consequência dos valores gerados para esse único conjunto de dados que foi analisado e não é uma deficiência do modelo e/ou método inferencial, uma vez que um resultado similar foi obtido na ACC. As semelhanças supracitadas entre os resultados de referência e a metodologia proposta sinalizam que o modelo não-paramétrico parece ser suficientemente flexível para acomodar uma certa variedade de distribuições para a covariável e, desta forma, evitar que se obtenham vieses nas inferências para o modelo condicional de interesse quando se adota um modelo paramétrico incorreto para a covariável. Contudo, tais conclusões devem ressalvar as devidas proporções, uma vez que como não foi viável realizar um estudo de Monte Carlo, por limitação de tempo, optou-se por analisar um único conjunto de dados com amostra grande e, assim, não se sabe como seriam os resultados de tais análises repetidas para conjuntos de dados variados.

A outra razão para se considerar a restrição (4.30) é que como ela é suficiente para identificar o modelo, evita-se num primeiro momento, lidar com possíveis problemas causados por inidentificabilidade. Com o objetivo de se explorar esta questão, repete-se o estudo de simulação utilizando

Tabela 4.2: Médias, desvios padrões (DP) e intervalos de credibilidade (IC) centrais a 95% *a posteriori* para β_0 e β_1 das análises de todos os dados disponíveis.

Distr. da Covariável		β_0			β_1		
Gerada	Suposta	Média	DP	IC 95%	Média	DP	IC 95%
X^N	Normal	6.22	0.15	[5.93; 6.51]	−0.515	0.012	[−0.538; −0.491]
	Log-normal	6.37	0.14	[6.09; 6.66]	−0.525	0.012	[−0.549; −0.502]
	Não-Param.	6.21	0.15	[5.92; 6.51]	−0.514	0.012	[−0.538; −0.491]
X^L	Normal	5.06	0.13	[4.82; 5.32]	−0.428	0.011	[−0.449; −0.407]
	Log-normal	6.01	0.14	[5.73; 6.29]	−0.501	0.012	[−0.525; −0.478]
	Não-Param.	6.00	0.14	[5.71; 6.28]	−0.500	0.012	[−0.524; −0.477]
X^C	Normal	4.72	0.12	[4.49; 4.96]	−0.395	0.010	[−0.416; −0.375]
	Log-normal	5.08	0.13	[4.83; 5.34]	−0.425	0.011	[−0.447; −0.404]
	Não-Param.	5.78	0.15	[5.49; 6.08]	−0.481	0.013	[−0.505; −0.456]
Valores Verdadeiros		$\beta_0 = 6.00$			$\beta_1 = -0.500$		

Tabela 4.3: Médias, desvios padrões (DP) e intervalos de credibilidade (IC) centrais a 95% *a posteriori* para β_0 e β_1 das análises de casos completos.

Distribuição Gerada	β_0			β_1		
para a Covariável	Média	DP	IC 95%	Média	DP	IC 95%
X^N	6.22	0.15	[5.93; 6.52]	−0.515	0.012	[−0.539; −0.492]
X^L	6.02	0.14	[5.74; 6.30]	−0.502	0.012	[−0.525; −0.479]
X^C	5.83	0.15	[5.54; 6.12]	−0.484	0.012	[−0.509; −0.460]
Valores Verdadeiros	$\beta_0 = 6.00$			$\beta_1 = -0.500$		

$\delta_0 = -6$, $\delta_1 = 0.5$, $\delta_2 = 1$ e $\delta_3 = 0.5$ para gerar valores para R e considerando o modelo (4.18) sem aplicar as restrições utilizadas anteriormente. Nos ajustes de modelos inidentificáveis em análises de dados categorizados com respostas incompletas, notou-se (veja Capítulo 3) que para se constatar a convergência foram necessárias cadeias muito maiores do que as que seriam requeridas em ajustes de modelos identificáveis, tais como de um modelo MAR; observou-se também que o cenário agrava-se ainda mais à medida que os tamanhos de amostra ficam maiores e que se consideram distribuições *a priori* mais vagas. Sendo assim, optou-se por, primeiramente, diminuir o tamanho de amostra para $n = 1\,000$, o que também possibilita a utilização de valores menores para M, reduzindo sensivelmente o tempo computacional, e, em segundo lugar, utilizar distribuições *a priori* mais concentradas para δ_j, $j = 1, 2, 3$, *i.e.*, com $\sigma_{\delta_j} = 1$. Vale a pena destacar que o único trabalho encontrado que considerou processos de Dirichlet na análise de dados contínuos incompletos, embora no contexto de MDP e não DPM, foi o de Scharfstein *et al.* (2003), em que os autores exploram duas vertentes de análises: ora utilizando valores altos para o parâmetro de precisão do DP ($\alpha = 10\,000$), o que na prática equivale a se adotar a distribuição paramétrica de referência, com distribuições *a priori* vagas para os demais parâmetros, ora empregando um valor baixo para o parâmetro de precisão ($\alpha = 1$) e uma distribuição *a priori* ainda mais informativa do que as aqui adotadas para o parâmetro inidenti-

ficável[4] do mecanismo de omissão (*i.e.*, com média da distribuição *a priori* normal diferente de zero e desvio padrão igual a 0.25). Além disso, embora eles sugiram que se use a abordagem de Gelman & Rubin (1992) para avaliar a convergência, não mencionam ter aplicado tal critério em suas análises e, pelo contrário, empregam uma análise informal, verificando visualmente se as estimativas para as densidades *a posteriori* dos parâmetros de interesse para diferentes valores fixados do parâmetro inidentificável parecem estar próximas das esperadas sob distribuições normais. Embora esses autores não relacionem as escolhas das distribuições *a priori* com a falta de identificabilidade do modelo e/ou com a dificuldade de se avaliar a convergência das cadeias, as escolhas que fizeram estão em linha com os comentários antecipados neste parágrafo. Para ter uma ideia da sensibilidade da escolha da média dessas distribuições *a priori* mais concentradas, nas análises a seguir, considerou-se primeiro $\mu_{\delta_j} = 0$ e depois $\mu_{\delta_j} = 1$, para $j = 1, 2, 3$. Com a intenção de contemplar casos em que o palpite *a priori* não está muito equivocado, mas também pode não acertar em cheio no valor verdadeiro dos parâmetros, note-se que para ambas as distribuições *a priori* os valores verdadeiros de δ_j, $j = 1, 2, 3$, estão contidos em regiões com valores das densidades *a priori* não muito baixos, embora os valores verdadeiros dos parâmetros sejam em geral diferentes das médias *a priori*. As demais escolhas dos hiperparâmetros foram realizadas conforme descrito nos parágrafos anteriores.

Ao confrontar as Tabelas 4.4 e 4.5, observa-se que apesar de os resumos *a posteriori* dependerem das distribuições *a priori*, esperavam-se maiores variações dos resultados; elas foram mais acentuadas nos casos em que os desvios padrões *a posteriori* são maiores. Os resultados obtidos com o modelo não-paramétrico para a covariável foram próximos dos correspondentes ao modelo paramétrico verdadeiro no caso da distribuição X^N, mas um pouco mais distantes no caso da distribuição X^L. Para esta distribuição, estranha-se também o fato de os desvios padrões *a posteriori* obtidos com o modelo não-paramétrico serem menores do que os obtidos com os modelos paramétricos, contradizendo o esperado de que há uma certa perda de eficiência para compensar o ganho de flexibilidade dos modelos não-paramétricos. Assim como antes, os intervalos de credibilidade contêm os valores verdadeiros dos parâmetros β_0 e β_1 apenas nos casos em que se especificaram modelos paramétricos corretamente ou se empregaram modelos não-paramétricos para a covariável. Ressalta-se mais uma vez que embora os resultados sejam muito específicos de análises baseadas num único conjunto de dados para cada caso, eles sinalizam que o modelo não-paramétrico pode ser uma alternativa viável para os casos em que não há informação sobre a forma distribucional da covariável. Por fim, a Tabela 4.6 ilustra os resultados equivocados que seriam obtidos com ACC; as diferenças entre as médias *a posteriori* e os valores verdadeiros dos parâmetros ocorrem nas mesmas direções das obtidas nas análises em que se supuseram distribuições paramétricas incorretas para as covariáveis.

[4]Como o mecanismo MAR gera um modelo identificável, o modelo (4.18)-(4.20) é condicionalmente identificável dado que valores para (δ_1, δ_3) são fixados. Contudo, similarmente ao que foi visto no Capítulo 3, o modelo também é condicionalmente identificável dado que se fixem valores para outros 2 parâmetros como (δ_2, δ_3).

Tabela 4.4: Médias, desvios padrões (DP) e intervalos de credibilidade (IC) centrais a 95% *a posteriori* para β_0 e β_1 das análises de todos os dados disponíveis com hiperparâmetros $\mu_{\delta_j} = 0$, $j = 1, 2, 3$.

Distr. da Covariável		β_0			β_1		
Gerada	Suposta	Média	DP	IC 95%	Média	DP	IC 95%
X^N	Normal	6.94	1.00	[5.14; 9.07]	−0.566	0.076	[−0.725; −0.426]
	Log-normal	9.37	0.68	[8.08; 10.74]	−0.707	0.053	[−0.816; −0.607]
	Não-Param.	6.97	1.04	[5.12; 9.17]	−0.567	0.079	[−0.732; −0.424]
X^L	Normal	8.87	0.84	[7.23; 10.53]	−0.720	0.066	[−0.850; −0.592]
	Log-normal	6.55	0.94	[4.99; 8.66]	−0.543	0.074	[−0.709; −0.417]
	Não-Param.	5.90	0.77	[4.58; 7.58]	−0.493	0.062	[−0.629; −0.384]
X^C	Normal	9.47	0.69	[8.15; 10.85]	−0.768	0.056	[−0.879; −0.662]
	Log-normal	8.63	0.73	[7.21; 10.08]	−0.707	0.059	[−0.823; −0.593]
	Não-Param.	8.55	1.35	[5.41; 10.51]	−0.680	0.095	[−0.828; −0.457]
Valores Verdadeiros		$\beta_0 = 6.00$			$\beta_1 = -0.500$		

Tabela 4.5: Médias, desvios padrões (DP) e intervalos de credibilidade (IC) centrais a 95% *a posteriori* para β_0 e β_1 das análises de todos os dados disponíveis com hiperparâmetros $\mu_{\delta_j} = 1$, $j = 1, 2, 3$.

Distr. da Covariável		β_0			β_1		
Gerada	Suposta	Média	DP	IC 95%	Média	DP	IC 95%
X^N	Normal	6.73	1.00	[4.96; 8.92]	−0.550	0.077	[−0.715; −0.412]
	Log-normal	9.33	0.68	[8.07; 10.71]	−0.705	0.053	[−0.813; −0.606]
	Não-Param.	6.78	1.08	[4.90; 9.18]	−0.553	0.082	[−0.732; −0.408]
X^L	Normal	8.73	0.86	[7.07; 10.43]	−0.711	0.067	[−0.842; −0.581]
	Log-normal	6.35	0.88	[4.84; 8.24]	−0.528	0.070	[−0.677; −0.405]
	Não-Param.	5.72	0.72	[4.42; 7.23]	−0.479	0.059	[−0.600; −0.372]
X^C	Normal	9.46	0.69	[8.12; 10.83]	−0.767	0.056	[−0.878; −0.660]
	Log-normal	8.61	0.74	[7.19; 10.06]	−0.705	0.059	[−0.822; −0.592]
	Não-Param.	8.21	1.52	[5.08; 10.47]	−0.660	0.108	[−0.827; −0.428]
Valores Verdadeiros		$\beta_0 = 6.00$			$\beta_1 = -0.500$		

Tabela 4.6: Médias, desvios padrões (DP) e intervalos de credibilidade (IC) centrais a 95% *a posteriori* para β_0 e β_1 das análises de casos completos.

Distribuição Gerada	β_0			β_1		
para a Covariável	Média	DP	IC 95%	Média	DP	IC 95%
X^N	9.42	0.66	[8.16; 10.76]	−0.712	0.051	[−0.816; −0.614]
X^L	10.17	0.69	[8.86; 11.57]	−0.777	0.055	[−0.888; −0.673]
X^C	9.17	0.65	[7.92; 10.49]	−0.703	0.053	[−0.810; −0.601]
Valores Verdadeiros	$\beta_0 = 6.00$			$\beta_1 = -0.500$		

4.5 Análise dos dados de embolia pulmonar

Para analisar os dados da Seção 4.1 foram ajustados algumas dezenas de modelos, dentre os quais opta-se por apresentar os resultados para aquele especificado por

$$R_i|(\delta_0, \delta_1, \delta_2, LN_i, LP_i) \overset{\text{ind.}}{\sim} Bern(\theta_i),\ \text{logito}(\theta_i) = \delta_0 + \delta_1 LN_i + \delta_2 LP_i, \quad i = 1, \ldots, n, \tag{4.36}$$

$$Y_i|(\beta_0, \{X_{ji}, \beta_j, j = 1, \ldots, 7\}) \overset{\text{ind.}}{\sim} Bern(\pi_i),\ \text{logito}(\pi_i) = \beta_0 + \sum_{j=1}^{7} \beta_j X_{ji}, \quad i = 1, \ldots, n, \tag{4.37}$$

em que Y_i é o indicador de embolia pulmonar, $X_{1i}, \ldots, X_{7i}$ denotam as variáveis explicativas (i) indicador de realização de cirurgia recente, (ii) indicador de ocorrência anterior de embolia pulmonar ou trombose venosa profunda, (iii) indicador de atelectasia laminar no raio-X do tórax, (iv) indicador de elevação da hemicúpula diafragmática no raio-X do tórax, (v) idade, em dezenas de anos, (vi) pulsação arterial, em centenas de batimentos por minuto (bpm) e (vii) pressão parcial de gás carbônico ($PaCO_2$), em kPa, R_i é o indicador de observação de $PaCO_2$ (X_{7i}) e

$$LN_i = \begin{cases} LC_i, \text{ se } LC_i < 0, \\ 0, \text{ caso contrário}, \end{cases} \tag{4.38}$$

$$LP_i = \begin{cases} LC_i, \text{ se } LC_i > 0, \\ 0, \text{ caso contrário} \end{cases} \tag{4.39}$$

$$LC_i = \text{logito}(\pi_i) - \text{logito}(\hat{\pi}), \tag{4.40}$$

em que $\hat{\pi}$ é a estimativa da prevalência de embolia pulmonar do hospital suposta conhecida.

Wicki, Perneger, Junod, Bounameaux & Perrier (2001) mencionam que $PaCO_2$ estava em falta para alguns pacientes porque a gasometria arterial não foi realizada ou foi executada enquanto os pacientes estavam respirando oxigênio. Perrier (comunicação pessoal) afirma que isso ocorreu para pacientes que estavam muito pouco doentes ou tão doentes que necessitavam da administração de oxigênio. Contudo, não foi registrado em qual dos dois casos os pacientes com dados de $PaCO_2$ faltantes seriam classificados. Os comentários de Perrier sugerem que é razoável supor que a probabilidade de se observar $PaCO_2$ (θ_i) pode (i) ser máxima para pacientes com probabilidade de embolia pulmonar (π_i) próxima da prevalência de embolia pulmonar (π) e (ii) diminuir, à medida que a probabilidade de embolia pulmonar fica mais distante da prevalência. Tendo isto em vista, a regressão segmentada (4.36) permite que θ_i decaia com velocidade diferente à medida que $\pi_i \longrightarrow 0$ e $\pi_i \longrightarrow 1$ e, assim, espera-se que $\delta_1 > 0$ e $\delta_2 < 0$. O único paciente para o qual havia informações sobre o exame de raio-X do tórax omissas não foi utilizado nas análises.

Conforme as seções anteriores, além de (4.36) e (4.37), necessita-se também especificar um modelo para X_{7i}. Poder-se-ia adotar um modelo condicional para X_{7i} dadas as demais variáveis explicativas. Contudo, em análises preliminares, apenas a pulsação arterial e a ocorrência anterior de embolia

Tabela 4.7: Médias, desvios padrões (DP) e intervalos de credibilidade (IC) centrais a 95% *a posteriori* para as análises de todos os dados de embolia pulmonar.

	Modelo Não-Paramétrico			Modelo Normal		
Parâmetros	Média	DP	IC 95%	Média	DP	IC 95%
intercepto	−2.476	0.642	[−3.727; −1.192]	−2.506	0.641	[−3.780; −1.261]
cirurgia recente	1.375	0.268	[0.852; 1.903]	1.381	0.269	[0.856; 1.912]
embolia anterior	1.080	0.177	[0.735; 1.429]	1.081	0.177	[0.736; 1.430]
raio-X - EHD	0.590	0.189	[0.221; 0.962]	0.592	0.189	[0.221; 0.965]
raio-X - AL	0.706	0.187	[0.339; 1.070]	0.706	0.187	[0.337; 1.073]
idade (×10 anos)	0.268	0.046	[0.179; 0.359]	0.270	0.046	[0.183; 0.361]
pulsação (×100 bpm)	1.158	0.331	[0.508; 1.809]	1.171	0.327	[0.527; 1.811]
$PaCO_2$	−0.405	0.101	[−0.609; −0.209]	−0.404	0.100	[−0.600; −0.208]
	Modelo Log-normal			Modelo Gama		
Parâmetros	Média	DP	IC 95%	Média	DP	IC 95%
intercepto	−2.530	0.630	[−3.761; −1.301]	−2.480	0.651	[−3.728; −1.189]
cirurgia recente	1.380	0.269	[0.853; 1.912]	1.380	0.268	[0.854; 1.908]
embolia anterior	1.081	0.177	[0.735; 1.430]	1.081	0.177	[0.737; 1.431]
raio-X - EHD	0.590	0.189	[0.219; 0.960]	0.592	0.190	[0.223; 0.963]
raio-X - AL	0.706	0.186	[0.338; 1.071]	0.707	0.187	[0.342; 1.073]
idade (×10 anos)	0.270	0.046	[0.183; 0.362]	0.270	0.045	[0.182; 0.359]
pulsação (×100 bpm)	1.163	0.332	[0.511; 1.811]	1.166	0.330	[0.523; 1.807]
$PaCO_2$	−0.398	0.098	[−0.597; −0.207]	−0.409	0.103	[−0.614; −0.214]

EHD: elevação da hemicúpula diafragmática, AL: atelectasia laminar.

Tabela 4.8: Médias, desvios padrões (DP) e intervalos de credibilidade (IC) centrais a 95% *a posteriori* para a análise de casos completos dos dados de embolia pulmonar.

Parâmetros	Média	DP	IC 95%
intercepto	−2.585	0.672	[−3.901; −1.273]
cirurgia recente	1.512	0.293	[0.943; 2.086]
embolia anterior	1.087	0.187	[0.724; 1.453]
raio-X - EHD	0.591	0.201	[0.198; 0.990]
raio-X - AL	0.732	0.200	[0.343; 1.126]
idade (10 anos)	0.288	0.048	[0.195; 0.384]
pulsação (100 bpm)	1.221	0.352	[0.538; 1.914]
$PaCO_2$	−0.429	0.102	[−0.631; −0.231]

EHD: elevação da hemicúpula diafragmática, AL: atelectasia laminar.

pulmonar ou trombose venosa profunda ajudaram a explicar a variabilidade de $PaCO_2$, mas, mesmo assim, muito fracamente, uma vez que não se observaram associações não-lineares nas análises descritivas (*i.e.*, gráficos de dispersão) e obteve-se um coeficiente de determinação de apenas 1% no ajuste de um modelo linear. Logo, adotaram-se os modelos marginais em vez de condicionais para X_{7i}. Foram considerados os modelos paramétricos normal, log-normal e gama além do modelo não-paramétrico. Empregaram-se distribuições *a priori* conforme descrito nas últimas seções; as médias e

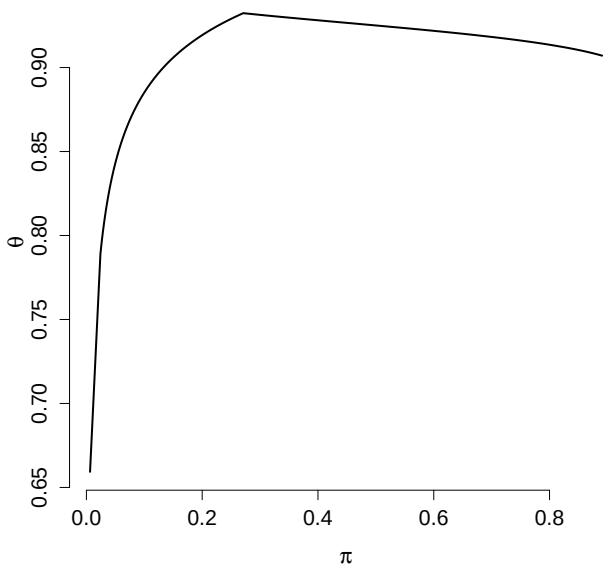


Figura 4.3: Estimativas de π_i (probabilidade de embolia pulmonar), obtidas a partir das médias *a posteriori* de $\{\beta_j\}$, e das estimativas de θ_i (probabilidade de se observar $PaCO_2$), obtidas a partir das médias *a posteriori* de $\{\delta_j\}$, calculadas para os dados observados.

variâncias das distribuições normais adotadas para $\beta_j, j = 0, \ldots, 7$, e $\delta_j, j = 0, 1, 2$ foram todas iguais a, respectivamente, 0 e 10^3. Apresentam-se resumos das distribuições *a posteriori* de $\beta_j, j = 0, \ldots, 7$, para tais análises na Tabela 4.7 e, na Tabela 4.8, para a análise de casos completos utilizando apenas o modelo (4.37).

Com exceção do parâmetro associado à variável $PaCO_2$, os desvios padrões *a posteriori* para os demais parâmetros são em geral bem menores nas análises que incluem todos os dados do que na análise de casos completos. Na Figura 4.3, as estimativas de θ_i (probabilidade de se observar $PaCO_2$) obtidas a partir das médias *a posteriori* de δ_0, δ_1 e δ_2 no modelo não-paramétrico e de estimativas de π_i (probabilidade de embolia pulmonar), calculadas para os dados observados, ilustram que a probabilidade de se observar $PaCO_2$ é bem maior para pacientes com alta probabilidade de embolia pulmonar do que para os que possuem uma baixa probabilidade. De qualquer forma, o fato de a probabilidade de observar $PaCO_2$ ser menor tanto para casos em que a probabilidade de embolia pulmonar é menor ou maior do que a prevalência faz com que a associação entre a existência de embolia pulmonar e os valores de $PaCO_2$ seja menor do que para modelos que não embutem tais suposições; note-se, por exemplo, a diferença entre as médias *a posteriori* da análise de casos completos e das análises que incluem todos os dados disponíveis. Apesar de as análises da Seção 4.1 terem indicado que os modelos paramétricos não se ajustam tão bem aos dados observados quanto o modelo não-paramétrico, isto parece não se ter propagado nas análises desta seção, uma vez que

as diferenças entre as medidas-resumos das distribuições *a posteriori* são muito pequenas. Assim, crê-se que qualquer dessas análises é mais indicada do que a análise de casos completos, uma vez que, ao embutir suposições sobre os dados faltantes, as primeiras devem fornecer resultados menos enviesados sobre a associação entre embolia pulmonar e $PaCO_2$, além de propiciarem resultados mais precisos sobre as demais associações.

4.6 Discussão

Este capítulo debruçou-se sobre a modelagem de respostas binárias no caso em que uma variável explicativa pode ter omissão informativa. Mostrou-se que a abordagem bayesiana com um modelo não-paramétrico via mistura por processo de Dirichlet para a covariável contínua é uma alternativa viável para evitar possíveis vieses causados nas inferências de interesse oriundos da escolha de uma distribuição paramétrica incorreta. Uma série de extensões podem ser consideradas, conforme se descreve a seguir.

Em primeiro lugar, duas ou mais variáveis explicativas contínuas podem sofrer omissão. Isto deve ser contemplado tanto no modelo para o mecanismo de omissão quanto no modelo para as variáveis explicativas. Ambos podem ser especificados com distribuições multivariadas ou com um produto de distribuições condicionais univariadas. No caso do mecanismo de omissão, alguns autores (*e.g.*, Lipsitz & Ibrahim, 1996; Ibrahim *et al.*, 1999) sugerem uma preferência pela última estratégia, ou seja, utilizar um produto de distribuições de Bernoulli em vez de distribuições de Bernoulli multivariadas. No caso do modelo para as variáveis explicativas, acredita-se também que talvez as misturas por processo de Dirichlet unidimensionais, como apresentadas por Ishwaran & James (2002), sejam um caminho mais fácil de se seguir do que, por exemplo, a versão multivariada do modelo não-paramétrico considerada por Escobar & West (1995), que utilizam uma distribuição normal multivariada juntamente com distribuições *a priori* normal multivariada e Wishart invertida. Além disso, a modelagem de Escobar & West (1995) constitui uma extensão do caso univariado descrito por West (1992) que, conforme discutido na Seção 4.2, apresenta algumas dificuldades quanto à escolha dos hiperparâmetros.

Em segundo lugar, pode existir interesse em se considerar outras distribuições para as respostas, além de casos multivariados. Não se acredita que a utilização destas outras distribuições forneça uma grande dificuldade adicional àquela que se obteria ao utilizá-las em análises de casos completos.

Em terceiro lugar, algumas das variáveis respostas podem estar sujeitas à omissão. Estes casos podem ser tratados modelando também os indicadores de observação destas variáveis.

Em quarto e último lugar, apesar de a mistura por processo de Dirichlet ser a análise bayesiana não-paramétrica mais empregada na literatura, pode-se querer avaliar outras alternativas não-paramétricas/semiparamétricas para a distribuição das covariáveis como a distribuição *a priori* de árvore de Pólya, que constitui uma generalização do processo de Dirichlet, resultando numa medida de probabilidade aleatória compatível com distribuições contínuas. Paddock (2002) utilizou este tipo

de distribuição *a priori* na análise de respostas com omissão ignorável. O uso deste tipo de distribuição *a priori* gera distribuições preditivas com descontinuidades, o que pode ser indesejável em algumas situações.

Provavelmente os maiores desafios quanto a aplicabilidade da modelagem deste capítulo, com ou sem as extensões supracitadas, são os casos em que as suposições sobre a omissão informativa geram modelos inidentificáveis e o tamanho da amostra for muito grande e/ou as distribuições *a priori* sobre todos os parâmetros forem muito vagas. Estas situações fazem com que as amostras das distribuições *a posteriori* obtidas pelo método MCMC fiquem muito autocorrelacionadas, exigindo cadeias muito longas para que seja possível detectar a convergência e também para que se obtenha erros de Monte Carlo suficientemente pequenos para se garantir a precisão desejada nas inferências de interesse. Particularmente, estes cenários, já observados no Capítulo 3, agravam-se com tamanhos de amostras grandes nas análises deste capítulo, porque a aproximação do processo de Dirichlet pela sua versão truncada necessita um maior número de componentes. Possivelmente, os outros esquemas de MCMC descritos na revisão de Griffin & Holmes (2010) podem atenuar esse problema.

Capítulo 5

Análise de sensibilidade clássica para a estimação de momentos em dados com omissão

Em praticamente todos os problemas com dados faltantes requerem-se suposições inverificáveis por critérios estatísticos, como omissão aleatória (MAR), para identificar modelos estatísticos apropriados. Tais suposições são usualmente questionáveis e estatísticos geralmente tratam do problema por meio de análises de sensibilidade. Especificamente para dados contínuos, Rubin (1977), Little (1994), Little & Wang (1996) e Daniels & Hogan (2000) propõem análises de sensibilidade sob suposições de normalidade, enquanto Rotnitzky, Robins & Scharfstein (1998), Scharfstein, Rotnitzky & Robins (1999) e Rotnitzky, Scharfstein, Su & Robins (2001) usam métodos de ponderação pelo inverso da probabilidade (IPW, *inverse probability weighted*) no contexto de modelos semiparamétricos para propósitos similares. Revisões de algumas dessas e outras abordagens são apresentadas em Daniels & Hogan (2007, Cap. 9 e 10), Molenberghs & Kenward (2007, Cap. 19-25) e Fitzmaurice, Davidian, Verbeke & Molenberghs (2008, Cap. 18, 20 e 22).

Apesar de tais desenvolvimentos metodológicos serem úteis em muitas situações, há casos em que a sua aplicação se afigura difícil. Para evitar tais dificuldades, deriva-se uma abordagem simples para estimar médias, desvios padrões e correlações. Adota-se a parametrização do modelo de mistura de padrões (Glynn *et al.*, 1986; Little & Rubin, 2002) e empregam-se as médias, desvios padrões e correlações inidentificáveis, ou funções deles, como parâmetros de sensibilidade. Essa estratégia é parecida com a adotada por Daniels & Hogan (2000), embora aqui não se suponha qualquer distribuição paramétrica para as respostas. Acredita-se que, em muitas aplicações, pode ser mais fácil eliciar informação sobre esses parâmetros de sensibilidade do que sobre as funções de viés de seleção usadas por Rotnitzky e colegas. Ao invés de métodos IPW, estimam-se os parâmetros identificáveis por meio de seus análogos amostrais.

Na Seção 5.1, apresenta-se o conjunto de dados que se utiliza para ilustrar os métodos descritos no restante do capítulo. Introduzem-se as ideias com o caso univariado na Seção 5.2 e considera-se a extensão multivariada na Seção 5.3. Avaliam-se os intervalos de incerteza empregados nas inferências na Seção 5.4. Na Seção 5.5, discutem-se as vantagens e desvantagens da abordagem

proposta relativamente aos demais métodos mencionados nos parágrafos anteriores.

5.1 Dados de faculdades americanas

O *U.S. News & World Report's Guide to America's Best Colleges 1995* coletou dados de estudantes de 1 302 faculdades e universidades americanas baseados em mais de 30 variáveis abrangendo características como admissão, custos, infraestrutura e desempenho. Allison (2001) considerou a estimação de médias, desvios padrões e correlações para sete dessas variáveis sob suposições de normalidade e MAR usando o algoritmo EM. Para a nossa exposição é suficiente enfocar apenas cinco delas, nomeadamente, GRADRAT (razão entre o número de graduandos do último ano e o número de alunos matriculados quatro anos antes $\times 100$), CSAT (média combinada das notas nas provas verbal e de matemática do *Scholastic Assessment Test*), ACT (média dos escores obtidos no *American College Testing*), RMBRD (custo anual total com alojamento e alimentação, em milhares de dólares) e um indicador de faculdades públicas e particulares. Para uma das faculdades o valor registrado para o GRADRAT é 118 e, portanto, foi considerado omisso. O indicador de tipo de administração da faculdade (público/particular) foi a única variável sem valores omissos; as outras quatro variáveis foram observadas simultaneamente para apenas 23% das faculdades e cada variável teve de 8% a 45% de valores faltantes. Há interesse em responder-se a duas questões:

i. faculdades públicas e particulares têm médias e desvios padrões de CSAT diferentes?

ii. GRADRAT, CSAT, ACT e RMBRD são correlacionados linearmente entre si?

Exibem-se estatísticas descritivas nas Tabelas 5.1 e 5.2. Como o levantamento dos dados abrangeu todas as faculdades americanas que satisfizeram os critérios do estudo, os dados compõem toda a população investigada e, portanto, calculam-se e discutem-se os erros padrões e intervalos de confiança e de incerteza com fins meramente ilustrativos.

Tabela 5.1: Estatísticas descritivas para CSAT.

Administração	CSAT observado			CSAT em falta		
da faculdade	Freq. (%)	Média	DP	Freq. (%)	Média	DP
pública	251 (0.53)	945.3	107.5	219 (0.47)	?	?
particular	528 (0.63)	978.8	129.2	304 (0.37)	?	?

Nota: ? denota valores não-observados, e DP, desvio padrão.

5.2 Caso univariado

Sejam Y_i a resposta da i-ésima unidade do estudo e R_i a variável indicadora que assume o valor 1 se Y_i é observado e 0, em caso contrário, $i = 1, \ldots, n$. Na estrutura de modelos de mistura de padrões, a distribuição conjunta (Y_i, R_i) é fatorada como o produto da distribuição marginal de R_i e a distribuição condicional de Y_i dado R_i. Como o interesse recai apenas nos momentos de Y_i, usa-se a

Tabela 5.2: Estatísticas descritivas para GRADRAT (G), CSAT (C), ACT (A) e RMBRD (R).

Padrão resp.				Freq.	GRADRAT		CSAT		ACT		RMBRD		Correlações					
G	C	A	R	(%)	Média	DP	Média	DP	Média	DP	Média	DP	G×C	G×A	G×R	C×A	C×R	A×R
o	o	o	o	296 (0.23)	62.6	19.0	988.9	120.4	22.90	2.74	4.25	0.99	0.56	0.58	0.43	0.92	0.45	0.47
o	o	o	f	158 (0.12)	58.0	17.0	961.5	110.6	22.30	2.53	?	?	0.56	0.54	?	0.87	?	?
o	o	f	o	158 (0.12)	64.6	18.2	976.7	134.7	?	?	4.73	1.19	0.72	?	0.50	?	0.53	?
o	f	o	o	123 (0.09)	53.0	14.5	?	?	21.39	1.91	3.25	0.78	?	0.47	0.26	?	?	0.39
f	o	o	o	17 (0.01)	?	?	879.6	98.1	20.76	2.41	3.96	1.14	?	?	?	0.83	−0.29	−0.27
o	o	f	f	119 (0.09)	62.7	19.1	949.9	124.9	?	?	?	?	0.55	?	?	?	?	?
o	f	f	o	157 (0.12)	64.8	20.8	?	?	?	?	4.27	1.26	?	?	0.58	?	?	?
o	f	o	f	82 (0.06)	51.9	16.7	?	?	21.12	1.98	?	?	?	0.36	?	?	?	?
f	f	o	o	11 (0.01)	?	?	?	?	20.82	2.27	3.22	0.85	?	?	?	?	?	−0.15
f	o	f	o	5 (0.00)	?	?	871.4	96.1	?	?	3.50	1.49	?	?	?	?	0.02	?
f	o	o	f	17 (0.01)	?	?	880.9	109.6	20.47	2.83	?	?	?	?	?	0.93	?	?
o	f	f	f	110 (0.08)	57.4	19.5	?	?	?	?	?	?	?	?	?	?	?	?
f	o	f	f	9 (0.01)	?	?	865.9	56.8	?	?	?	?	?	?	?	?	?	?
f	f	o	f	10 (0.01)	?	?	?	?	19.90	1.37	?	?	?	?	?	?	?	?
f	f	f	o	16 (0.01)	?	?	?	?	?	?	3.20	1.14	?	?	?	?	?	?
f	f	f	f	14 (0.01)	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Análise				estat.	60.4	18.8	968.0	123.6	22.12	2.58	4.15	1.17	0.59	0.56	0.50	0.91	0.44	0.47
de Casos				freq.	1 203		779		714		783		731	659	734	488	476	447
Disponíveis				(%)	(0.92)		(0.60)		(0.55)		(0.60)		(0.56)	(0.51)	(0.56)	(0.37)	(0.37)	(0.34)

Nota: ? denota valores não-observados, DP, desvio padrão, o, observado e f, em falta.

abordagem de modelos de mistura de padrões e propriedades de esperança condicional para escrever

$$\mu = \mathrm{E}(Y_i) = \mathrm{E}[\mathrm{E}(Y_i|R_i)] = \gamma_1\mu_{(1)} + \gamma_0\mu_{(0)}, \tag{5.1}$$

$$\begin{aligned}\sigma^2 &= \mathrm{Var}(Y_i) = \mathrm{E}[\mathrm{Var}(Y_i|R_i)] + \mathrm{Var}[\mathrm{E}(Y_i|R_i)] \\ &= \gamma_1\sigma^2_{(1)} + \gamma_0\sigma^2_{(0)} + \gamma_1(\mu_{(1)} - \mu)^2 + \gamma_0(\mu_{(0)} - \mu)^2,\end{aligned} \tag{5.2}$$

em que $\gamma_r = P(R_i = r)$, $\mu_{(r)} = \mathrm{E}(Y_i|R_i = r)$ e $\sigma^2_{(r)} = \mathrm{Var}(Y_i|R_i = r)$, para $r = 0, 1$.

Pode-se estimar γ_1 ($\gamma_0 = 1-\gamma_1$), $\mu_{(1)}$ e $\sigma_{(1)}$ pelas suas contrapartidas amostrais, *i.e.*, a proporção de unidades observada, $\hat{\gamma}_1$, e a média e o desvio padrão amostrais das respostas observadas, $\hat{\mu}_{(1)}$ e $\hat{\sigma}_{(1)}$. Entretanto, $\boldsymbol{\omega} = \mu_{(0)}$ ou $\boldsymbol{\omega} = (\mu_{(0)}, \sigma_{(0)})$ também são necessários para estimar μ ou σ, respectivamente, e ambos $\mu_{(0)}$ e $\sigma_{(0)}$ não são identificados pelos dados observados. Por conseguinte, é útil recordar que a incerteza estatística é uma combinação de imprecisão estatística e ignorância estatística (*vide* Seção 3.1.3). No caso que se discute aqui, a ignorância estatística está relacionada com a média e o desvio padrão das unidades não-observadas. Fixando um valor $\boldsymbol{\omega}^S$ para $\boldsymbol{\omega}$, pode-se calcular $\hat{\gamma}_1$, $\hat{\mu}_{(1)}$ e $\hat{\sigma}_{(1)}$ e usando (5.1) e (5.2), pode-se obter uma estimativa não-enviesada $\hat{\mu}(\boldsymbol{\omega}^S)$ de $\mu(\boldsymbol{\omega}^S)$ e uma estimativa consistente $\hat{\sigma}(\boldsymbol{\omega}^S)$ de $\sigma(\boldsymbol{\omega}^S)$. Como isso permanece válido para qualquer $\boldsymbol{\omega}^S$, pode-se realizar uma análise de sensibilidade, obtendo estimativas e regiões de confiança para o alvo percorrendo um conjunto de valores $\boldsymbol{\Omega}_{\boldsymbol{\omega}}$ que se espera conter o verdadeiro valor para o parâmetro de sensibilidade $\boldsymbol{\omega}$. Na Seção 3.1.3, descrevem-se a Região de Ignorância Estimada (RIgE) e as Regiões de Incertezas Estimadas (RInE) forte, ponto a ponto e fraca que serão utilizadas no restante deste capítulo.

Para dados categorizados com omissão, o conjunto $\boldsymbol{\Omega}_{\boldsymbol{\omega}}$ pode cobrir uma grade detalhada de todo o espaço paramétrico de $\boldsymbol{\omega}$, mas para dados contínuos, essa estratégia é claramente inexequível; *e.g.*, note-se em (5.1) que $\hat{\mu}(\mu_{(0)}) \longrightarrow \infty$ conforme $\mu_{(0)} \longrightarrow \infty$ independentemente de $0 < \hat{\gamma}_1 < 1$ e $\hat{\mu}_{(1)}$. Consequentemente, como não é possível realizar a análise sem uma cuidadosa avaliação de $\boldsymbol{\Omega}_{\boldsymbol{\omega}}$, o analista dos dados deve procurar alternativas de parametrizações para os parâmetros de sensibilidade e selecionar aquela que possibilite a eliciação mais facilmente executada. Por exemplo, ao invés de utilizar $\mu_{(0)}$ como parâmetro de sensibilidade, pode-se preferir usar α, β ou p, em que $\mu_{(0)} = \alpha + \mu_{(1)}$, $\mu_{(0)} = \beta\mu_{(1)}$, principalmente para variáveis positivas, e $\mu_{(0)} = F^{-1}_{(1)}(p)$, o p-ésimo quantil da distribuição das unidades observadas. A variância do estimador de μ depende de qual parâmetro de sensibilidade é usado, *e.g.*, para os primeiros três casos, tem-se

$$\mathrm{Var}[\hat{\mu}(\mu_{(0)})] = \frac{\gamma_1\sigma^2_{(1)}}{n} + \frac{\gamma_1(1-\gamma_1)(\mu_{(1)} - \mu_{(0)})^2}{n}, \tag{5.3}$$

$$\mathrm{Var}[\hat{\mu}(\alpha)] = \sigma^2_{(1)}\mathrm{E}\left(\frac{1}{n_1}\right) + \frac{\gamma_1(1-\gamma_1)\alpha^2}{n}, \tag{5.4}$$

$$\mathrm{Var}[\hat{\mu}(\beta)] = \sigma^2_{(1)}\left[\beta^2\mathrm{E}\left(\frac{1}{n_1}\right) + \frac{2\beta(1-\beta)}{n} + \frac{\gamma_1(1-\beta)^2}{n}\right] + \frac{\gamma_1(1-\gamma_1)\mu^2_{(1)}(1-\beta)^2}{n}, \tag{5.5}$$

em que $n_1 = \sum_{i=1}^{n} R_i$ denota o número de unidades observadas.

Uma vez que não se pode realizar uma análise sensata se todas as respostas estiverem omissas, *i.e.*, se $n_1 = 0$, poder-se-ia ter calculado (5.3)–(5.5) admitindo que n_1 segue uma distribuição binomial positiva (Stephan, 1945) ao invés da distribuição binomial com parâmetros n e γ_1. Entretanto, as alterações requeridas para tais propósitos geram fórmulas desnecessariamente mais complicadas que só melhoram a precisão nos casos em que $n\gamma_1$ é pequeno. Aproximações simples para o primeiro momento negativo da distribuição binomial positiva são discutidos, por exemplo, por Grab & Savage (1954) e Mendenhall & Lehman (1960). Emprega-se aqui $\mathrm{E}(1/n_1) \cong 1/(n\gamma_1 + 1 - \gamma_1)$, que geralmente tem precisão de pelo menos dois dígitos decimais se $n\gamma_1 > 10$ (Grab & Savage, 1954). Todavia, comparar (5.3)–(5.5) é mais fácil utilizando a aproximação mais grosseira $1/n\gamma_1$; conclui-se, então, que as seguintes relações valem para n grande:

$$\mathrm{Var}[\hat{\mu}(\mu_{(0)})] < \mathrm{Var}[\hat{\mu}(\alpha)] \leq \mathrm{Var}[\hat{\mu}(\beta)], \quad \text{if } \beta \leq \beta_1 \text{ or } \beta \geq 1, \tag{5.6}$$

$$\mathrm{Var}[\hat{\mu}(\mu_{(0)})] \leq \mathrm{Var}[\hat{\mu}(\beta)] < \mathrm{Var}[\hat{\mu}(\alpha)], \quad \text{if } \beta_1 < \beta \leq \beta_2 \text{ or } \beta_3 \leq \beta < 1, \tag{5.7}$$

$$\mathrm{Var}[\hat{\mu}(\beta)] < \mathrm{Var}[\hat{\mu}(\mu_{(0)})] < \mathrm{Var}[\hat{\mu}(\alpha)], \quad \text{if } \beta_2 < \beta < \beta_3, \tag{5.8}$$

em que $\beta_1 = (-1-\gamma_1)/(1-\gamma_1)$, $\beta_2 = (-\gamma_1^4 - \gamma_1)/(1-\gamma_1)$ e $\beta_3 = (\gamma_1^4 - \gamma_1)/(1-\gamma_1)$. Como $\beta_1 < \beta_2 < \beta_3 < 0$ e β é geralmente positivo, (5.8) dificilmente ocorrerá na prática.

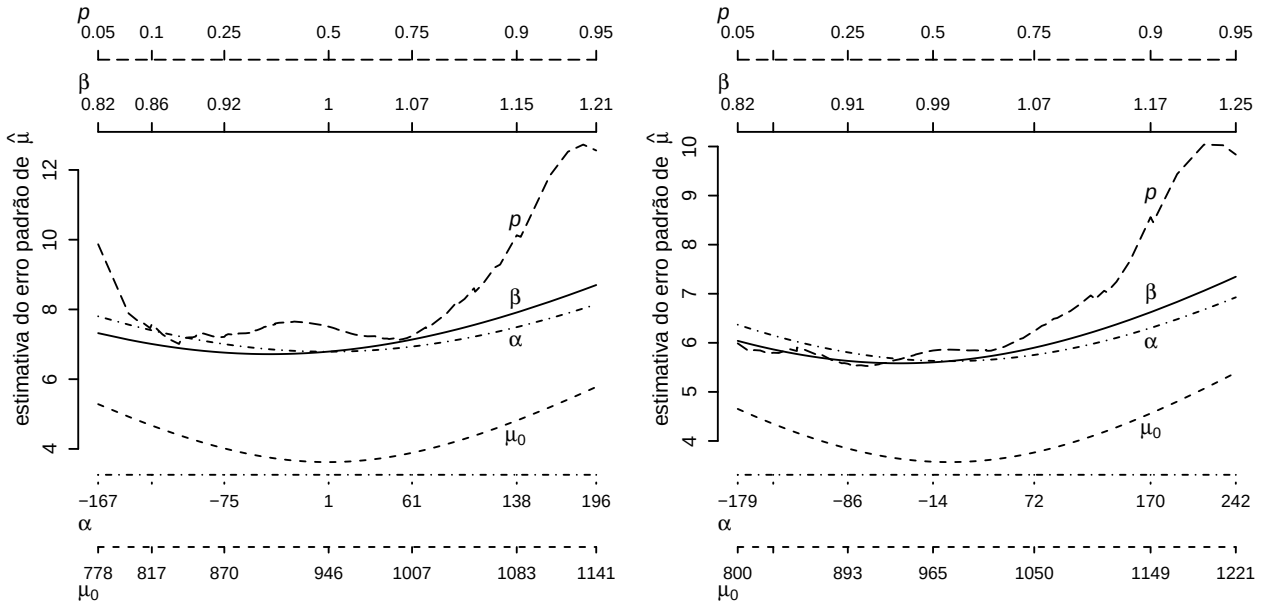
Usando expressões assintóticas para variâncias e covariâncias de estatísticas de ordem (Sen, Singer & Pedroso de Lima, 2009, p.223), obtém-se

$$\begin{aligned}\mathrm{Var}\,[\hat{\mu}(p)] \cong{}& \frac{\gamma_1\sigma_{(1)}^2}{n} + \frac{p(1-p)}{\left\{f_{(1)}\left[F_{(1)}^{-1}(p)\right]\right\}^2}\left[\frac{\gamma_1 - 2}{n} + \mathrm{E}\left(\frac{1}{n_1}\right)\right] + \frac{\gamma_1(1-\gamma_1)\left[\mu_{(1)} - F_{(1)}^{-1}(p)\right]^2}{n} \\ &+ \frac{2}{n}\left[\mathrm{E}\left(\frac{1}{n_1}\right) - \frac{1}{n}\right]\left\{\sum_{j=1}^{k}\frac{p_j(1-p)}{f_{(1)}\left[F_{(1)}^{-1}(p_j)\right]f_{(1)}\left[F_{(1)}^{-1}(p)\right]} + \sum_{j=k+1}^{n_1}\frac{p(1-p_j)}{f_{(1)}\left[F_{(1)}^{-1}(p)\right]f_{(1)}\left[F_{(1)}^{-1}(p_j)\right]}\right\},\end{aligned} \tag{5.9}$$

em que $f_{(1)}$ denota a densidade das unidades observadas, $p_k < p < p_{k+1}$ e $p_j = j/n_1 + o(n_1^{-1/2})$, $j = 1, \ldots, n_1$, tal que $0 < p_1 < p_2 < \ldots < p_{n_1} < 1$, *e.g.*, p_j pode ser igual a $(j - 0.5)/n_1$ ou a uma das outras três definições discutidas por Hyndman & Fan (1996) que satisfaçam sua Propriedade 5; supõe-se também que $f_{(1)}$ e $F_{(1)}$ são contínuos em $F_{(1)}^{-1}(p)$ e em $F_{(1)}^{-1}(p_j)$, $j = 1, \ldots, n_1$, e que $f_{(1)}\left[F_{(1)}^{-1}(p)\right] > 0$ e $f_{(1)}\left[F_{(1)}^{-1}(p_j)\right] > 0$, $j = 1, \ldots, n_1$.

A Figura 5.1 retrata as raízes quadradas das estimativas de (5.3)–(5.5) e (5.9) para os dados da Tabela 5.1, obtidas por substituição dos parâmetros pelos seus análogos amostrais; $f_{(1)}$ em (5.9) é substituída por uma estimativa da densidade obtida pelo método do núcleo gaussiano (Silverman, 1986). Os quatro eixos horizontais indicam equivalências entre os quatro parâmetros de sensibilidade para estimar μ; por exemplo, $p = 0.25$, $\beta \cong 0.92$, $\alpha \cong -75$ e $\mu_{(0)} = 870$ levam ao mesmo valor

para a estimativa $\hat{\mu}$ para faculdades públicas, mas a diferentes estimativas do seu erro padrão. Os erros padrões obtidos quando $\mu_{(0)}$ é escolhido como parâmetro de sensibilidade são menores do que aqueles baseados nos outros parâmetros de sensibilidade porque como α, β e p relacionam a média não-observada com a média observada ou com o quantil da distribuição observada, essas abordagens incluem mais incerteza sobre $\mu_{(0)}$ do que a suposição de que $\mu_{(0)}$ é conhecido. Por esta razão, quando há um palpite plausível para $\mathbf{\Omega}_{\mu_{(0)}}$, $\mu_{(0)}$ tem a vantagem sobre outros parâmetros de sensibilidade de gerar um estimador mais preciso para μ. Os espaços paramétricos de α e β não são limitados (assim como $\mu_{(0)}$) e as precisões das suas estimativas não são tão diferentes, sugerindo que estes parâmetros de sensibilidade podem ser igualmente escolhidos exclusivamente com propósitos interpretativos. O parâmetro de sensibilidade p, por outro lado, tem um espaço paramétrico limitado e a análise correspondente pode apenas fornecer respostas razoáveis se houver crença de que o mecanismo de omissão não é tão "extremo" ao ponto de a média da distribuição não-observada poder ser menor do que o mínimo (ou maior do que o máximo) valor observado.



(a) (b)

Figura 5.1: Estimativas dos erros padrões para $\hat{\mu}$ de CSAT usando $\mu_{(0)}$, α, β ou p como parâmetro de sensibilidade para faculdades (a) públicas e (b) particulares.

Examinando a estimação de σ em (5.2), além de se empregar $\sigma_{(0)}$ como parâmetro de sensibilidade, pode-se alternativamente trabalhar com λ, em que $\sigma_{(0)} = \lambda\sigma_{(1)}$. Quer com $\big(\mu_{(0)}, \sigma_{(0)}\big)$ como parâmetro de sensibilidade ou quer com qualquer outra parametrização, uma maneira simples de se obter uma estimativa para a variância de $\hat{\sigma}$ é empregar o *bootstrap* não-paramétrico[1] (Efron & Gong, 1983). Como na discussão anterior, espera-se que $\mathrm{Var}[\hat{\sigma}(\sigma_{(0)})] < \mathrm{Var}[\hat{\sigma}(\lambda)]$ quando $\hat{\sigma}(\sigma_{(0)}) = \hat{\sigma}(\lambda)$,

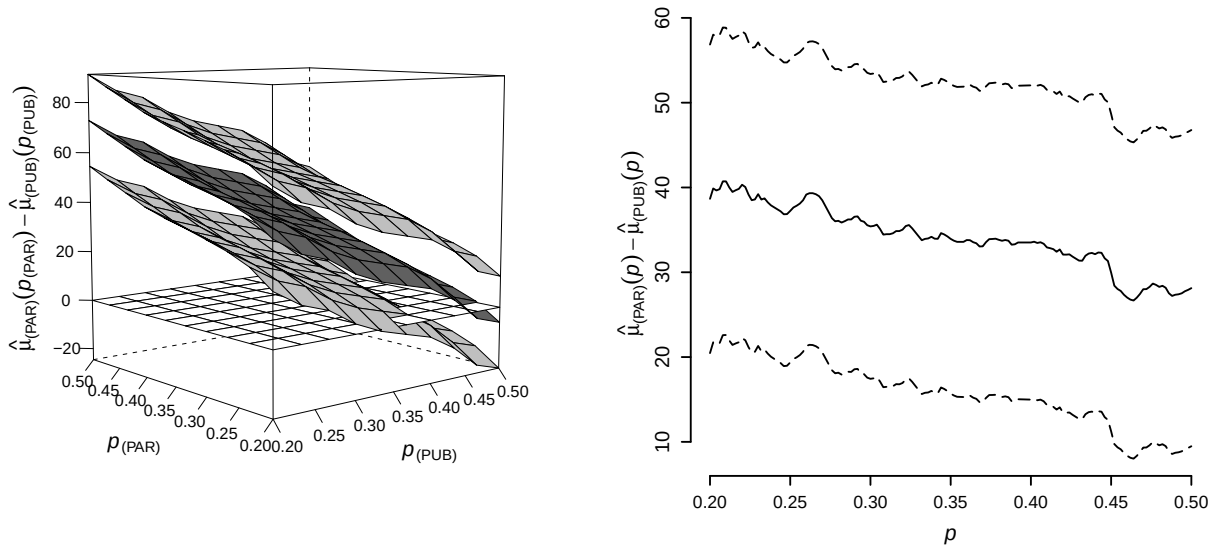
[1]Sempre que se empregar neste capítulo o *bootstrap* para estimar erros padrões, geram-se 2 000 réplicas da estatística.

para qualquer parametrização empregada para $\mu_{(0)}$. Por exemplo, fixando $\sigma_0 = 134.3$ ou $\lambda = 1.25$ para faculdades públicas, ambos com $\mu_{(0)} = 870$, $\alpha \cong -75$, $\beta \cong 0.92$ ou $p = 0.25$, obtém-se a mesma estimativa para σ (126.4), mas as estimativas para os erros padrões de $\hat{\sigma}$ são iguais a 2.6, 5.5, 2.3, 5.2, 2.3, 5.1, 2.9, 5.8, respectivamente, ao empregar-se $(\mu_{(0)}, \sigma_0)$, $(\mu_{(0)}, \lambda)$, ..., (p, λ). Apesar de não se ter uma explicação de por que os erros padrões são um pouco maiores quando se empregar $\mu_{(0)}$ do que os obtidos usando α e β, a principal conclusão neste caso é que a parametrização para o parâmetro de sensibilidade de $\mu_{(0)}$ parece ter menos impacto no erro padrão de $\hat{\sigma}$ do que a parametrização de $\sigma_{(0)}$.

Uma forma de avaliar a questão (i) da Seção 5.1 é por meio das funções paramétricas $\mu_{(\mathrm{PAR})} - \mu_{(\mathrm{PUB})}$ e $\sigma_{(\mathrm{PAR})}/\sigma_{(\mathrm{PUB})}$. Como não se tem oportunidade de se obter informação de especialistas, prefere-se usar a parametrização (p, λ) a qual se especula que

1. a média de CSAT para ambas faculdades públicas e particulares que não informaram o CSAT podem estar entre os quantis 20% e 50% das distribuições observadas de CSAT e
2. os desvios padrões de CSAT para faculdades que não reportaram o CSAT podem estar no intervalo compreendido entre 0.80 a 1.25 vezes os correspondentes desvios padrões para faculdades que o informaram.

Primeiramente, considera-se parâmetros de sensibilidade individuais para cada nível de administração de faculdade, $\mathbf{\Omega}_{p_{(\mathrm{PUB})}} = \mathbf{\Omega}_{p_{(\mathrm{PAR})}} = [0.2; 0.5]$ e $\mathbf{\Omega}_{\lambda_{(\mathrm{PUB})}} = \mathbf{\Omega}_{\lambda_{(\mathrm{PAR})}} = [0.80; 1.25]$, e, em seguida,



(a) (b)

Figura 5.2: Estimativas e intervalos de 95% de confiança para $\mu_{(\mathrm{PAR})} - \mu_{(\mathrm{PUB})}$ com (a) um parâmetro de sensibilidade para cada nível de administração de faculdade, $p_{(\mathrm{PUB})}$ e $p_{(\mathrm{PAR})}$ e (b) um parâmetro de sensibilidade comum, p, para ambas faculdades públicas e particulares.

Tabela 5.3: RIgE e RInE a 95% para $\mu_{(\mathrm{PAR})} - \mu_{(\mathrm{PUB})}$ e $\sigma_{(\mathrm{PAR})}/\sigma_{(\mathrm{PUB})}$ para CSAT.

Parâmetro	Par. Sens.[a]	RIgE	RInE Fraca	RInE Ponto a Ponto	RInE Forte
$\mu_{(\mathrm{PAR})} - \mu_{(\mathrm{PUB})}$	individual	[−6.1; 72.9]	[−10.4; 77.2]	[−21.6; 88.2]	[−24.6; 91.2]
	comum	[26.5; 41.0]	[13.5; 53.7]	[10.8; 56.2]	[7.9; 59.1]
$\frac{\sigma_{(\mathrm{PAR})}}{\sigma_{(\mathrm{PUB})}}$	individual	[0.93; 1.54]	[0.91; 1.58]	[0.84; 1.69]	[0.83; 1.71]
	comum	[1.16; 1.24]	[1.06; 1.34]	[1.05; 1.35]	[1.03; 1.37]

[a]Parâmetros de sensibilidade individuais e comuns para os níveis de administração de faculdade.

simplifica-se o modelo para o caso em que os parâmetros de sensibilidade são iguais para ambos níveis de administração de faculdade, ou seja,, $p = p_{(\mathrm{PUB})} = p_{(\mathrm{PAR})}$ e $\lambda = \lambda_{(\mathrm{PUB})} = \lambda_{(\mathrm{PAR})}$, e consequentemente, $\mathbf{\Omega}_p = [0.2; 0.5]$ e $\mathbf{\Omega}_\lambda = [0.80; 1.25]$. Retratam-se estimativas e intervalos de confiança para $\mu_{(\mathrm{PAR})} - \mu_{(\mathrm{PUB})}$ na Figura 5.2. Interessantemente, como a função de distribuição empírica é uma função em escada, $\hat{\mu}_{(\mathrm{PAR})} - \hat{\mu}_{(\mathrm{PUB})}$ é uma função monótona quando $p_{(\mathrm{PUB})}$ e $p_{(\mathrm{PAR})}$ são empregados, mas é uma função não-monótona quando apenas p é usado. Esta falta de monotonicidade não teria sido observada se $\mu_{(0)}$, α ou β tivesse sido empregado. Apresentam-se as RIgE e as RInE para ambas funções paramétricas na Tabela 5.3. Se houvesse crença de que os parâmetros de sensibilidade são iguais para os dois níveis de administração das faculdades, concluir-se-ia que a média e o desvio padrão de CSAT para faculdades particulares são maiores do que os valores correspondentes obtidos para faculdades públicas. A mesma conclusão seria obtida com a comumente empregada análise de dados completos, que é válida sob a suposição de $\mu_{(0)} = \mu_{(1)}$ e $\sigma_{(0)} = \sigma_{(1)}$ para cada nível de administração de faculdade, fornecendo estimativas e intervalos de 95% de confiança para $\mu_{(\mathrm{PAR})} - \mu_{(\mathrm{PUB})}$ e $\sigma_{(\mathrm{PAR})}/\sigma_{(\mathrm{PUB})}$ iguais a, respectivamente, 33.5 [16.2; 50.8] e 1.20 [1.07; 1.34]. Contudo, como não se tem informação que sustente qualquer dessas fortes suposições, confia-se nos resultados que empregam um conjunto de parâmetros de sensibilidade para cada nível de administração de faculdade, embora eles não permitam concluir se há uma diferença entre as médias ou entre os desvios padrões de faculdades públicas e particulares. Reduções no número de parâmetros de sensibilidade devem ser conduzidas com cautela para não se correr o risco de excluir o verdadeiro modelo subjacente; veja também a discussão da Capítulo 3 no contexto de dados categorizados incompletos.

5.3 Caso multivariado

Seja $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iJ})'$ em que Y_{ij} denota a j-ésima resposta da i-ésima unidade do estudo, $i = 1, \ldots, n$, $j = 1, \ldots, J$. Além disso, defina-se um vetor de indicadores de resposta $\mathbf{R}_i = (R_{i1}, \ldots, R_{iJ})'$ com $R_{ij} = 1$ se Y_{ij} é observado e $R_{ij} = 0$ em caso contrário. As relações (5.1) e (5.2) são então as versões univariadas de

$$\boldsymbol{\mu} = \sum_{\mathbf{r} \in \mathcal{R}} \gamma_{\mathbf{r}} \boldsymbol{\mu}_{(\mathbf{r})}, \tag{5.10}$$

$$\boldsymbol{\Sigma} = \sum_{\mathbf{r}\in\mathcal{R}} \gamma_{\mathbf{r}} \boldsymbol{\Sigma}_{(\mathbf{r})} + \sum_{\mathbf{r}\in\mathcal{R}} \gamma_{\mathbf{r}} (\boldsymbol{\mu}_{(\mathbf{r})} - \boldsymbol{\mu})(\boldsymbol{\mu}_{(\mathbf{r})} - \boldsymbol{\mu})', \tag{5.11}$$

em que $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_J)' = \mathrm{E}(\mathbf{Y}_i)$, $\boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{Y}_i)$, $\gamma_{\mathbf{r}} = P(\mathbf{R}_i = \mathbf{r})$, $\mathcal{R}$ contém todos os valores observados de $\mathbf{r} = (r_1, \ldots, r_J)'$, $\boldsymbol{\mu}_{(\mathbf{r})} = (\mu_{1(\mathbf{r})}, \ldots, \mu_{J(\mathbf{r})})' = \mathrm{E}(\mathbf{Y}_i|\mathbf{R}_i = \mathbf{r})$ e $\boldsymbol{\Sigma}_{(\mathbf{r})} = \mathrm{Cov}(\mathbf{Y}_i|\mathbf{R}_i = \mathbf{r})$. Frequentemente, prefere-se trabalhar com correlações, $\psi_{jk(\mathbf{r})} = \mathrm{Corr}(Y_{ij}, Y_{ik}|\mathbf{R}_i = \mathbf{r})$, ao invés de covariâncias, $\sigma_{jk(\mathbf{r})} = \mathrm{Cov}(Y_{ij}, Y_{ik}|\mathbf{R}_i = \mathbf{r})$, $j \neq k$, e por conseguinte escreve-se $\boldsymbol{\Sigma}_{(\mathbf{r})} = \mathbf{D}_{\boldsymbol{\sigma}_{(\mathbf{r})}} \boldsymbol{\Psi}_{(\mathbf{r})} \mathbf{D}_{\boldsymbol{\sigma}_{(\mathbf{r})}}$, em que $\mathbf{D}_{\boldsymbol{\sigma}_{(\mathbf{r})}}$ denota uma matriz diagonal com os elementos de $\boldsymbol{\sigma}_{(\mathbf{r})}$ na diagonal principal, $\boldsymbol{\sigma}_{(\mathbf{r})} = (\sigma_{j(\mathbf{r})}, j = 1, \ldots, J)'$, $\sigma^2_{j(\mathbf{r})} = \mathrm{Var}(Y_{ij}|\mathbf{R}_i = \mathbf{r})$ e $\boldsymbol{\Psi}_{(\mathbf{r})} = \mathrm{Corr}(\mathbf{Y}_i|\mathbf{R}_i = \mathbf{r})$; as definições correspondentes de variâncias, covariâncias e correlações incondicionais seguem analogamente.

Na versão tetravariada da Tabela 5.2, há 16 padrões de omissão e 136 parâmetros inidentificáveis para estimar $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ usando (5.10) e (5.11); dentre estes parâmetros de sensibilidade, há 32 médias, 32 desvios padrões e 72 correlações, conforme indicado pelos valores não-observados na Tabela 5.2. Para $J = 10$ variáveis sujeitas à omissão, há $2^J = 1\,024$ padrões de omissão potenciais que geram 44 800 parâmetros inidentificáveis: $J \times 2^{J-1} = 5\,120$ médias, 5 120 desvios padrões e $\sum_{j=0}^{J-1} \binom{J}{j} \left[\binom{J}{2} - \binom{j}{2}\right] = 34\,560$ correlações. O desafio aqui não é apenas de que o número de parâmetros de sensibilidade cresce exponencialmente dependendo dos padrões de omissão e do número de variáveis, mas também que há opções adicionais de parametrização. Por exemplo, no caso tetravariado, ao invés de usar $\mu_{1(0,r_2,r_3,r_4)}$, $r_2, r_3, r_4 = 0, 1$, como parâmetros de sensibilidade para estimar μ_1, pode-se preferir empregar outros parâmetros de sensibilidade que são funções destes parâmetros inidentificáveis, nomeadamente, $\mu_{1(1,r_2,r_3,r_4)}$, $r_2, r_3, r_4 = 0, 1$, ou ainda usar outras opções, como quantis da distribuição de Y_{i1} condicionalmente a $\mathbf{R}_i = (1, r_2, r_3, r_4)$, para alguns valores de r_2, r_3, r_4. Outra alternativa é usar todos os dados disponíveis da j-ésima variável, ou seja, a média $\mu_{j(1)}$ ou o quantil da distribuição de Y_{ij} condicionalmente a $R_{ij} = 1$, $F_{j(1)}$, *e.g.*, $\mu_{1(0,r_2,r_3,r_4)} = \alpha_{1(0,r_2,r_3,r_4)} + \mu_{1(1)}$, $\mu_{1(0,r_2,r_3,r_4)} = \beta_{1(0,r_2,r_3,r_4)}\mu_{1(1)}$ ou $\mu_{1(0,r_2,r_3,r_4)} = F^{-1}_{j(1)}(p_{1(0,r_2,r_3,r_4)})$. Analogamente, e como na seção anterior, há também a possibilidade de se considerarem reparameterizações da forma $\sigma_{1(0,r_2,r_3,r_4)} = \lambda_{1(0,r_2,r_3,r_4)}\sigma_{1(1)}$, em que $\sigma^2_{j(1)} = \mathrm{Var}(Y_{ij}|R_{ij} = 1)$.

O limite inferior (superior) do intervalo de ignorância para cada função paramétrica de interesse pode ser obtido minimizando (maximizando) a função correspondente com relação aos parâmetros de sensibilidade apropriados. No caso tetravariado, por exemplo, quando o alvo é a média μ_j, as otimizações são realizadas sobre apenas as 8 médias inidentificáveis de $\{\mu_{j(\mathbf{r})}\}$ e, para os desvios padrões, elas são efetuadas com relação às 8 médias inidentificáveis de $\{\mu_{j(\mathbf{r})}\}$ e aos 8 desvios padrões inidentificáveis de $\{\sigma_{j(\mathbf{r})}\}$ (16 parâmetros de sensibilidade). Para cada correlação, o número de parâmetros de sensibilidade a serem empregados na otimização aumenta para 44 (16 médias, 16 desvios padrões e 12 correlações). Portanto, na prática, todos os 136 parâmetros de sensibilidade seriam utilizados simultaneamente apenas se o alvo fosse alguma função específica de $(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\Psi})$.

Por causa do grande número de alternativas de parametrização que podem ser combinadas para um conjunto de variáveis, calcula-se a variância de $\hat{\boldsymbol{\mu}}$ apenas para o caso em que os parâmetros

de sensibilidade são os elementos inidentificáveis de $\boldsymbol{\mu}_{(\mathbf{r})}$. Primeiramente, note-se que, com alguma álgebra, (5.10) e (5.11) podem ser convenientemente reescritos em formulação matricial, útil para implementação computacional, nomeadamente,

$$\begin{aligned}\boldsymbol{\mu} &= \boldsymbol{\gamma}_*'\boldsymbol{\mu}_{(*)},\\ \boldsymbol{\Sigma} &= \boldsymbol{\gamma}_*'\boldsymbol{\Sigma}_{(*)}(\mathbf{1}_P \otimes \mathbf{I}_J) + \boldsymbol{\gamma}_*'\mathbf{S}_{(*)}(\mathbf{1}_P \otimes \mathbf{I}_J),\end{aligned}$$

em que $\boldsymbol{\gamma}_* = \boldsymbol{\gamma} \otimes \mathbf{I}_J$, $\boldsymbol{\gamma} = (\gamma_{\mathbf{r}}, \mathbf{r} \in \mathcal{R})'$, $\otimes$ denota o produto de Kronecker, $\mathbf{I}_J$ representa uma matriz identidade de posto J, $\boldsymbol{\mu}_{(*)} = (\boldsymbol{\mu}_{(\mathbf{r})}', \mathbf{r} \in \mathcal{R})'$, $\boldsymbol{\Sigma}_{(*)}$ e $\mathbf{S}_{(*)}$ são matrizes diagonais em blocos com blocos iguais a $\boldsymbol{\Sigma}_{(\mathbf{r})}$ e $\mathbf{S}_{(\mathbf{r})} = (\boldsymbol{\mu}_{(\mathbf{r})} - \boldsymbol{\mu})(\boldsymbol{\mu}_{(\mathbf{r})} - \boldsymbol{\mu})'$, $\mathbf{r} \in \mathcal{R}$, respectivamente, $\mathbf{1}_P$ denota um vetor $P \times 1$ com todos os elementos iguais a 1 e P representa o número de padrões de omissão, *i.e.*, a cardinalidade de $\mathcal{R}$. Ao empregar como parâmetros de sensibilidade os componentes inidentificáveis de $\boldsymbol{\mu}_{(\mathbf{r})}$, empilhados no vetor $\boldsymbol{\mu}^{\mathrm{IN}}$, a matriz de covariância de $\hat{\boldsymbol{\mu}}$ é especificada como

$$\mathrm{Cov}\left[\hat{\boldsymbol{\mu}}(\boldsymbol{\mu}^{\mathrm{IN}})\right] = \frac{1}{n}\boldsymbol{\gamma}_*'\boldsymbol{\Sigma}_{(*)}^{I}(\mathbf{1}_P \otimes \mathbf{I}_J) + \frac{1}{n}\boldsymbol{\mu}_D'[(\mathbf{D}_{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\boldsymbol{\gamma}') \otimes (\mathbf{1}_J\mathbf{1}_J')]\boldsymbol{\mu}_D,$$

em que $\boldsymbol{\mu}_D = (\mathbf{D}_{\boldsymbol{\mu}_{(\mathbf{r})}}, \mathbf{r} \in \mathcal{R})'$ e $\boldsymbol{\Sigma}_{(*)}^{I}$ é obtido de $\boldsymbol{\Sigma}_{(*)}$ por substituição dos parâmetros inidentificáveis por 0. Ao considerarem-se outras parametrizações para os parâmetros de sensibilidade, recorre-se ao *bootstrap* não-paramétrico para obter estimativas para os erros padrões de $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\sigma}}$, $\hat{\boldsymbol{\Psi}}$ e funções deles.

Para avaliar a questão (ii) da Seção 5.1, as conjeturas arbitrárias sobre os parâmetros inidentificáveis são que

1. as médias não-observadas de GRADRAT e das outras variáveis podem variar, respectivamente, entre os quantis 30%–50% e 20%–50% das correspondentes distribuições observadas,
2. os desvios padrões não-observados de todas as variáveis podem estar entre 0.80 e 1.25 vezes os desvios padrões das correspondentes distribuições observadas e
3. as correlações lineares não-observadas de (a) RMBRD *vs.* as outras três variáveis podem variar de 0.0 a 0.5, (b) GRADRAT *vs.* CSAT e GRADRAT *vs.* ACT podem estar entre 0.2 e 0.8 e (c) CSAT *vs.* ACT podem variar de 0.5 a 1.0.

As reparametrizações subjacentes empregadas para os parâmetros de sensibilidade são $\mu_{j(\mathbf{r})} = F_{j(1)}^{-1}(p_{j(\mathbf{r})})$ e $\sigma_{j(\mathbf{r})} = \lambda_{j(\mathbf{r})}\sigma_{j(1)}$, $\forall \mathbf{r}$ tal que $r_j = 0$, e as suposições sugeridas são

1. $\boldsymbol{\Omega}_{p_{j(\mathbf{r})}}$ é $= [0.3; 0.5]$ para $j = 1$ e $= [0.2; 0.5]$ para $j = 2, 3, 4$, $\forall \mathbf{r}$ tal que $r_j = 0$,
2. $\boldsymbol{\Omega}_{\lambda_{j(\mathbf{r})}} = [0.80; 1.25]$, $\forall j, \mathbf{r}$ tal que $r_j = 0$,
3. $\boldsymbol{\Omega}_{\psi_{jk(\mathbf{r})}}$, $\forall \mathbf{r}$ tal que $r_j = 0$ e/ou $r_k = 0$, é (a) $= [0.0; 0.5]$ para $j = 1, 2, 3$, $k = 4$, (b) $= [0.2; 0.8]$ para $j = 1$, $k = 2, 3$, e (c) $= [0.5; 1.0]$ para $j = 2$, $k = 3$.

Na Tabela 5.4, exibem-se estimativas para os intervalos de ignorância e de incerteza para médias, desvios padrões e correlações das quatro variáveis. Conclui-se que as variáveis de cada par são

positivamente correlacionadas (linearmente); as magnitudes das correlações, contudo, são difíceis de se avaliar, devido à ignorância causada pelos dados faltantes. Conforme esperado, os intervalos são mais extensos para pares de variáveis que tiveram simultaneamente mais dados omissos (vide Tabela 5.2). Análises de casos completos e de casos disponíveis também apontam para associações positivas entre as variáveis, mas como elas não consideram a incerteza causada pelos dados faltantes, seus intervalos de confiança a 95% (não apresentados) são muito mais estreitos que qualquer das RInE.

Tabela 5.4: RIgE e RInE a 95% para médias, desvios padrões (DP) e correlações de GRADRAT (G), CSAT (C), ACT (A) e RMBRD (R).

Variável	Parâmetro	RIgE	RInE Fraca	RInE Ponto a Ponto	RInE Forte
GRADRAT	Média	$[59.6; 60.3]$	$[58.8; 61.1]$	$[58.7; 61.2]$	$[58.5; 61.4]$
	DP	$[18.7; 19.3]$	$[18.3; 19.7]$	$[18.1; 19.9]$	$[18.0; 20.0]$
CSAT	Média	$[926.7; 963.6]$	$[924.4; 965.8]$	$[918.9; 971.1]$	$[917.4; 972.5]$
	DP	$[114.7; 146.0]$	$[112.8; 148.3]$	$[108.7; 153.3]$	$[107.6; 154.7]$
ACT	Média	$[21.16; 22.07]$	$[21.17; 22.06]$	$[21.08; 22.15]$	$[21.06; 22.16]$
	DP	$[2.43; 2.95]$	$[2.38; 3.01]$	$[2.31; 3.10]$	$[2.28; 3.13]$
RMBRD	Média	$[3.75; 4.09]$	$[3.73; 4.11]$	$[3.69; 4.15]$	$[3.68; 4.17]$
	DP	$[1.09; 1.38]$	$[1.08; 1.40]$	$[1.04; 1.44]$	$[1.03; 1.46]$
G×C	Corr.	$[0.37; 0.69]$	$[0.38; 0.69]$	$[0.35; 0.71]$	$[0.34; 0.72]$
G×A	Corr.	$[0.31; 0.66]$	$[0.31; 0.66]$	$[0.28; 0.69]$	$[0.28; 0.69]$
G×R	Corr.	$[0.24; 0.52]$	$[0.24; 0.52]$	$[0.22; 0.55]$	$[0.21; 0.55]$
C×A	Corr.	$[0.59; 0.95]$	$[0.60; 0.95]$	$[0.57; 0.97]$	$[0.57; 0.97]$
C×R	Corr.	$[0.11; 0.55]$	$[0.12; 0.54]$	$[0.08; 0.57]$	$[0.07; 0.58]$
A×R	Corr.	$[0.05; 0.46]$	$[0.06; 0.46]$	$[0.02; 0.49]$	$[0.02; 0.50]$

5.4 Avaliação dos intervalos de incerteza

Para construir RInE, Vansteelandt *et al.* (2006) supõem que os valores dos parâmetros de sensibilidade que correspondem aos limites inferior e superior do intervalo de ignorância são independentes dos dados observados. Essa suposição é satisfeita quando o alvo é a média, mas ela pode não ser satisfeita para desvios padrões e correlações. Considere-se, por exemplo, o alvo σ no caso univariado da Seção 5.2 e admita que $\mathbf{\Omega}_{\sigma_{(0)}} = [\sigma^L_{(0)}; \sigma^U_{(0)}]$ e $\mathbf{\Omega}_{\mu_{(0)}} = [\mu^L_{(0)}; \mu^U_{(0)}]$ são especificados. Depois de substituir os parâmetros identificáveis em (5.2) pelas suas contrapartidas amostrais, note-se que $\sigma^L_{(0)}$ e $\sigma^U_{(0)}$ são os valores de $\sigma_{(0)}$ no conjunto $\mathbf{\Omega}_{\sigma_{(0)}}$ que, respectivamente, minimiza e maximiza $\hat{\sigma}(\mu_{(0)}, \sigma_{(0)})$ independentemente dos dados e de $\mu_{(0)}$. Contudo, como

$$\underset{\mu_{(0)} \in \mathbf{\Omega}_{\mu_{(0)}}}{\arg\min}\ \hat{\sigma}(\mu_{(0)}, \sigma_{(0)}) = \begin{cases} \mu^L_{(0)}, & \text{se } \hat{\mu}_{(1)} < \mu^L_{(0)}, \\ \mu^U_{(0)}, & \text{se } \hat{\mu}_{(1)} > \mu^U_{(0)}, \\ \hat{\mu}_{(1)}, & \text{em caso contrário,} \end{cases}$$

$$\underset{\mu_{(0)}\in\mathbf{\Omega}_{\mu_{(0)}}}{\arg\max}\ \hat{\sigma}(\mu_{(0)},\sigma_{(0)}) = \begin{cases} \mu_{(0)}^{U}, & \text{se } (\hat{\mu}_{(1)}-\mu_{(0)}^{U})^2 > (\hat{\mu}_{(1)}-\mu_{(0)}^{L})^2, \\ \mu_{(0)}^{L}, & \text{em caso contrário} \end{cases}$$

claramente dependem dos dados por meio de $\hat{\mu}_{(1)}$, a suposição é violada. Observa-se entretanto que, para este caso específico com apenas dois padrões de omissão, a suposição é satisfeita para desvios padrões se o parâmetro de sensibilidade for trocado de $\mu_{(0)}$ por α ou β. Para as diferenças de médias e as razões de desvios padrões consideradas na Seção 5.2, a suposição é satisfeita quando há um conjunto de parâmetros de sensibilidade para cada nível de administração de faculdade e, adicionalmente para a razão de desvios padrões, se a parametrização com α ou β for empregada; em caso contrário, a suposição é violada.

Vansteelandt *et al.* (2006) apresentam outro exemplo em que a suposição falha e, por conseguinte, obtêm estimativas *bootstrap* para avaliar a probabilidade de cobertura dos intervalos de incerteza. Seguindo suas ideias, reamostra-se o conjunto de dados e repetem-se as análises $B = 5\,000$ vezes. As estimativas das probabilidades de cobertura são:

1. para a RInE forte, a proporção das B RInE fortes que contêm a RIgE da análise original;
2. para a RInE ponto a ponto, o mínimo entre a proporção das B RInE ponto a ponto que contêm o limite inferior da RIgE da análise original e a proporção correspondente que contém o limite superior;
3. para a RInE fraca, a média dos B comprimentos das intersecções entre as RInE fracas e a RIgE da análise original dividida pelo comprimento dessa RIgE.

Avaliam-se também as coberturas para os casos em que os erros padrões e os intervalos de incerteza são estimados em escalas transformadas, para os quais se espera que a normalidade assintótica, suposta para a construção das RInE, seria uma boa aproximação; posteriormente, as RInE são transformadas de volta para as escalas originais. Considera-se a função logito para a média de GRADRAT, o logaritmo tanto para a média de RMBRD quanto para todos os desvios padrões e para a razão dos desvios padrões e a transformação-z de Fisher para todas as correlações, *i.e.*, $0.5\log[(1+\psi_{jk})/(1-\psi_{jk})]$.[2]

Apresentam-se as estimativas de cobertura das RInE para as análises multivariadas e univariadas consideradas nas seções anteriores nas Tabelas 5.5 e 5.6. Para a análise multivariada (Tabela 5.5), as escalas transformadas não fazem com que as RInE fiquem mais próximas do nível nominal de 95% e, em alguns casos, os resultados são até piores do que os obtidos na escala original. Por outro lado, a cobertura para as RInE da razão de desvios padrões (Tabela 5.6) é muito melhor na escala logarítmica quando não se considera a redução no número de parâmetros de sensibilidade. Com exceção das RInE ponto a ponto e forte para a média de RMBRD, para as correlações de GRADRAT e CSAT *vs.* ACT

[2]Essas RInE não são apresentadas porque diferem por zero ou apenas uma única unidade do último dígito dos resultados apresentados para as escalas originais nas Tabelas 5.3 e 5.4 e, assim, não alteram qualquer das conclusões.

e para a razão de desvios padrões, as coberturas não são tão distantes de 95%. As RInE fracas são em geral muito mais próximas do nível nominal do que as outras duas alternativas. É interessante notar que, embora Vansteelandt *et al.* (2006) afirmem que a RInE fraca, contrariamente às RInE ponto a ponto e forte, não podem geralmente ser estimadas em escalas monotonamente transformadas enquanto retêm o nível de cobertura original, para estas análises as coberturas das RInE fraca são, em geral, mais próximas do nível nominal nos casos para os quais a escala transformada também melhora a cobertura das RInE ponto a ponto e forte.

Tabela 5.5: Estimativas *bootstrap* de cobertura (×100) para RInE a 95% de médias, desvios padrões (DP) e correlações de GRADRAT (G), CSAT (C), ACT (A) e RMBRD (R) para as análises da Tabela 5.4.

Variável	Parâmetro	RInE Fraca		RInE Ponto a Ponto		RInE Forte	
		original	transf.[a]	original	transf.[a]	original	transf.[a]
GRADRAT	Média	93.9	94.3	93.6	94.1	93.5	93.9
	DP	94.8	94.7	94.4	94.6	94.7	94.4
CSAT	Média	95.0		94.4		94.2	
	DP	95.6	95.4	95.0	95.1	96.0	95.4
ACT	Média	95.3		95.7		96.8	
	DP	95.0	94.7	93.6	93.5	94.7	93.8
RMBRD	Média	93.6	93.7	88.6	88.5	89.8	89.8
	DP	94.7	94.6	92.8	93.5	94.0	93.8
G×C	Corr.	95.8	95.9	96.4	95.4	96.8	95.8
G×A	Corr.	93.7	94.0	89.7	90.7	91.7	90.8
G×R	Corr.	94.9	95.1	94.3	93.8	94.4	94.4
C×A	Corr.	95.1	96.4	86.4	82.0	90.8	86.5
C×R	Corr.	95.5	95.7	95.8	95.8	95.8	95.8
A×R	Corr.	95.9	96.1	94.2	94.0	96.4	96.4

[a]Escalas transformadas: logito para a média de GRADRAT, logaritmo para a média de RMBRD e todos os DP e transformação-z de Fisher para todas as correlações.

Tabela 5.6: Estimativas *bootstrap* de cobertura (×100) para RInE a 95% de $\mu_{(\mathrm{PRI})} - \mu_{(\mathrm{PUB})}$ e $\sigma_{(\mathrm{PRI})}/\sigma_{(\mathrm{PUB})}$ de CSAT para as análises da Tabela 5.3.

Parâmetro	Par. Sens.[a]	Escala	RInE Fraca	RInE Ponto a Ponto	RInE Forte
$\mu_{(\mathrm{PRI})} - \mu_{(\mathrm{PUB})}$	individual	original	94.5	93.2	92.7
	comum	original	95.0	94.3	93.5
$\frac{\sigma_{(\mathrm{PRI})}}{\sigma_{(\mathrm{PUB})}}$	individual	original	92.6	82.5	85.0
		logaritmo	94.7	94.3	94.3
	comum	original	94.4	94.2	94.5
		logaritmo	94.7	94.2	94.6

[a]Parâmetros de sensibilidade individuais e comuns para os níveis de administração de faculdade.

Tabela 5.7: Estimativas *bootstrap* de cobertura ($\times 100$) para RInE a 95% de $\mu_{(\mathrm{PRI})} - \mu_{(\mathrm{PUB})}$ e $\sigma_{(\mathrm{PRI})}/\sigma_{(\mathrm{PUB})}$ de CSAT.

Parâmetro	Par. Sens.[a]	Escala	RInE Fraca	RInE Ponto a Ponto	RInE Forte
$\mu_{(\mathrm{PRI})} - \mu_{(\mathrm{PUB})}$	$\mu_{(0)}$	original	95.0	95.1	95.3
	α	original	95.0	94.6	94.9
	β	original	95.0	94.6	94.8
	p	original	94.5	93.2	92.7
$\frac{\sigma_{(\mathrm{PRI})}}{\sigma_{(\mathrm{PUB})}}$	$(\mu_{(0)}, \sigma_{(0)})$	original	93.7	83.1	84.8
		logaritmo	94.9	94.4	94.8
	$(\mu_{(0)}, \lambda)$	original	92.9	82.7	85.1
		logaritmo	94.8	94.6	94.7
	$(\alpha, \sigma_{(0)})$	original	93.7	81.5	84.2
		logaritmo	94.8	94.1	94.2
	(α, λ)	original	92.9	82.6	84.7
		logaritmo	94.8	94.5	94.6
	$(\beta, \sigma_{(0)})$	original	93.7	81.5	84.1
		logaritmo	94.8	94.1	94.2
	(β, λ)	original	92.9	82.7	84.7
		logaritmo	94.8	94.6	94.6
	$(p, \sigma_{(0)})$	original	93.2	81.4	83.7
		logaritmo	94.5	94.2	93.7
	(p, λ)	original	92.6	82.5	85.0
		logaritmo	94.7	94.3	94.3

[a]Um destes parâmetros de sensibilidade para cada nível de administração de faculdade.

Na Tabela 5.7, mostram-se as estimativas de cobertura para as RInE correspondentes à análise univariada variando as parametrizações para o caso de um conjunto de parâmetros de sensibilidade para cada nível de administração da faculdade. Os conjuntos de valores para os parâmetros de sensibilidade são escolhidos de tal forma que as RIgE coincidam com as correspondentes da Tabela 5.3. Ao comparar os resultados das Tabelas 5.5, 5.6 e 5.7, não há uma evidência clara de que as coberturas estão mais próximas do nível nominal quando os valores dos parâmetros de sensibilidade que correspondem aos limites inferior e superior do intervalo de ignorância são independentes dos dados observados. De fato, os resultados podem ser mais influenciados pela qualidade da aproximação normal para as distribuições dos estimadores, que por sua vez dependem da escolha da parametrização do parâmetro de sensibilidade. Para ambas as funções paramétricas na Tabela 5.7, as coberturas de todas as RInE são as mais próximas do nível nominal para $\mu_{(0)}$, as mais distantes de 95% para p e muito similares para α e β. Para $\sigma_{(\mathrm{PRI})}/\sigma_{(\mathrm{PUB})}$, as coberturas estão um pouco mais próximas do nível nominal quando $\sigma_{(0)}$ é escolhido em detrimento de λ para as RInE fracas quando se emprega $\mu_{(0)}$, α, β ou p, e também para as RInE ponto a ponto obtidas na escala original com $\mu_{(0)}$; os papéis de $\sigma_{(0)}$ e λ invertem-se para todas as RInE fortes e para as outras RInE ponto a ponto.

5.5 Discussão

Modelos de seleção e modelos de mistura de padrões são provavelmente as estruturas mais comuns na modelagem de dados incompletos. No caso univariado, Scharfstein, Daniels & Robins (2003) mostraram que a suposição

$$\text{logito } P(R_i = 0|Y_i = y) = \text{constante} + q(y) \tag{5.12}$$

sob o modelo de seleção, em que q é a chamada função de viés de seleção, é equivalente à restrição

$$f_{(0)}(y) = f_{(1)}(y)\frac{\exp[q(y)]}{\int_{-\infty}^{\infty}\exp[q(s)]f_{(1)}(s)ds}, \ \forall y, \tag{5.13}$$

sob o modelo de mistura de padrões. Além da interpretação direta dessas expressões, eles também observaram que, se $q(y) = \delta\log(y)$, por exemplo, segue de (5.12) que $\exp(\delta)$ é a razão de chances de omissão entre indivíduos que diferem por uma unidade de $\log(y)$; de (5.13), conclui-se que $\delta > 0$ (< 0) indica que a distribuição de Y_i para respostas faltantes é mais (menos) fortemente ponderada na direção de valores grandes de Y_i do que a distribuição de Y_i para respostas observadas. Esses discernimentos são fundamentais para levar a cabo uma análise de sensibilidade. Todavia, tanto a forma funcional de $q(y)$ quanto a faixa de valores considerados para δ são difíceis de se avaliar.

Quando o alvo de inferência é a média, o desvio padrão, a correlação ou alguma função deles, pode-se empregar suas contrapartidas inidentificáveis como parâmetros de sensibilidade. Estes parâmetros de sensibilidade são mais fáceis de se eliciar do que as funções de viés de seleção porque as primeiras são diretamente relacionadas com os parâmetros de interesse. Entretanto, há conexões entre ambas as estratégias; por exemplo, para uma $q(y)$ especificada, pode-se usar (5.13) para calcular os resultados correspondentes para $\mu_{(0)}$, α, β, p, $\sigma_{(0)}$ e λ considerados na Seção 5.2.

Algumas dessas ideias de parametrização foram previamente consideradas na literatura. Por exemplo, Rubin (1977) usa (5.1) para desenvolver uma solução bayesiana supondo normalidade e Daniels & Hogan (2000) consideram um modelo de mistura de padrões de distribuições normais multivariadas em que a identificação do modelo é realizada por meio de $\mathbf{b}_{(d)} = \boldsymbol{\mu}_{(d)} - \boldsymbol{\mu}_{(d+1)}$ e $\mathbf{C}_{(d)} = \boldsymbol{\Sigma}_{(d)}^{1/2}\boldsymbol{\Sigma}_{(d+1)}^{-1/2}$, $d = 1, \ldots, J$, em que $d = 1 + \sum_{j=1}^{J} r_j$ é o indicador de abandono e $\mathbf{b}_{(d)}$ e $\mathbf{C}_{(d)}$ são, respectivamente, vetores e matrizes pré-especificados. Neste capítulo estendeu-se esses resultados eliminando a suposição de normalidade e permitindo uma maior flexibilidade na identificação do modelo. Primeiro, não se consideraram apenas diferenças absolutas de médias de padrões de omissão mas também diferenças relativas e a possibilidade de relacionar médias não-observadas com quantis das distribuições observadas. Segundo, substituíram-se as funções de matrizes de covariância de difícil eliciação pelas diferenças relativas de desvios padrões e pelas correlações. Terceiro, no caso multivariado, mostrou-se que uma média (ou um desvio padrão) inidentificável pode ser relacionada com uma (um) identificável não apenas em casos de um único padrão de omissão mas também em

casos com conjuntos de padrões de omissão em que a variável correspondente é observada. Com essas alternativas, é mais fácil extrair informação de especialistas ou de dados históricos e, consequentemente, produzir análises de sensibilidade mais plausíveis. Quando o interesse recai apenas em funções das médias, uma vantagem desta abordagem é que não é necessário especificar valores para parâmetros de sensibilidade de desvios padrões (e correlações), ao contrário de Daniels & Hogan (2000) que têm que identificar as distribuições normais multivariadas e, como consequência, mostram que os desvios padrões *a posteriori* das funções das médias são influenciados pela escolha desses parâmetros de sensibilidade. Todavia, uma desvantagem da abordagem apresentada aqui é que não há uma forma de se fixar os parâmetros de sensibilidade num valor que corresponda à suposição MAR.

Capítulo 6

Considerações finais

Apesar da vastidão de trabalhos que se debruçam sobre a análise de dados com omissão, não existe uma abordagem universalmente aceita e, provavelmente, nunca existirá. O campo deve continuar a expandir-se exponencialmente, porque além de os modelos que continuam a ser propostos para análises de conjuntos de dados completos necessitarem de desenvolvimentos adicionais e possuirem particularidades que devem ser estudadas para serem utilizados em análises de dados com omissão, também as escolhas realizadas para se fazer inferências são em geral mais abundantes e muito mais controversas nas análises de dados faltantes. Resume-se as principais contribuições e conclusões deste trabalho na Seção 6.1 e descreve-se algumas sugestões de temas para pesquisas futuras na Seção 6.2.

6.1 Contribuições e conclusões deste trabalho

Mostrou-se que os modelos sobreparametrizados e parâmetros inidentificáveis, obtidos em uma série de campos da estatística, possuem as seguintes especificidades:

- não é apenas o uso de um modelo identificável incorreto que pode levar a inferências de interesse equivocadas; a adoção de modelos sobreparametrizados parcimoniosos errados também é um fator que precisa ser levado em consideração;
- distribuições *a priori* têm um impacto muito maior nas inferências sobre parâmetros inidentificáveis e, por isso, escolhas usualmente consideradas não-informativas ou levemente informativas podem na verdade ser mais informativas do que se imagina;
- consequentemente, inclusões de informação, seja por meio do modelo estrutural ou das distribuições *a priori*, devem ser cautelosamente contempladas e nunca encaradas arbitrariamente;
- a incerteza estatística, seja ela medida pelo comprimento do intervalo de incerteza da inferência clássica ou pelo desvio padrão *a posteriori* ou tamanho do intervalo de credibilidade da inferência bayesiana, não desaparece à medida que o tamanho da amostra tende para infinito;
- os testes de hipóteses têm um poder muito baixo em certas situações sendo, portanto, inadmissível considerar a não-rejeição de uma hipótese como sua aceitação para qualquer tamanho

de amostra;

- as cadeias de distribuições *a posteriori* geradas via métodos de Monte Carlo baseados em cadeias de Markov tornam-se altamente autocorrelacionadas à medida que o tamanho de amostra cresce e as distribuições *a priori* são mais difusas, fazendo com que seja mais difícil constatar a convergência e garantir a precisão desejada nas inferências de interesse.

Quando variáveis explicativas também sofrem omissão, necessita-se especificar um modelo para elas mesmo que se tenha interesse apenas no modelo condicional das variáveis respostas dadas estas covariáveis. Explorando o caso de respostas binárias, mostrou-se que escolhas distribucionais incorretas para o primeiro modelo podem fazer com que se obtenha resultados enviesados para o último e que isto pode ser evitado ao utilizar-se um modelo não-paramétrico bayesiano via mistura por processo de Dirichlet para as covariáveis. Apesar da flexibilidade desta abordagem e da possibilidade de combiná-la com modelos de omissão informativa e não-informativa, o enfoque empregado de se aproximar o processo de Dirichlet por sua versão truncada pode ser inviável para tamanhos de amostra muito grandes (*e.g.*, $n > 10\,000$).

Não é raro que o interesse de uma análise recaia apenas em momentos de uma distribuição. Nesse contexto, apresentou-se uma nova análise de sensibilidade clássica para a estimação destes parâmetros quando há respostas incompletas unindo duas ideias que até o momento apareciam separadamente na literatura estatística: (i) evitar a adoção de suposições sobre a forma distribucional paramétrica das respostas e (ii) empregar parâmetros de sensibilidade de fácil interpretação diretamente relacionados com os parâmetros de interesse. Essa abordagem é principalmente atrativa na análise de dados contínuos ao evitar a recorrente suposição de normalidade. Explorou-se algumas alternativas de parâmetros de sensibilidade, dentre as quais se destaca a que relaciona a média dos dados não-observados com quantis da distribuição dos dados observados, dado que esta estratégia não parece explorada na literatura.

6.2 Sugestões de temas para pesquisas futuras

Algumas possibilidades de extensões para a análise bayesiana semiparamétrica já foram elencadas na Seção 4.6. Tais sugestões também podem ser enfocadas sob o ponto de vista clássico.

Zhang (2010) propõe que se use uma abordagem de verossimilhança perfilada para representar e interpretar a evidência estatística em análises de dados incompletos sem utilizar suposições inverificáveis. Seria interessante investigar como os resultados dessas análises se comparam com os das análises bayesianas e das análises de sensibilidades clássicas de Vansteelandt *et al.* (2006) exploradas no Capítulo 3.

Poleto, Singer & Paulino (2011c) estenderam a modelagem de Paulino (1991) referente à distribuição multinomial para a distribuição produto de multinomiais possibilitando a inclusão de variáveis explicativas sempre observadas na análise de respostas com ou sem omissão em tabelas

de contingência. Estes autores utilizaram a abordagem de máxima verossimilhança para o ajuste de modelos estritamente lineares e log-lineares e a metodologia de mínimos quadrados generalizados para o ajuste de modelos funcionais lineares mais gerais. Os resultados foram desenvolvidos em formulação matricial adequada para a implementação computacional, realizada com a construção de uma biblioteca para o pacote estatístico R com a finalidade de facilitar o traçado destas inferências. Extensão e implementação similares podem ser realizadas para as análises bayesianas descritas no Capítulo 3. Em ambos os paradigmas inferenciais pode-se também contemplar casos de omissão nas variáveis explicativas.

Com relação às abordagens clássicas, quando é muito difícil trabalhar com a verossimilhança, como no caso de dados correlacionados, pode ser mais fácil modificar quer a verossimilhança quer as equações de estimação para o traçado de certas inferências. Quando há dados com omissão, tais análises em geral necessitam correções para resultarem em inferências consistentes se o mecanismo não for completamente aleatório. Um trabalho importante sobre este assunto é o de Robins, Rotnitzky & Zhao (1995), em que se propõe uma ponderação para as equações de estimação generalizadas que, juntamente com as extensões e refinamentos de trabalhos posteriores, permitem traçar inferências não apenas em mecanismos de omissão aleatória, mas também naqueles com omissão informativa. No caso de omissão aleatória, desenvolveram-se estimadores que possuem uma atraente propriedade denominada "robustez dupla": eles são consistentes tanto se o modelo para o mecanismo de omissão ou o modelo para os dados completos forem corretamente especificados (mas não necessariamente se ambos estiverem incorretos). Para mais detalhes, veja a revisão de Rotnitzky (2008). Molenberghs, Kenward, Verbeke & Birhanu (2011) adaptam essas ideias para efetuar correções em pseudo-verossimilhanças sob omissão ignorável. Há campo para estender muitos desses trabalhos para casos com omissão informativa; também frequentemente não são contempladas análises de sensibilidade.

Referências bibliográficas

Agresti, A. & Min, Y. (2005). Frequentist performance of Bayesian confidence intervals for comparing proportions in 2×2 contingency tables. *Biometrics* **61**, 515–523. doi: 10.1111/j.1541-0420.2005.031228.x. 32

Allison, P. D. (2001). *Missing Data*. Thousand Oaks: Sage. 3, 68

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics* **2**, 1152–1174. doi: 10.1214/aos/1176342871. 49, 50

Baker, S. G. & 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). 11

Baker, S. G., Rosenberger, W. F. & DerSimonian, R. (1992). Closed-form estimates for missing counts in two-way contingency tables. *Statistics in Medicine* **11**, 643–657. doi: 10.1002/sim.4780110509. 2, 42

Basu, D. (1977). On the elimination of nuisance parameters. *Journal of the American Statistical Association* **72**, 355–366. 10

Bedrick, E. J., Christensen, R. & Johnson, W. (1996). A new perspective on priors for generalized linear models. *Journal of the American Statistical Association* **91**, 1450–1460. 30

Blackwell, D. & MacQueen, J. B. (1973). Ferguson distributions via polya urn schemes. *The Annals of Statistics* **1**, 353–355. doi: 10.1214/aos/1176342372. 51

Chambers, R. L. & Skinner, C. J. (2003). *Analysis of Survey Data*. New York: Wiley. 3

Chen, H. Y. (2002). Double-semiparametric method for missing covariates in Cox regression models. *Journal of the American Statistical Association* **97**, 565–575. doi: 10.1198/016214502760047096. 45

Chen, H. Y. (2004). Nonparametric and semiparametric models for missing covariates in parametric regression. *Journal of the American Statistical Association* **99**, 1176–1189. doi: 10.1198/016214504000001727. 45, 46

Chen, H. Y. (2009). Estimation and inference based on neumann series approximation to locally efficient score in missing data problems. *Scandinavian Journal of Statistics* **36**, 713–734. doi: 10.1111/j.1467-9469.2009.00646.x. 45

Chen, H. Y. & Little, R. J. A. (1999). Proportional hazards regression with missing covariates. *Journal of the American Statistical Association* **94**, 896–908. 45

Clarke, P. S. (2002). On boundary solutions and identifiability in categorical regression with non-ignorable non-response. *Biometrical Journal* **44**, 701–717. doi: 10.1002/1521-4036(200209)44:6<701::AID-BIMJ701>3.0.CO;2-1. 11

Clarke, P. S. & 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. doi: 10.1111/j.1369-7412.2003.04973.x. 11

Congdon, P. (2005). *Bayesian Models for Categorical Data*. New York: Wiley. 3

Congdon, P. (2006). *Bayesian Satistical Modelling*, 2ª ed. New York: Wiley. 3, 48, 51

Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **48**, 133–169. 19

Copas, J. B. & Li, H. G. (1997). Inference for non-random samples (with discussion). *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **59**, 55–95. doi: 10.1111/1467-9868.00055. 21

Creemers, A., Hens, N., Aerts, M., Molenberghs, G., G., V. & Kenward, M. G. (2010). A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Biometrical Journal* **52**, 111–125. doi: 10.1002/bimj.200800235. 19

Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G. & Kenward, M. G. (2009). Shared-parameter models and missingness at random. *Submetido para publicação*. 15, 16, 17

Daniels, M. J. & Hogan, J. W. (2000). Reparameterizing the pattern mixture model for sensitivity analyses under informative dropout. *Biometrics* **56**, 1241–1248. doi: 10.1111/j.0006-341X.2000.01241.x. 67, 81, 82

Daniels, M. J. & Hogan, J. W. (2007). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. London: Chapman & Hall / CRC. 3, 21, 67

De Leeuw, J. & Meijer, E. (2008). *Handbook of Multilevel Analysis*. New York: Springer. 3

Dickey, J. M. (1983). Multiple hypergeometric functions: probabilistic interpretations and statistical uses. *Journal of the American Statistical Association* **78**, 628–637. 25

Diggle, P. J., Heagerty, P., Liang, K.-Y. & Zeger, S. L. (2002). *Analysis of Longitudinal Data*, 2ª ed. Oxford: Oxford University Press. 3

Diggle, P. J. & Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis (with discussion). *Journal of the Royal Statistical Society. Series C: Applied Statistics* **43**, 49–93. 11, 44

Efron, B. & Gong, G. (1983). A leisure look at the bootstrap, the jackknife and cross-validation. *The American Statistician* **37**, 36–48. 72

Escobar, M. D. & West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**, 577–588. 50, 51, 53, 64

Escobar, M. D. & West, M. (1998). Computing nonparametric hierarchical models. Em D. Dey, P. Müller & D. Sinha, eds., *Practical Nonparametric and Semiparametric Bayesian Statistics (Lecture Notes in Statististics 133)*. New York: Springer, 1–22. 51

Fay, R. E. (1986). Causal models for patterns of nonresponse. *Journal of the American Statistical Association* **81**, 354–365. 40

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**, 209–230. doi: 10.1214/aos/1176342360. 48

Fitzmaurice, G., Davidian, M., Verbeke, G. & Molenberghs, G. (2008). *Longitudinal Data Analysis*. Boca Raton: Chapman & Hall. 3, 67

Fitzmaurice, G. M., Laird, N. M. & Ware, J. H. (2004). *Applied Longitudinal Analysis*. New York: Wiley. 3

Fleiss, J. L., Levin, B. & Paik, M. C. (2003). *Statistical Methods for Rates and Proportions*, 3ª ed. New York: Wiley. 3

Forster, J. J. & Smith, P. W. F. (1998). Model-based inference for categorical survey data subject to non-ignorable non-response (with discussion). *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **60**, 57–70, 89–102. doi: 10.1111/1467-9868.00108,doi: 10.1111/1467-9868.00110. 22, 33

Gelman, A. & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press. 3

Gelman, A. & Meng, X.-L. (2004). *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*. New York: Wiley. 3

Gelman, A. & Rubin, D. B. (1992). A single series from the gibbs sampler provides a false sense of security. Em J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, eds., *Bayesian Statistics 4*. Oxford: Oxford University Press, 625–631. 57, 59

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion). Em J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, eds., *Bayesian Statistics 4*. Oxford: Oxford University Press, 169–193. 32, 57

Gibbons, L. E. & Hosmer, D. W. (1991). Conditional logistic regression with missing data. *Communications in Statistics - Simulation and Computation* **20**, 109–120. doi: 10.1080/03610919108812942. 45

Glonek, G. F. V. (1999). On identifiability in models for incomplete binary data. *Statistics & Probability Letters* **41**, 191–197. doi: 10.1016/S0167-7152(98)00142-4. 11

Glynn, R. J., Laird, N. M. & Rubin, D. B. (1986). Selection modeling versus mixture modeling with nonignorable nonresponse (with discussion). Em H. Wainer, ed., *Drawing Inferences from Self-Selected Samples*. Mahwah: Lawrence Erlbaum Associates, 115–151. 7, 23, 67

Grab, E. L. & Savage, I. R. (1954). Tables of the expected value of 1/X for positive Bernoulli and Poisson variables. *Journal of the American Statistical Association* **49**, 169–177. 71

Griffin, J. & Holmes, C. (2010). Computational issues arising in Bayesian nonparametric hierarchical models. Em N. L. Hjort, C. Holmes, P. Müller & S. G. Walker, eds., *Bayesian Nonparametrics*. Cambridge: Cambridge University Press, 208–222. 52, 65

Groves, R. M., Dillman, D. A., Eltinge, J. L. & Little, R. J. A. (2002). *Survey Nonresponse*. New York: Wiley. 3

Gustafson, P. (2005). On model expansion, model contraction, identifiability and prior information: two illustrative scenarios involving mismeasured variables (with discussion). *Statistical Science* **20**, 111–140. doi: 10.1214/088342305000000098. 21

Hagenaars, J. A. & McCutcheon, A. L. (2002). *Applied Latent Class Analysis*. Cambridge: Cambridge University Press. 3

Halloran, M. E. & Berry, D. (1999). *Statistical Models in Epidemiology, the Environment and Clinical Trials*. New York: Springer. 3

Harrell, F. E. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression and Survival Analysis*. New York: Springer. 3

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* **5**, 475–492. 6

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica* **47**, 153–161. 6

Hedeker, D. & Gibbons, R. D. (2006). *Longitudinal Data Analysis*. New York: Wiley. 3

Heidelberger, P. & Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research* **31**, 1109–1144. 32, 57

Horton, N. J. & Laird, N. M. (1999). Maximum likelihood analysis of generalized linear models with missing covariates. *Statistical Methods in Medical Research* **8**, 37–50. doi: 10.1177/096228029900800104. 45

Horton, N. J. & Laird, N. M. (2001). Maximum likelihood analysis of logistic regression models with incomplete covariate data and auxiliary information. *Biometrics* **57**, 34–42. doi: 10.1111/j.0006-341X.2001.00034.x. 45

Huang, L., Chen, M.-H. & Ibrahim, J. G. (2005). Bayesian analysis for generalized linear models with nonignorably missing covariates. *Biometrics* **61**, 767–780. doi: 10.1111/j.1541-0420.2005.00338.x. 45

Hyndman, R. J. & Fan, Y. (1996). Sample quantiles in statistical packages. *The American Statistician* **50**, 361–365. 71

Ibrahim, J. G. (1990). Incomplete data in generalized linear models. *Journal of the American Statistical Association* **85**, 765–769. 45

Ibrahim, J. G., Chen, M.-H., Lipsitz, S. R. & Herring, A. H. (2005). Missing-data methods for generalized linear models: a comparative review. *Journal of the American Statistical Association* **100**, 332–346. doi: 10.1198/016214504000001844. 46

Ibrahim, J. G., Lipsitz, S. R. & 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. doi: 10.1111/1467-9868.00170. 45, 64

Ishwaran, H. & James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* **96**, 161–173. doi: 10.1198/016214501750332758. 52

Ishwaran, H. & James, L. F. (2002). Approximate Dirichlet process computing finite normal mixtures: smoothing and prior information. *Journal of Computational and Graphical Statistics* **11**, 508–532. doi: 10.1198/106186002411. 50, 51, 53, 64

Jansen, I., Molenberghs, G., Aerts, M., Thijs, H. & Van Steen, K. (2003). A local influence approach applied to binary data from a psychiatric study. *Biometrics* **59**, 409–418. doi: 10.1111/1541-0420.00048. 19

Jiang, T. J. & Dickey, J. M. (2008). Bayesian methods for categorical data under informative censoring. *Bayesian Analysis* **3**, 541–554. doi: 10.1214/08-BA321. 11, 22, 26

Kenward, M. G., Goetghebeur, E. & Molenberghs, G. (2001). Sensitivity analysis for incomplete categorical data. *Statistical Modelling* **1**, 31–48. doi: 10.1177/1471082X0100100104. 18, 28, 29, 30, 42

Kenward, M. G. & Molenberghs, G. (1998). Likelihood based frequentist inference when data are missing at random. *Statistical Science* **13**, 236–247. doi: 10.1214/ss/1028905886. 10

Kenward, M. G., Molenberghs, G. & Thijs, H. (2003). Pattern-mixture models with proper time dependence. *Biometrika* **90**, 53–71. doi: 10.1093/biomet/90.1.53. 11, 14, 15, 44

Korn, E. L. & Graubard, B. I. (1999). *Analysis of Health Surveys*. New York: Wiley. 3

Lehtonen, R. & Pahkinen, E. (2004). *Practical Methods for Design and Analysis of Complex Surveys*, 2ª ed. New York: Wiley. 3

Leonard, T. (1996). On exchangeable sampling distributions for uncontrolled data. *Statistics & Probability Letters* **26**, 1–6. doi: 10.1016/0167-7152(94)00245-2. 51

Levy, P. S. & Lemeshow, S. (1999). *Sampling of Populations: Methods and Applications*, 3ª ed. New York: Wiley. 3

Lindsey, J. K. (1996). *Parametric Statistical Inference*. Oxford: Oxford University Press. 10

Lindsey, J. K. (1999). *Models for Repeated Measurements*, 2ª ed. Oxford: Oxford University Press. 3, 17

Lipsitz, S. R. & Ibrahim, J. G. (1996). A conditional model for incomplete covariates in parametric regression models. *Biometrika* **83**, 916–922. doi: 10.1093/biomet/83.4.916. 45, 64

Lipsitz, S. R., Ibrahim, J. G., Chen, M.-H. & Peterson, H. (1999). Non-ignorable missing covariates in generalized linear models. *Statistics in Medicine* **18**, 2435–2448. doi: 10.1002/(SICI)1097-0258(19990915/30)18:17/18<2435::AID-SIM267>3.0.CO;2-B. 45

Lipsitz, S. R., Parzen, M. & Ewell, M. (1998). Inference using conditional logistic regression with missing covariates. *Biometrics* **54**, 295–303. 45

Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* **88**, 125–134. 7, 12

Little, R. J. A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika* **81**, 471–483. doi: 10.1093/biomet/81.3.471. 12, 67

Little, R. J. A. & Rubin, D. B. (2002). *Statistical Analysis with Missing Data*, 2ª ed. New York: Wiley. 3, 4, 7, 23, 67

Little, R. J. A. & Wang, Y.-X. (1996). Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics* **52**, 98–111. 67

Lunn, D. J., Spiegelhalter, D., Thomas, A. & Best, N. (2009). The BUGS project: evolution, critique and future directions (with discussion). *Statistics in Medicine* **28**, 3049–3082. doi: 10.1002/sim.3680. 4, 51

Lunn, D. J., Thomas, A., Best, N. & Spiegelhalter, D. (2000). WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* **10**, 325–337. doi: 10.1023/A:1008929526011. 51

Mendenhall, W. & Lehman, J., E. H. (1960). An approximation to the negative moments of the positive binomial useful in life testing. *Technometrics* **2**, 227–242. 71

Miranda, A. & Rabe-Hesketh, S. (2010). Missing ordinal covariates with informative selection. Relatório técnico, NCRM Working Paper. Department of Quantitative Social Science, Institute of Education, University of London. 45

Müller, P. & Quintana, F. A. (2004). Nonparametric Bayesian data analysis. *Statistical Science* **19**, 95–110. doi: 10.1214/088342304000000017. 48

Molenberghs, G., Beunckens, C., Sotto, C. & Kenward, M. G. (2008). Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **70**, 371–388. doi: 10.1111/j.1467-9868.2007.00640.x. 18, 39

Molenberghs, G., Goetghebeur, E., Lipsitz, S. R. & Kenward, M. G. (1999). Nonrandom missingness in categorical data: strengths and limitations. *The American Statistician* **53**, 110–118. 11, 40

Molenberghs, G. & Kenward, M. G. (2007). *Missing Data in Clinical Studies.* New York: Wiley. 3, 4, 19, 67

Molenberghs, G., Kenward, M. G. & 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. doi: 10.1111/1467-9876.00217. 18, 21, 28, 29, 31

Molenberghs, G., Kenward, M. G., Verbeke, G. & Birhanu, T. (2011). Pseudo-likelihood estimation for incomplete data. *Statistica Sinica* **21**, 187–206. 85

Molenberghs, G., Michiels, B., Kenward, M. G. & Diggle, P. J. (1998). Monotone missing data and pattern mixture models. *Statistica Neerlandica* **52**, 153–161. doi: 10.1111/1467-9574.00075. 14

Molenberghs, G. & Verbeke, G. (2005). *Models for Discrete Longitudinal Data.* New York: Springer. 3

Mouchart, M. (2007). Ignorable missingness. *Manuscrito preliminar não publicado.* 17

Neal, R. M. (2003). Slice sampling (with discussion). *The Annals of Statistics* **31**, 705–767. doi: 10.1214/aos/1056562461. 30

Neath, A. A. & Samaniego, F. J. (1997). On the efficacy of Bayesian inference for nonidentifiable models. *The American Statistician* **51**, 225–232. 21, 22

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. 21

Paddock, S. M. (2002). Bayesian nonparametric multiple imputation of partially observed data with ignorable nonresponse. *Biometrika* **89**, 529–538. doi: 10.1093/biomet/89.3.529. 64

Paik, M. C. (2004). Nonignorable missingness in matched case-control data analyses. *Biometrics* **60**, 306–314. doi: 10.1111/j.0006-341X.2004.00174.x. 45

Paulino, C. D. (1988). *Análise de dados categorizados incompletos: fundamentos, métodos e aplicações.* Tese de doutorado. IME–USP. 8, 11

Paulino, C. D. (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. 84

Paulino, C. D. & Pereira, C. A. B. (1992). Bayesian analysis of categorical data informatively censored. *Communications in Statistics - Theory and Methods* **21**, 2689–2705. doi: 10.1080/03610929208830937. 8, 11, 22, 23, 25

Paulino, C. D. & Pereira, C. A. B. (1994). On identifiability of parametric statistical models. *Statistical Methods & Applications: Journal of the Italian Statistical Society* **3**, 125–151. doi: 10.1007/BF02589044. 21

Paulino, C. D. & Pereira, C. A. B. (1995). Bayesian methods for categorical data under informative general censoring. *Biometrika* **82**, 439–446. doi: 10.1093/biomet/82.2.439. 8, 11, 21, 22, 23, 24, 25

Paulino, C. D. & Singer, J. M. (2006). *Análise de Dados Categorizados.* São Paulo: Edgard Blücher. 3

Plummer, M., Best, N., Cowles, K. & Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R News* **6(1)**, 7–11. 4

Poleto, F. Z. (2006). *Análise de dados categorizados com omissão.* Dissertação de mestrado. IME–USP. 1, 5, 19, 45

Poleto, F. Z., Singer, J. M. & Paulino, C. D. (2011a). Comparing diagnostic tests with missing data. *Journal of Applied Statistics* **38**, 1207–1222. doi: 10.1080/02664763.2010.491860. 10

Poleto, F. Z., Singer, J. M. & Paulino, C. D. (2011b). Missing data mechanisms and their implications on the analysis of categorical data. *Statistics and Computing* **21**, 31–43. doi: 10.1007/s11222-009-9143-x. 10, 11

Poleto, F. Z., Singer, J. M. & Paulino, C. D. (2011c). A product-multinomial framework for categorical data analysis with missing responses. *Submetido para publicação.* Relatório Técnico RT–MAE–2007–07. Instituto de Matemática e Estatística, Universidade de São Paulo. 84

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing.* Vienna. URL `http://www.R-project.org`. 4

Raftery, A. E. & Lewis, S. M. (1992). How many iterations in the Gibbs sampler? Em J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, eds., *Bayesian Statistics 4.* Oxford: Oxford University Press, 763–773. 32, 57

Robins, J. M., Rotnitzky, A. & 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. 85

Rotnitzky, A. (2008). Inverse probability weighted methods. Em G. Fitzmaurice, M. Davidian, G. Verbeke & G. Molenberghs, eds., *Longitudinal Data Analysis.* Boca Raton: Chapman & Hall, 453–476. 85

Rotnitzky, A., Robins, J. M. & Scharfstein, D. O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* **93**, 1321–1339. 67

Rotnitzky, A., Scharfstein, D., Su, T.-L. & Robins, J. M. (2001). Methods for conducting sensitivity analysis of trials with potentially nonignorable competing causes of censoring. *Biometrics* **57**, 103–113. doi: 10.1111/j.0006-341X.2001.00103.x. 67

Rubin, D. B. (1974). Characterizing the estimation of parameters in incomplete-data problems. *Journal of the American Statistical Association* **69**, 467–474. 6

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592. doi: 10.1093/biomet/63.3.581. 8, 12, 17, 39

Rubin, D. B. (1977). Formalizing subjective notions about the effect of nonrespondents in sample surveys. *Journal of the American Statistical Association* **72**, 538–543. 67, 81

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley. 3, 4, 8

Satten, G. A. & Carroll, R. J. (2000). Conditional and unconditional categorical regression models with missing covariates. *Biometrics* **56**, 384–388. doi: 10.1111/j.0006-341X.2000.00384.x. 45

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data.* Boca Raton: Chapman & Hall / CRC. 3, 4

Scharfstein, D. O., Daniels, M. J. & Robins, J. M. (2003). Incorporating prior beliefs about selection bias into the analysis of randomized trials with missing outcomes. *Biostatistics* **4**, 495–512. doi: 10.1093/biostatistics/4.4.495. 12, 26, 27, 31, 37, 46, 58, 81

Scharfstein, D. O., Rotnitzky, A. & Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion). *Journal of the American Statistical Association* **94**, 1096–1146. 21, 67

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization.* New York: Wiley. 48

Sen, P. K., Singer, J. M. & Pedroso de Lima, A. C. (2009). *From Finite Sample to Asymptotic Methods in Statistics.* Cambridge: Cambridge University Press. 71

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639–650. 49

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Boca Raton: Chapman & Hall. 51, 71

Smith, B. J. (2007). boa: an R package for MCMC output convergence assessment and posterior inference. *Journal of Statistical Software* **21**. 4

Smith, P. W. F., Skinner, C. J. & 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. doi: 10.1111/1467-9876.00172. 11

Soares, P. (2004). *Análise bayesiana de dados deficientemente categorizados.* Tese de doutorado. Instituto Superior Técnico, Universidade Técnica de Lisboa. 5, 8, 11, 31

Soares, P. & Paulino, C. D. (2001). Incomplete categorical data analysis: a Bayesian perspective. *Journal of Statistical Computation and Simulation* **69**, 157–170. doi: 10.1080/00949650108812088. 8, 11, 22, 25, 33, 39

Soares, P. & Paulino, C. D. (2007). Log-linear models for coarse categorical data. Em M. I. Gomes, D. Pestana & P. Silva, eds., *Proceedings of the 56th Session of the International Statistical Institute.* 26, 30

Srivastava, M. S. (2002). *Methods of Multivariate Statistics.* New York: Wiley. 3

Särndal, C.-E. & Lundström, S. (2005). *Estimation in Surveys with Nonresponse.* New York: Wiley. 3

Särndal, C.-E., Swensson, B. & Wretman, J. H. (1991). *Model Assisted Survey Sampling.* New York: Springer. 3

Stephan, F. F. (1945). The expected value and variance of the reciprocal and other negative powers of a positive Bernoullian variate. *The Annals of Mathematical Statistics* **16**, 50–61. doi: 10.1214/aoms/1177731170. 71

Stubbendick, A. L. & Ibrahim, J. G. (2003). Maximum likelihood methods for nonignorable missing responses and covariates in random effects models. *Biometrics* **59**, 1140–1150. doi: 10.1111/j.0006-341X.2003.00131.x. 45

Stubbendick, A. L. & Ibrahim, J. G. (2006). Likelihood-based inference with nonignorable missing responses and covariates in models for discrete longitudinal data. *Statistica Sinica* **16**, 1143–1167. 45

Tan, M. T., Tian, G.-L. & Ng, K. W. (2010). *Bayesian Missing Data Problems: EM, Data Augmentation and Noniterative Computation.* Boca Raton: Chapman & Hall. 3

Tanner, M. A. & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association* **82**, 528–550. 26

Thijs, H., Molenberghs, G., Michiels, B., Verbeke, G. & Curran, D. (2002). Strategies to fit pattern-mixture models. *Biostatistics* **3**, 245–265. doi: 10.1093/biostatistics/3.2.245. 13

Tian, G.-L., Ng, K. W. & Geng, Z. (2003). Bayesian computation for contingency tables with incomplete cell-counts. *Statistica Sinica* **13**, 189–206. 22, 25, 26

Tsiatis, A. A. (2006). *Semiparametric Theory and Missing Data*. New York: Springer. 3

Twisk, J. W. R. (2002). *Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide*. Cambridge: Cambridge University Press. 3

Vach, W. (1994). *Logistic Regression with Missing Values in the Covariates*. New York: Springer. 2, 3

Vach, W. (1997). Some issues in estimating the effect of prognostic factors from incomplete covariate data. *Statistics in Medicine* **16**, 57–72. doi: 10.1002/(SICI)1097-0258(19970115)16:1<57::AID-SIM471>3.0.CO;2-S. 45

Vach, W. & 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. doi: 10.1002/sim.4780141205. 45

Vach, W. & Schumacher, M. (1993). Logistic regression with incompletely observed categorical covariates: a comparison of three approaches. *Biometrika* **80**, 353–362. doi: 10.1093/biomet/80.2.353. 45

Van Der Laan, M. J. & Robins, J. M. (2002). *Unified Methods for Censored and Longitudinal Data and Causality*. New York: Springer. 3

Van Steen, K., Molenberghs, G., Verbeke, G. & Thijs, H. (2001). A local influence approach to sensitivity analysis of incomplete longitudinal ordinal data. *Statistical Modelling* **1**, 125–142. doi: 10.1177/1471082X0100100203. 19

Vansteelandt, R. A., S. & Robins, J. (2007). Estimation of regression models for the mean of repeated outcomes under nonignorable nonmonotone nonresponse. *Biometrika* **94**, 841–860. doi: 10.1093/biomet/asm070. 17

Vansteelandt, S., Goetghebeur, E., Kenward, M. G. & Molenberghs, G. (2006). Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica* **16**, 953–979. 18, 21, 23, 28, 30, 31, 77, 78, 79, 84

Verbeke, G. & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer. 3, 13

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. 11, 31

Walker, S. G., Damien, P., Laud, P. W. & Smith, A. F. M. (1999). Bayesian nonparametric inference for random distributions and related functions (with discussion). *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **61**, 485–527. doi: 10.1111/1467-9868.00190. 49

Wang, C.-P. & Ghosh, M. (2000). Bayesian analysis of bivariate competing risks models. *Sankhya: The Indian Journal of Statistics* **62**, 388–401. 26

Weiss, R. E. (2005). *Modeling Longitudinal Data*. New York: Springer. 3

West, M. (1992). Modelling with mixtures (with discussion). Em J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, eds., *Bayesian Statistics 4*. Oxford: Oxford University Press, 503–524. 50, 51, 52, 53, 64

Wicki, J., Perneger, T. V., Junod, A. F., Bounameaux, H. & Perrier, A. (2001). Assessing clinical probability of pulmonary embolism in the emergency ward. *Archives of Internal Medicine* **161**, 92–97. 46, 61

Wilks, S. S. (1932). Moments and distributions of estimates of population parameters from fragmentary samples. *Annals of Mathematical Statistics* **3**, 163–195. doi: 10.1214/aoms/1177732885. 3

Wu, M. C. & Bailey, K. R. (1988). Analysing changes in the presence of informative right censoring caused by death and withdrawal. *Statistics in Medicine* **7**, 337–346. doi: 10.1002/sim.4780070134. 7

Wu, M. C. & Bailey, K. R. (1989). Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics* **45**, 939–955. 7

Wu, M. C. & Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**, 175–188. 7

Zhang, Z. (2010). Profile likelihood and incomplete data. *International Statistical Review* **78**, 102–116. doi: 10.1111/j.1751-5823.2010.00107.x. 84

Zhang, Z. & Rockette, H. E. (2005). On maximum likelihood estimation in parametric regression with missing covariates. *Journal of Statistical Planning and Inference* **134**, 206–223. doi: 10.1016/j.jspi.2004.04.006. 45

Zhang, Z. & Rockette, H. E. (2006). Semiparametric maximum likelihood for missing covariates in parametric regression. *Annals of the Institute of Statistical Mathematics* **58**, 687–706. doi: 10.1007/s10463-006-0047-7. 45

Zhang, Z. & Rockette, H. E. (2007). An EM algorithm for regression analysis with incomplete covariate information. *Journal of Statistical Computation and Simulation* **77**, 163–173. doi: 10.1080/10629360600565202. 45

Zhao, Y. (2009). Regression analysis with covariates missing at random: a piece-wise nonparametric model for missing covariates. *Communications in Statistics - Theory and Methods* **38**, 3736–3744. doi: 10.1080/03610920802618392. 45

Índice de autores

Aerts, M. 15–17, 19

Agresti, A. 32

Allison, P. D. 3, 68

Antoniak, C. E. 49, 50

Bailey, K. R. 7

Baker, S. G. 2, 11, 42

Basu, D. 10

Bedrick, E. J. 30

Berry, D. 3

Best, N. 4, 51

Beunckens, C. 18, 39

Birhanu, T. 85

Blackwell, D. 51

Blettner, M. 45

Bounameaux, H. 46, 61

Carroll, R. J. 7, 45

Chambers, R. L. 3

Chen, H. Y. 45, 46

Chen, M.-H. 45, 46, 64

Christensen, R. 30

Clarke, P. S. 11

Congdon, P. 3, 48, 51

Cook, R. D. 19

Copas, J. B. 21

Cowles, K. 4

Creemers, A 19

Curran, D. 13

Damien, P. 49

Daniels, M. J. 3, 12, 21, 26, 27, 31, 37, 46, 58, 67, 81, 82

Davidian, M. 3, 67

De Leeuw, J. 3

DerSimonian, R. 2, 42

Dickey, J. M. 11, 22, 25, 26

Diggle, P. J. 3, 11, 14, 44

Dillman, D. A. 3

Efron, B. 72

Eltinge, J. L. 3

Escobar, M. D. 50, 51, 53, 64

Ewell, M. 45

Fan, Y. 71

Fay, R. E. 40

Ferguson, T. S. 48

Fitzmaurice, G. 3, 67

Fitzmaurice, G. M. 3

Fleiss, J. L. 3

Forster, J. J. 22, 33

G., Verbeke 19

Gelman, A. 3, 57, 59

Geng, Z. 22, 25, 26

Geweke, J. 32, 57

Ghosh, M. 26

Gibbons, L. E. 45

Gibbons, R. D. 3
Glonek, G. F. V. 11
Glynn, R. J. 7, 23, 67
Goetghebeur, E. 11, 18, 21, 23, 28–31, 40, 42, 77–79, 84
Gong, G. 72
Grab, E. L. 71
Graubard, B. I. 3
Griffin, J. 52, 65
Groves, R. M. 3
Gustafson, P. 21

Hagenaars, J. A. 3
Halloran, M. E. 3
Harrell, F. E. 3
Heagerty, P. 3
Heckman, J. J. 6
Hedeker, D. 3
Heidelberger, P. 32, 57
Hens, N. 15–17, 19
Herring, A. H. 46
Hill, J. 3
Hogan, J. W. 3, 21, 67, 81, 82
Holmes, C. 52, 65
Horton, N. J. 45
Hosmer, D. W. 45
Huang, L. 45
Hyndman, R. J. 71

Ibrahim, J. G. 45, 46, 64
Ishwaran, H. 50–53, 64

James, L. F. 50–53, 64
Jansen, I. 19
Jiang, T. J. 11, 22, 26
Johnson, W. 30
Junod, A. F. 46, 61

Kenward, M. G. 3, 4, 10, 11, 14–19, 21, 23, 28–31, 39, 40, 42, 44, 67, 77–79, 84, 85
Korn, E. L. 3

Laird, N. M. 3, 7, 11, 23, 45, 67
Laud, P. W. 49
Lehman, Jr., E. H. 71
Lehtonen, R. 3
Lemeshow, S. 3
Leonard, T. 51
Levin, B. 3
Levy, P. S. 3
Lewis, S. M. 32, 57
Li, H. G. 21
Liang, K.-Y. 3
Lindsey, J. K. 3, 10, 17
Lipsitz, S. R. 11, 40, 45, 46, 64
Little, R. J. A. 3, 4, 7, 12, 23, 45, 67
Lundström, S. 3
Lunn, D. J. 4, 51

MacQueen, J. B. 51
McCutcheon, A. L. 3
Meijer, E. 3
Mendenhall, W. 71
Meng, X.-L. 3
Michiels, B. 13, 14
Min, Y. 32
Miranda, A. 45
Müller, P. 48
Molenberghs, G. 3, 4, 10, 11, 13–19, 21, 23, 28–31, 39, 40, 42, 44, 67, 77–79, 84, 85
Mouchart, M. 17

Neal, R. M. 30
Neath, A. A. 21, 22
Ng, K. W. 3, 22, 25, 26

Nordheim, E. V. 21

Paddock, S. M. 64
Pahkinen, E. 3
Paik, M. C. 3, 45
Parzen, M. 45
Paulino, C. D. 3, 8, 10, 11, 21–26, 30, 33, 39, 84
Pedroso de Lima, A. C. 71
Pereira, C. A. B. 8, 11, 21–25
Perneger, T. V. 46, 61
Perrier, A. 46, 61
Peterson, H. 45
Plummer, M. 4
Poleto, F. Z. 1, 5, 10, 11, 19, 45, 84

Quintana, F. A. 48

R Development Core Team 4
Rabe-Hesketh, S. 45
Raftery, A. E. 32, 57
Robins, J. 17
Robins, J. M. 3, 12, 21, 26, 27, 31, 37, 46, 58, 67, 81, 85
Rockette, H. E. 45
Rosenberger, W. F. 2, 42
Rotnitzky, A. 21, 67, 85
Rubin, D. B. 3, 4, 6–8, 12, 17, 23, 39, 57, 59, 67, 81

Samaniego, F. J. 21, 22
Satten, G. A. 45
Savage, I. R. 71
Schafer, J. L. 3, 4
Scharfstein, D. 67
Scharfstein, D. O. 12, 21, 26, 27, 31, 37, 46, 58, 67, 81
Schumacher, M. 45
Scott, D. W. 48
Sen, P. K. 71
Sethuraman, J. 49
Silverman, B. W. 51, 71
Singer, J. M. 3, 10, 11, 71, 84
Skinner, C. J. 3, 11
Smith, A. F. M. 49
Smith, B. J. 4
Smith, P. W. F. 11, 22, 33
Soares, P. 5, 8, 11, 22, 25, 26, 30, 31, 33, 39
Sotto, C. 18, 39
Spiegelhalter, D. 4, 51
Srivastava, M. S. 3
Särndal, C.-E. 3
Stephan, F. F. 71
Stubbendick, A. L. 45
Su, T.-L. 67
Swensson, B. 3

Tan, M. T. 3
Tanner, M. A. 26
Thijs, H. 11, 13–15, 19, 44
Thomas, A. 4, 51
Tian, G.-L. 3, 22, 25, 26
Tsiatis, A. A. 3
Twisk, J. W. R. 3

Vach, W. 2, 3, 45
Van Der Laan, M. J. 3
Van Steen, K. 19
Vansteelandt, Rotnitzky A., S. 17
Vansteelandt, S. 18, 21, 23, 28, 30, 31, 77–79, 84
Verbeke, G. 3, 13, 15–17, 19, 67, 85
Vines, K. 4

Walker, S. 11, 31
Walker, S. G. 49
Wang, C.-P. 26

Wang, Y.-X. 67
Ware, J. H. 3
Weiss, R. E. 3
Welch, P. D. 32, 57
West, M. 50–53, 64
Wicki, J. 46, 61
Wilks, S. S. 3
Wong, W. H. 26
Wretman, J. H. 3
Wu, M. C. 7

Zeger, S. L. 3
Zhang, Z. 45, 84
Zhao, L. P. 85
Zhao, Y. 45