

**Modelos baseados no planejamento
para análise de populações finitas**

Luz Mery González García

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

Área de Concentração: Estatística
Orientador: Prof. Dr. Julio da Motta Singer

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CNPq e da
Universidad Nacional de Colombia

São Paulo, 27 de maio de 2008

Modelos baseados no planejamento para análise de populações finitas

Este exemplar corresponde à redação
final da tese devidamente corrigida
e defendida por Luz Mery González García
e aprovada pela Comissão Julgadora.

Banca Examinadora:

- Prof. Dr. Julio da Motta Singer (presidente) - IME-USP.
- Prof. Dr. Heleno Bolfarine - IME-USP.
- Prof. Dr. Damiano Nóbrega da Silva - UFRN.
- Prof. Dr. Cristiano Ferraz - UFPE.
- Profa. Dra. Viviana Beatriz Lencina - UNT.

A Deus pelo amor e energia positiva.

A meus pais, Mary e Libardo, pela compreensão e o apoio.

A meu namorado, Nelson, pelo apoio, o carinho e a paciência.

Agradecimentos

Em primeiro lugar, quero agradecer aos meus pais, Mary e Libardo, e ao meu namorado, Nelson, quem na distância caminharam comigo para conseguir este logro, pela força e pelo constante apoio emocional.

Sou muito grata ao Prof. Julio Singer por acreditar em mim e me apoiar, não apenas para iniciar meus estudos de doutorado no IME, mas também durante todo o curso e toda a elaboração deste trabalho, pela paciência, pela dedicação, pelas discussões e pela confiança.

Ao Prof. Ed Stanek, a Viviana, a Silvina, a Wenjun e a Bo pelas idéias e discussões sobre os avanços do trabalho.

Ao CNPq e à Universidad Nacional de Colombia (UNAL, Colômbia) que me apoiaram financeiramente durante estes quatro anos.

A todos meus colegas do IME que compartilharam comigo este caminho, em especial a Gladys, Raydonal e Patricia pela amizade e disposição nos momentos alegres e difíceis.

Aos professores e a os funcionários do IME pelo trato afetuoso e pela ajuda que dedicam aos estudantes.

Aos meus colegas no Departamento de Estatística (UNAL, Colômbia), em especial a Luis Alberto López, pela sua confiança, e a Emilse por seu apoio.

À Senhora Lighia e ao Claus por seu apreço e por me acolher na sua casa.

A Domingo, Sandra, Cristian, Liliam e Mauricio por me receber e acolher na chegada ao Brasil.

E, finalmente, a todos os que me acompanharam, me apoiaram e compartilharam comigo estes quatro anos.

Resumo

Estudamos o problema de obtenção de estimadores/preditores ótimos para combinações lineares de respostas coletadas de uma população finita por meio de amostragem aleatória simples. Nesse contexto, estendemos o modelo misto para populações finitas proposto por Stanek, Singer & Lencina (2004, *Journal of Statistical Planning and Inference*) para casos em que se incluem erros de medida (endógenos e exógenos) e informação auxiliar. Admitindo que as variâncias são conhecidas, mostramos que os estimadores/preditores propostos têm erro quadrático médio menor dentro da classe dos estimadores lineares não viciados. Por meio de estudos de simulação, comparamos o desempenho desses estimadores/preditores empíricos, i.e., obtidos com a substituição das componentes de variância por estimativas, com aquele de competidores tradicionais. Também, estendemos esses modelos para análise de estudos com estrutura do tipo pré-teste/pós-teste. Também por intermédio de simulação, comparamos o desempenho dos estimadores empíricos com o desempenho do estimador obtido por meio de técnicas clássicas de análise de medidas repetidas e com o desempenho do estimador obtido via análise de covariância por meio de mínimos quadrados, concluindo que os estimadores/preditores empíricos apresentaram um menor erro quadrático médio e menor vício. Em geral, sugerimos o emprego dos estimadores/preditores empíricos propostos para dados com distribuição assimétrica ou amostras pequenas.

Palavras-chave: coeficiente de regressão, erros de medida, estimador ótimo, estudos pré-teste/pós-teste, informação auxiliar, medida de associação, modelo de permutação aleatória.

Abstract

We consider optimal estimation of finite population parameters with data obtained via simple random samples. In this context, we extend a finite population mixed model proposed by Stanek, Singer & Lencina (2004, *Journal of Statistical Planning and Inference*) by including measurement errors (endogenous or exogenous) and auxiliary information. Assuming that variance components are known, we show that the proposed estimators/predictors have the smallest mean squared error in the class of unbiased estimators. Using simulation studies, we compare the performance of the empirical estimators/predictors obtained by replacing variance components with estimates with the performance of a traditional estimator. We also extend the finite population mixed model to data obtained via pretest-posttest designs. Through simulation studies, we compare the performance of the empirical estimator of the difference in gain between groups with the performance of the usual repeated measures estimator and with the performance of the usual analysis of covariance estimator obtained via ordinary least squares. The empirical estimator has smaller mean squared error and bias than the alternative estimators under consideration. In general, we recommend the use of the proposed estimators/ predictors for either asymmetric response distributions or small samples.

Keywords: auxiliary information, BLUP, measure of association, measurement error, optimal estimator, pretest-posttest designs, random permutation model, regression coefficient, response error.

Sumário

Lista de Figuras	xi
Lista de Tabelas	xv
1 Introdução	1
1.1 Motivação do Problema 1	1
1.2 Motivação do Problema 2	4
1.3 Objetivos	11
2 Estimação e previsão em populações finitas com informação auxiliar e erros de medida	13
2.1 Modelos que contemplam erros de medida	15
2.2 Combinações lineares e preditores ótimos	17
2.3 Particularização dos resultados para estimação de uma medida de associação	27
2.4 Extensão para casos com múltiplas variáveis auxiliares	34
2.5 Discussão	35
3 Estimação e previsão para experimentos pré-teste/pós-teste com populações finitas	41
3.1 Modelo misto para estudos pré-teste/pós-teste em populações finitas com dois estratos	42
3.1.1 Combinações lineares de interesse e preditores ótimos	46
3.1.2 Exemplo	53

3.2	Modelo misto para estudos pré-teste/pós-teste em populações finitas com dois tipos de intervenção	55
3.2.1	Combinações lineares de interesse e preditores ótimos	61
3.2.2	Exemplo	69
3.2.3	Estudo por simulação dos estimadores propostos	70
3.3	Discussão	74
4	Considerações finais	77
4.1	Conclusões	77
4.2	Pesquisa Futura	78
	Apêndices	79
A	Cálculo do vetor de médias e da matriz de covariâncias dos modelos propostos nos Capítulos 2 e 3	79
B	Obtenção dos preditores para as combinações lineares descritas no Capítulo 2	87
C	Resumo das simulações do Capítulo 2.	91
D	Obtenção dos preditores para as combinações lineares descritas no Capítulo 3	97
E	Resumo das simulações relativas ao Capítulo 3, Seção 3.2.	105
F	Dados dos exemplos	111
G	Interpretação da medida de associação entre duas variáveis	115
	Referências Bibliográficas	117

Lista de Figuras

1.1	Diagrama de dispersão das notas em Matemática na FUVEST vs habilidades em Cálculo (que assumimos conhecidas e iguais às notas de Cálculo I), no ano 2004 dos estudantes do IME ($N = 104$), e reta da regressão obtida via mínimos quadrados ($y = 5.3253 + 0.0574x$).	3
1.2	Perfis individuais da pressão arterial média de cães, antes (pré-teste) e depois (pós-teste) de uma aplicação de $MgSO_4$	5
1.3	Perfis individuais da resistência homogênea do sistema respiratório de ratos.	6
1.4	Perfis médios e médias amostrais (círculos) da pressão arterial média de cães, antes (pré-teste) e depois (pós-teste) de uma aplicação de $MgSO_4$	7
1.5	Valores das médias amostrais (círculos) da resistência homogênea do sistema respiratório de ratos.	8
1.6	Diagrama de dispersão da pressão arterial média de cães antes (pré-teste) vs depois (pós-teste) de uma aplicação de $MgSO_4$, e retas esperadas segundo o modelo (1.5).	9
1.7	Diagrama de dispersão da resistência homogênea do sistema respiratório, antes (pré-teste) vs depois (pós-teste) de submeter os ratos a ventilação com ar sintético, ou com Hélio e Oxigênio, e retas esperadas segundo o modelo (1.6).	10
2.1	Gráfico de dispersão e reta de regressão calculada via mínimos quadrados ordinários ($y = 1.4265 + 0.4395x$) para o exemplo ilustrativo apresentado na Tabela 2.1.	29
2.2	Diagrama de dispersão de x vs y^\bullet para uma população com $\rho_{yx}^\bullet = 0.5$ e tamanho 2000, e diagramas de dispersão de ρ_{yx}^\bullet vs ρ_{yx} , ρ_{yx}^\bullet vs $Var_S(\hat{B}^*)$ e ρ_{yx}^\bullet vs $Var_S(\bar{B}^*)$ para 200 populações geradas. Dados simétricos.	38

- 2.3 Diagrama de dispersão de x vs y^\bullet para uma população com $\rho_{yx}^\bullet = 0.5$ e tamanho 2000, e diagramas de dispersão de ρ_{yx}^\bullet vs ρ_{yx} , ρ_{yx}^\bullet vs $Var_S(\hat{B}^*)$ e ρ_{yx}^\bullet vs $Var_S(\bar{B}^*)$ para 200 populações geradas. Dados assimétricos in x 39
- 2.4 Diagrama de dispersão de x vs y^\bullet para uma população com $\rho_{yx}^\bullet = 0.5$ e tamanho 2000, e diagramas de dispersão de ρ_{yx}^\bullet vs ρ_{yx} , ρ_{yx}^\bullet vs $Var_S(\hat{B}^*)$ e ρ_{yx}^\bullet vs $Var_S(\bar{B}^*)$ para 200 populações geradas. Dados assimétricos em x e y^\bullet 40
- C.1 EQM ($\times 10^{-2}$), média do vício relativo e contribuição relativa do vício ao EQM de \hat{B}^* , \hat{B}_e^* e \hat{B}_{MQO} para diferentes valores de ρ_{yx}^\bullet , n (o primeiro número entre parênteses) e σ_e^\bullet (o segundo número entre parênteses). Variâncias iguais para os erros de medida endógenos. Dados com distribuição simétrica. 91
- C.2 EQM ($\times 10^{-2}$), média do vício relativo e contribuição relativa do vício ao EQM de \hat{B}^* , \hat{B}_e^* e \hat{B}_{MQO} para diferentes valores de ρ_{yx}^\bullet , n (o primeiro número entre parênteses) e σ_e^\bullet (o segundo número entre parênteses). Variâncias iguais para os erros de medida endógenos. Dados com distribuição assimétrica in x 92
- C.3 EQM ($\times 10^2$) e média do vício relativo de \hat{B}^* , \hat{B}_e^* e \hat{B}_{MQO} para diferentes valores de ρ_{yx}^\bullet e n . Variâncias diferentes para os erros de medida endógenos. Dados com distribuição simétrica. 93
- C.4 EQM e média do vício relativo de \hat{B}^* , \hat{B}_e^* e \hat{B}_{MQO} para diferentes valores de ρ_{yx}^\bullet e n . Variâncias diferentes para os erros de medida endógenos. Dados com distribuição assimétrica in x 93
- C.5 EQM, média do vício relativo e contribuição relativa do vício ao EQM de \tilde{B} , \tilde{B}_e e \tilde{B}_{MQO} para diferentes valores de ρ_{yx}^\bullet e n . Variâncias diferentes para os erros de medida exógenos. Dados com distribuição simétrica. 94
- C.6 EQM, média do vício relativo e contribuição relativa do vício ao EQM de \tilde{B} , \tilde{B}_e e \tilde{B}_{MQO} para diferentes valores de ρ_{yx}^\bullet e n . Variâncias diferentes para os erros de medida exógenos. Dados com distribuição assimétrica in x 95
- F.1 Boxplot da pressão arterial média (mmHg) de cães, antes (pré-teste) e depois (pós-teste) de uma aplicação de $MgSO_4$, e efeito da aplicação. 112
- F.2 Boxplot da resistência homogênea do sistema respiratório (RHSR) de ratos antes (pré-teste), depois (pós-teste) e efeito da ventilação com Ar Sintético, ou com Hélio e Oxigênio. 113

Lista de Tabelas

2.1	População finita considerando variâncias iguais para os erros de medida endógenos.	28
2.2	Estimativas de B sob o modelo misto para populações finitas, com σ_{yx} e σ_x^2 conhecidas ou desconhecidas, estimativas de mínimos quadrados ordinários e média do EQM para $N = 5$ e $n = 3$ com variâncias iguais para os erros de medida endógenos.	30
2.3	População finita considerando variâncias diferentes para os erros de medida endógenos.	30
2.4	Estimativas de B sob o modelo misto para populações finitas, com σ_{yx} e σ_x^2 conhecidas ou desconhecidas, estimativas de mínimos quadrados ordinários e média do EQM para $N = 5$ e $n = 3$ com variâncias diferentes para os erros de medida endógenos.	31
2.5	Estimativas de B sob o modelo misto para populações finitas, com σ_{yx} e σ_x^2 conhecidas ou desconhecidas, estimativas de mínimos quadrados ordinários e média do EQM para $N = 5$ e $n = 3$ com variâncias diferentes para os erros de medida exógenos.	31
2.6	EQM ($\times 10^{-3}$) e média da contribuição relativa do vício ao EQM (dentro do parênteses) das estimativas.	32
3.1	Medidas resumo para os dados do Exemplo 1.1.	53
3.2	Estimativas da diferença dos efeitos, na pressão arterial média em cães, da aplicação de M_gSO_4 entre o grupo submetido à droga Indometacina e o grupo submetido à droga Nifedipina, variância estimada e intervalos de confiança, considerando o caso com erros de medida endógenos, com erros de medida exógenos e sem erros de medida.	54
3.3	Medidas resumo ($\times 10^{-3}$) para os dados de Exemplo 1.2.	69

3.4	Estimativas ($\times 10^{-3}$) da diferença dos efeitos, na resistência respiratória de ratos, da ventilação com uma mistura de Oxigênio e Hélio e da ventilação com ar sintético, variância estimada ($\times 10^{-3}$) e intervalos de confiança ($\times 10^{-3}$), considerando erros de medida endógenos, erros de medida exógenos e sem erros de medida.	71
E.1	EQM ($\times 10^{-2}$) e contribuição relativa do vício ao EQM (entre parênteses) das estimativas, considerando variâncias iguais para os erros de medida endógenos. Dados com distribuição simétrica.	106
E.2	EQM e contribuição relativa do vício ao EQM (entre parênteses) das estimativas, considerando variâncias iguais para os erros de medida endógenos. Dados com distribuição assimétrica.	107
E.3	EQM ($\times 10^{-2}$) considerando variâncias diferentes para os erros de medida endógenos. Dados com distribuição simétrica.	108
E.4	EQM e contribuição relativa do vício ao EQM (entre parênteses) considerando variâncias diferentes para os erros de medida endógenos. Dados com distribuição assimétrica.	108
E.5	EQM ($\times 10^{-2}$) das estimativas considerando variâncias diferentes para os erros de medida exógenos. Dados com distribuição simétrica.	109
E.6	EQM e contribuição relativa do vício ao EQM (entre parênteses) das estimativas considerando variâncias diferentes para os erros de medida exógenos. Dados com distribuição assimétrica.	109
F.1	Pressão arterial média (mmHg) de cães antes e depois de uma aplicação de $MgSO_4$	111
F.2	Resistência homogênea do sistema respiratório (RHSR) de ratos.	112
F.3	Notas em Matemática na FUVEST e habilidades em Cálculo (que assumimos conhecidas e iguais às notas de Cálculo I), de 104 estudantes do IME no ano 2004.	114

Capítulo 1

Introdução

Neste trabalho, apresentamos extensões dos modelos mistos para populações finitas propostos por Stanek & Singer (2004) para resolver dois problemas. O primeiro é a estimação/predição de combinações lineares de respostas obtidas de uma população finita por meio de amostragem aleatória simples na presença de informação auxiliar e erros de medida. O segundo é a análise de dados provenientes de experimentos do tipo pré-teste/pós-teste na presença de erros de medida.

Salientamos que neste trabalho vamos usar, como notação de quantidades numéricas, o ponto para separar unidades e décimais, em concordância com notação usada em publicações internacionais.

1.1 Motivação do Problema 1

O Instituto de Matemática e Estatística (IME), da Universidade de São Paulo, recebe cada ano em torno de 120 estudantes novos nos Bacharelados de Ciências da Computação, Estatística, Matemática e Matemática Aplicada. O ingresso deles depende da nota obtida na prova FUVEST, que avalia seu desempenho em Física, Matemática, Português, etc. Essa nota é baseada em provas planejadas para quantificar a habilidade (um valor latente) dos candidatos em tais matérias. Utilizando a nota FUVEST como informação auxiliar, nosso interesse é:

- a) estimar a habilidade média em Cálculo (que não é medida com a nota FUVEST e sim com a nota de Cálculo I) no primeiro semestre de 2004 (a nota de Cálculo I é observada via uma ou mais provas e, possivelmente, mede a habilidade em Cálculo com erro de medida),
- b) prever a habilidade em Cálculo para um estudante selecionado,

- c) avaliar a associação entre as notas obtidas na FUVEST e as habilidades em Cálculo, e
- d) estimar e/ou prever qualquer combinação linear das habilidades em Cálculo usando como informação auxiliar tanto as notas em Matemática (FUVEST) quanto outras informações disponíveis.

Autores como Cassel, Särndal & Wretman (1976), Deville & Särndal (1992), Särndal, Swensson & Wretman (1992), Chen & Qin (1993), Wu & Sitter (2001) e Li, Stanek & Singer (2008) têm estudado o uso de informação auxiliar para melhorar a precisão de estimadores de parâmetros como a média populacional [ítem a)], por exemplo. Por outro lado, a incorporação de erros de medida tem sido abordada por Bolfarine (1991), Bolfarine & Zacks (1992), Särndal, Swensson & Wretman (1992), Stanek & Singer (2004) e Stanek, Lencina, Singer, González, Li & San Martino (2008), entre outros, em que Särndal, Swensson & Wretman (1992) tratam os erros de medida por meio de um modelo estatístico para medições (efetuadas sobre elementos de uma amostra de uma população finita); Bolfarine & Zacks (1992), por meio de modelos de super-população e Stanek, Lencina, Singer, González, Li & San Martino (2008), por meio do modelo misto para populações finitas com erro de medida. Os últimos consideram a predição de um valor latente para um indivíduo selecionado [ítem b)].

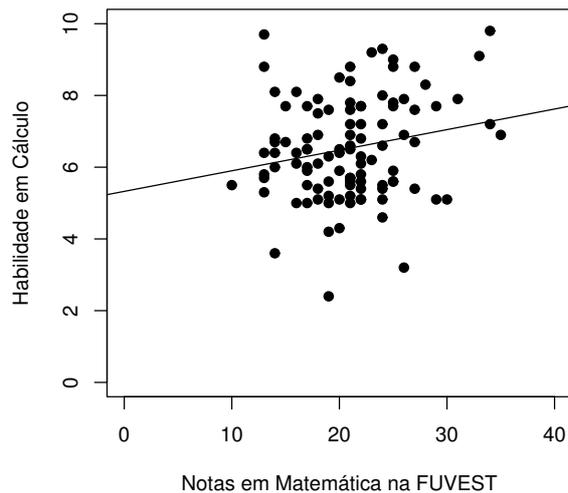
Para contemplar o ítem c), definimos uma “medida de associação” entre duas variáveis de uma população finita. Associando a cada estudante, rotulado por s , $s = 1, \dots, N$, na população finita de tamanho N , os valores fixos y_s e x_s que representam o valor latente correspondente à habilidade em Cálculo e a nota obtida em Matemática na prova FUVEST, respectivamente, definimos

$$B = \frac{\sum_{s=1}^N (y_s - \mu_y)(x_s - \mu_x)}{\sum_{s=1}^N (x_s - \mu_x)^2}, \quad (1.1)$$

em que $\mu_y = N^{-1} \sum_{s=1}^N y_s$ e $\mu_x = N^{-1} \sum_{s=1}^N x_s$, como uma medida de associação entre as duas variáveis, y e x . Assumindo, para efeitos ilustrativos, que além de conhecer a nota obtida em Matemática na prova FUVEST, também conhecemos os valores latentes das habilidades em Cálculo (lembrando que na realidade, estes valores são observados com erro), que aqui são tomados como sendo os mesmos das notas de Cálculo I, na Figura 1.1 mostramos o diagrama de dispersão para as notas dos 104 ingressantes em 2004 e a correspondente reta da regressão $y = A + Bx$ obtida via mínimos quadrados. Os dados estão apresentados na Tabela F.3 (pág. 114, Apêndice F).

Muitos autores têm considerado a estimação do parâmetro B . Em particular, Kish & Frankel (1974) e Särndal, Swensson & Wretman (1992) estudaram o problema sob o enfoque baseado no planejamento, enquanto Fuller (1975), Holt, Smith & Winter (1980), Pfeiffermann & Smith (1985),

Figura 1.1: Diagrama de dispersão das notas em Matemática na FUVEST vs habilidades em Cálculo (que assumimos conhecidas e iguais às notas de Cálculo I), no ano 2004 dos estudantes do IME ($N = 104$), e reta da regressão obtida via mínimos quadrados ($y = 5.3253 + 0.0574x$).



Bolfarine & Zacks (1992) e Bolfarine, Zacks, Elian & Rodrigues (1994) estudaram-no a partir do enfoque de super-população. Sob o primeiro enfoque, a distância entre os pontos (x_s, y_s) e $(x_s, A + Bx_s)$, $s = 1, \dots, N$, em que $A = \mu_y - \mu_x B$, corresponde à falta de ajuste “lack of fit”, enquanto sob o segundo enfoque, ela corresponde a um erro de medida.

No primeiro caso, dada uma amostra aleatória simples de tamanho n da população finita (fixa) de tamanho N , o estimador de B é aquele obtido via mínimos quadrados ordinários (MQO), \hat{B}_{MQO} , e coincide com o estimador Horvitz-Thompson (ou estimador- π)^a. Esta equivalência ocorre para qualquer plano amostral que tenha probabilidades iguais de inclusão amostral. No segundo, o estimador também é dado por \hat{B}_{MQO} . Comentários sobre a interpretação do parâmetro B in (1.1), no contexto de populações finitas, estão apresentados no Apêndice G.

^a Se π_s é a probabilidade de inclusão da unidade s numa amostra, e \mathcal{S} é o conjunto de rótulos obtidos na amostra (ou unidades na amostra), o estimador- π de B é definido como

$$\hat{B}_\pi = \left\{ \sum_{s \in \mathcal{S}} \pi_s^{-1} (x_s - \tilde{x}_\pi)^2 \right\}^{-1} \left\{ \sum_{s \in \mathcal{S}} \pi_s^{-1} (x_s - \tilde{x}_\pi) (y_s - \tilde{y}_\pi) \right\},$$

em que $\tilde{x}_\pi = N^{-1} \sum_{s \in \mathcal{S}} \pi_s^{-1} x_s$ e $\tilde{y}_\pi = N^{-1} \sum_{s \in \mathcal{S}} \pi_s^{-1} y_s$. \tilde{x}_π e \tilde{y}_π são os estimadores π de μ_x e μ_y , respectivamente. Sob amostragem aleatória simples sem reposição, $\pi_s = n/N$ para todo s e neste caso $\hat{B}_\pi = \hat{B}_{MQO} = \left\{ \sum_{s \in \mathcal{S}} (x_s - \bar{x})^2 \right\}^{-1} \left\{ \sum_{s \in \mathcal{S}} (x_s - \bar{x}) (y_s - \bar{y}) \right\}$, com $\bar{x} = N^{-1} \sum_{s \in \mathcal{S}} x_s$ e $\bar{y} = N^{-1} \sum_{s \in \mathcal{S}} y_s$. Para detalhes adicionais sobre o estimador- π , ver Särndal, Swensson & Wretman (1992), por exemplo.

1.2 Motivação do Problema 2

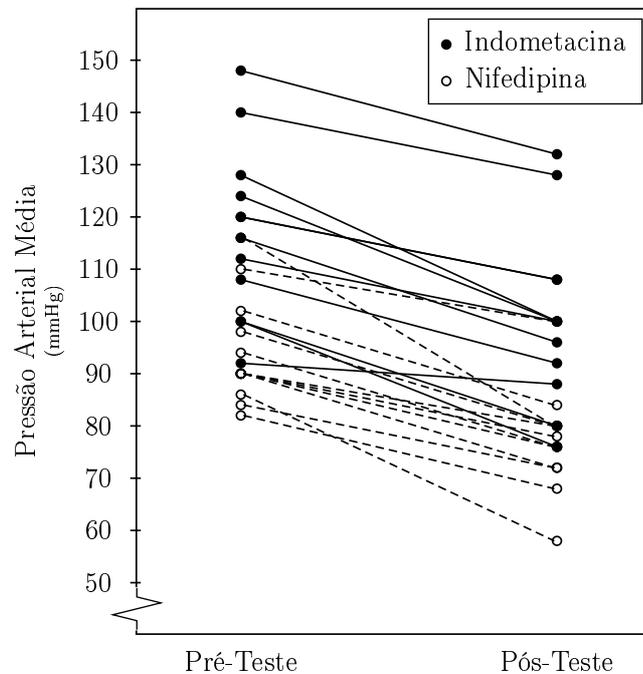
Estudos do tipo pré-teste/pós-teste são frequentemente usados em áreas como Biologia, Medicina, Odontologia e Psicologia, entre outras. O objetivo, em geral, é avaliar o efeito de algum tipo de intervenção (um tratamento, por exemplo) na distribuição de alguma resposta. O procedimento consiste na observação da variável resposta antes e depois da intervenção. Consideramos estudos do tipo pré-teste/pós-teste sob dois cenários que descrevemos por meio de exemplos.

Exemplo 1.1 (Problema da pressão arterial em cães): Num estudo realizado no Instituto de Ciências Biomédicas da Universidade de São Paulo para avaliar o efeito de $MgSO_4$ na pressão arterial média-PAM (mmHg) de cães, a resposta foi medida antes e depois da aplicação em dois grupos de animais, o primeiro ($n = 12$) previamente tratado com indometacina e o outro ($n = 12$) previamente tratado com nifedipina. Para detalhes, ver Singer, Seoanes & Ogando (1988). Os dados estão apresentados na Tabela F.1 (pág. 111, Apêndice F) e um gráfico de perfis individuais, na Figura 1.2. Neste exemplo, temos *uma intervenção* (aplicação de $MgSO_4$) e *uma população com dois estratos*, que correspondem aos dois grupos experimentais. O interesse é:

- a) avaliar a diferença da PAM média dos dois grupos no pré-teste,
- b) avaliar o efeito (PAM média pós-teste - PAM média pré-teste) da aplicação de $MgSO_4$ dentro de cada grupo,
- c) avaliar a diferença dos efeitos da aplicação de $MgSO_4$ entre os dois grupos.

Exemplo 1.2 (Problema da resistência respiratória de ratos): Consideramos dados obtidos num estudo realizado na Faculdade de Medicina da Universidade de São Paulo para avaliar a resistência respiratória de ratos com respeito a gases de diferentes densidades. Apesar do estudo levar em conta várias variáveis respostas, vamos considerar apenas a variável resistência homogênea do sistema respiratório (RHSR). Primeiramente, a RHSR de cada um dos 29 ratos foi medida logo após serem submetidos à ventilação com ar sintético. Posteriormente, 15 dos ratos escolhidos aleatoriamente foram submetidos novamente à ventilação com ar sintético (C) e os outros 14 ratos foram submetidos à ventilação com uma mistura de Hélio e Oxigênio (T). Usamos a letra C para o ar sintético porque representa o grupo ‘controle’ e a letra T para a mistura de Hélio e Oxigênio porque representa o ‘tratamento’ de interesse. Ao término da intervenção, uma segunda medição da RHSR foi realizada. Os dados estão apresentados na Tabela F.2 (pág. 112, Apêndice F) e um gráfico de perfis individuais, na Figura 1.3. Neste exemplo, temos *uma população não estratificada* medida num ponto de referência (pré-teste), que corresponde à medição realizada logo após a ventilação com ar sintético. A cada

Figura 1.2: Perfis individuais da pressão arterial média de cães, antes (pré-teste) e depois (pós-teste) de uma aplicação de $MgSO_4$.



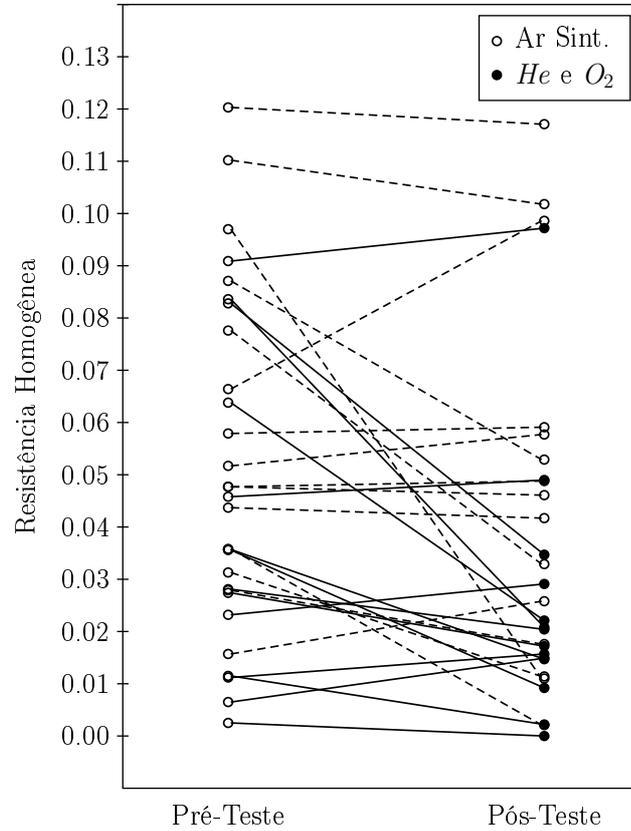
indivíduo dessa população foi aplicado um de *dois tipos de intervenção* de forma aleatória. Essas intervenções correspondem à ventilação com C ou à ventilação com T . Neste cenário, o interesse é:

- avaliar o efeito (RHSR média pós-teste - RHSR média pré-teste) da ventilação com C ,
- avaliar o efeito (RHSR média pós-teste - RHSR média pré-teste) da ventilação com T ,
- avaliar se os efeitos da ventilação com T e com C são equivalentes.

Diferentes autores têm estudado diversos estimadores para os parâmetros relacionados com os itens a), b) e c) nos Exemplos 1.1 e 1.2 sob o enfoque baseado em modelos; entre eles estão os trabalhos de Brogan & Kutner (1980), Laird (1983), Stanek (1988), Knoke (1991), Singer & Andrade (1997), Bonate (2000), Yang & Tsiatis (2001), Leon, Tsiatis & Davidian (2003) e Aoki, Achcar, Bolfarine & Singer (2003).

Modelos lineares para *medidas repetidas* e modelos de *análise de covariância* são os mais utilizados na análise desse tipo de dados. A seguir discutimos os Exemplos 1.1 e 1.2 sob estes modelos.

Figura 1.3: Perfis individuais da resistência homogênea do sistema respiratório de ratos.



• Modelo para medidas repetidas

No Exemplo 1.1, supondo que são observados n_h cães do estrato h , $h = 1, 2$, o modelo linear básico pode ser escrito como

$$Y_{hit}^* = \mu_{ht} + \epsilon_{hit}, \quad h = 1, 2, \quad i = 1, \dots, n_h \quad \text{e} \quad t = 1, 2, \quad (1.2)$$

em que ϵ_{hit} é um componente aleatório com média zero, e μ_{h1} e μ_{h2} correspondem ao valor esperado da pressão arterial média (PAM) pré-teste e pós-teste, respectivamente. Portanto, Y_{hi1}^* e Y_{hi2}^* são variáveis aleatórias correspondentes à PAM pré-teste e pós-teste do cão i do estrato h , respectivamente. As componentes μ_{ht} definidas em (1.2) podem ser re-expressas como

$$\mu_{ht} = \mu + \alpha_h + \gamma_t + (\alpha\gamma)_{ht}, \quad h = 1, 2 \quad \text{e} \quad t = 1, 2,$$

com $\alpha_1 + \alpha_2 = 0$, $\gamma_1 + \gamma_2 = 0$, $(\alpha\gamma)_{1t} + (\alpha\gamma)_{2t} = 0$ e $(\alpha\gamma)_{h1} + (\alpha\gamma)_{h2} = 0$. Nesse caso μ simboliza a média geral, α_h representa o desvio entre a média da PAM dos cães do estrato h e a média geral, μ , γ_1 (respectivamente, γ_2) representa o desvio entre a média da PAM dos cães

no pré-teste (respectivamente, pós-teste) e a média geral μ , e $(\alpha\gamma)_{ht} = \mu_{ht} - \mu - \alpha_h - \gamma_t$.

Para modelar a estrutura de covariância intra-unidades amostrais, podemos considerar

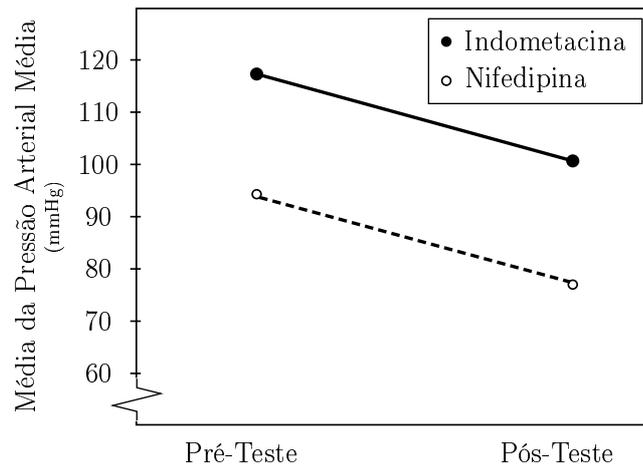
$$\epsilon_{hit} = \pi_{i(h)} + (\pi\gamma)_{it(h)} + e_{m(hit)}, \quad i = 1, \dots, n_h, \quad t = 1, 2, \quad h = 1, 2 \quad \text{e} \quad m = 1,$$

com $(\pi\gamma)_{i1(h)} + (\pi\gamma)_{i2(h)} = 0^b$, $i = 1, \dots, n_h$, $h = 1, 2$, em que $\pi_{i(h)}$, $(\pi\gamma)_{it(h)}$ e $e_{m(hit)}$ são variáveis aleatórias tais que

$$\begin{aligned} E[\pi_{i(h)}] &= E[(\pi\gamma)_{it(h)}] = E[e_{m(hit)}] = 0, \\ \text{Var}[\pi_{i(1)}] &= \text{Var}[\pi_{i(2)}] = \sigma_\pi^2, \quad \text{Var}[e_{m(hit)}] = \sigma_{ee}^2, \\ \text{Var}[(\pi\gamma)_{it(1)}] &= \text{Var}[(\pi\gamma)_{it(2)}] = \sigma_{\pi\gamma}^2/2, \\ \text{Cov}[(\pi\gamma)_{i1(h)}, (\pi\gamma)_{i2(h)}] &= -\sigma_{\pi\gamma}^2/2, \end{aligned}$$

e $\pi_{i(h)}$, $(\pi\gamma)_{it(h)}$ e $e_{m(hit)}$ são não correlacionadas duas a duas, $i = 1, \dots, n_h$, $t = 1, 2$ e $h = 1, 2$. Um gráfico de perfis individuais é apresentado na Figura 1.2 e um gráfico dos perfis médios, na Figura 1.4.

Figura 1.4: Perfis médios e médias amostrais (círculos) da pressão arterial média de cães, antes (pré-teste) e depois (pós-teste) de uma aplicação de $MgSO_4$.



No Exemplo 1.2, considerando que na primeira medição nenhum dos n ratos observados tinham sido submetidos a qualquer intervenção, o modelo linear para as medidas realizadas no

^b Ao impor a restrição $(\pi\gamma)_{i1(h)} + (\pi\gamma)_{i2(h)} = 0$, $i = 1, \dots, n_h$, $h = 1, 2$, estamos considerando o modelo com parâmetros restritos “the constrained parameters (CP) model” que facilita a interpretação dos parâmetros inseridos no modelo. Porém, o modelo com parâmetros irrestritos “the unconstrained parameters (UP) model”, também pode ser utilizado, ver Voss (1999) e Lencina, Singer & Stanek (2005).

pré-teste [ver Bonate (2000, p. 120)] é

$$Y_{i1}^* = \mu_1 + e_{i1}, \quad i = 1, \dots, n, \quad (1.3)$$

em que e_{i1} é um componente aleatório, tal que $E(e_{i1}) = 0$, $Var(e_{i1}) = \sigma_1^2$ e $Cov(e_{i1}, e_{i'1}) = 0$ para $i, i' = 1, \dots, n$, $i \neq i'$, e μ_1 é RHSR média (populacional) no pré-teste. O modelo linear para a RHSR pós-teste é

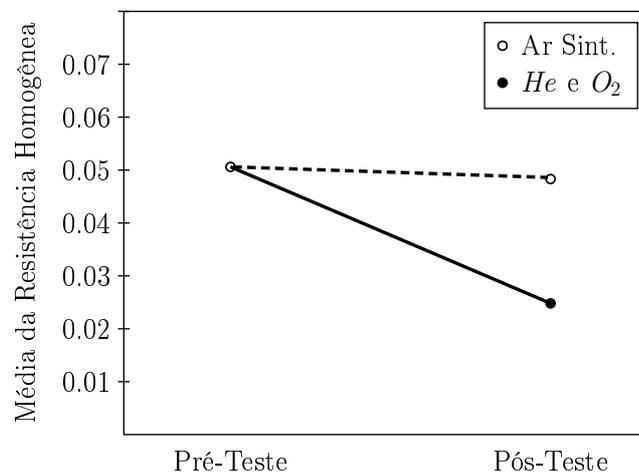
$$Y_{i2}^{h*} = \mu_h + e_{i2}^h, \quad i = 1, \dots, n, \quad (1.4)$$

se o rato i é submetido à ventilação com h , $h = C$ ou $h = T$, em que e_{i2}^C e e_{i2}^T são componentes aleatórios tais que $E(e_{i2}^C) = E(e_{i2}^T) = 0$, $Var(e_{i2}^C) = (\sigma_2^C)^2$, $Var(e_{i2}^T) = (\sigma_2^T)^2$ e e_{i2}^C e e_{i2}^T são não correlacionadas duas a duas para $i = 1, \dots, n$, e μ_C e μ_T correspondem à RHSR média pós-teste associada à ventilação com C ou com T , respectivamente. O modelo (1.4) pode ser reparametrizado usando

$$\mu_C = \mu_1 + \alpha_C + \gamma \quad \text{e} \quad \mu_T = \mu_1 + \alpha_T + \gamma,$$

com $\alpha_C + \alpha_T = 0$, em que γ representa o desvio entre a média geral pós-teste ($\mu_1 + \gamma$) e a média geral pré-teste (μ_1), e α_C (respectivamente, α_T) representa o desvio entre a média da RHSR pós-teste quando os ratos são submetidos a ventilação com C (respectivamente, T) e a média geral pós-teste ($\mu_1 + \gamma$). Um gráfico dos perfis individuais é apresentado na Figura 1.3 e um gráfico das médias amostrais $\hat{\mu}_1$, $\hat{\mu}_C$ e $\hat{\mu}_T$, na Figura 1.5.

Figura 1.5: Valores das médias amostrais (círculos) da resistência homogênea do sistema respiratório de ratos.



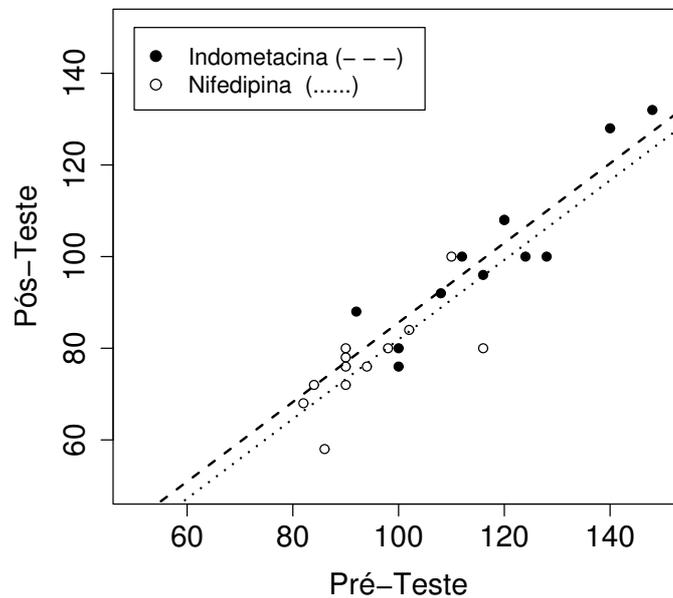
- **Modelo de análise de covariância**

Para o Exemplo 1.1, admitindo a PAM pré-teste Y_{hi1}^* , $i = 1, \dots, n_h$, $h = 1, 2$, fixada, o modelo é

$$Y_{hi2}^* = \mu_2 + \tau_h + \beta (Y_{hi1}^* - \bar{Y}_1) + e_{hi}, \quad h = 1, 2, \quad i = 1, \dots, n_h, \quad (1.5)$$

com $\bar{Y}_1 = (n_1 + n_2)^{-1} \sum_{h=1}^2 \sum_{i=1}^{n_h} Y_{hi1}^*$, $\tau_1 + \tau_2 = 0$, em que e_{hi} é um erro (de medida) aleatório tal que $E(e_{hi}) = 0$, $Var(e_{hi}) = \sigma_{Ae}^2$ e as variáveis e_{hi} são não correlacionadas duas a duas para $i = 1, \dots, n_h$, $h = 1, 2$, e μ_2 , τ_h e β correspondem, respectivamente, à média geral da PAM pós-teste, ao efeito do estrato h na PAM média e ao coeficiente de regressão linear comum (de Y_{hi2}^* em função de Y_{hi1}^*). O diagrama de dispersão para a PAM dos cães e as retas $E(\widehat{Y_{hi2}^*}) = \hat{\mu}_2 + \hat{\tau}_h + \hat{\beta} (Y_{hi1}^* - \bar{Y}_1)$, $h = 1, 2$, obtidas via mínimos quadrados estão apresentados na Figura 1.6.

Figura 1.6: Diagrama de dispersão da pressão arterial média de cães antes (pré-teste) vs depois (pós-teste) de uma aplicação de $MgSO_4$, e retas esperadas segundo o modelo (1.5).



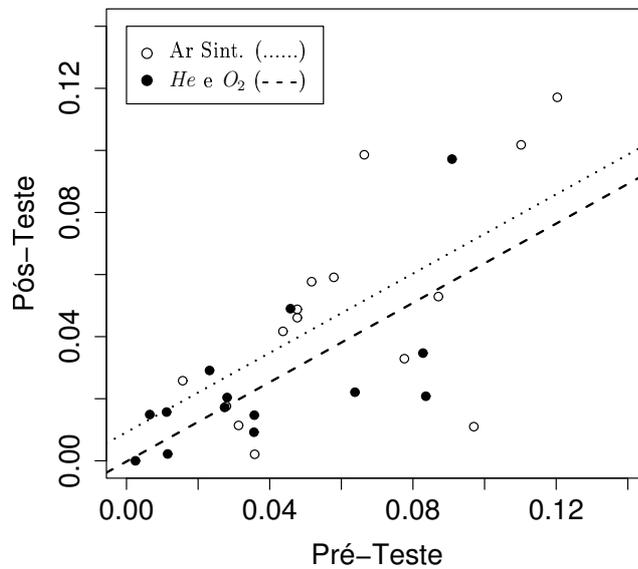
Para o Exemplo 1.2, admitindo a RHSR pré-teste Y_{i1}^* , $i = 1, \dots, n$, fixada, o modelo é

$$Y_{i2}^{h*} = \mu_2 + \tau_h + \beta (Y_{i1}^* - \bar{Y}_1) + e_{ih}, \quad i = 1, \dots, n, \quad (1.6)$$

se o rato i é submetido à ventilação com h , $h = C$ ou $h = T$, com $\bar{Y}_1 = n^{-1} \sum_{i=1}^n Y_{i1}^*$,

$\tau_C + \tau_T = 0$, em que e_{i_C} e e_{i_T} são erros (de medida) aleatórios tal que $E(e_{i_C}) = E(e_{i_T}) = 0$, $Var(e_{i_C}) = Var(e_{i_T}) = \sigma_{Ae}^2$ e e_{i_C} e e_{i_T} são não correlacionadas duas a duas, e μ_2 , β e τ_C (respectivamente, τ_T) correspondem, respectivamente, à média geral da RHSR pós-teste, ao coeficiente de regressão linear comum (de Y_{i2}^{h*} em função de Y_{i1}^*) e ao efeito da ventilação com C (respectivamente, T) na RHSR média. O diagrama de dispersão para a RHSR dos ratos e as retas $E(Y_{i2}^{h*}) = \hat{\mu}_2 + \hat{\tau}_h + \hat{\beta}(Y_{i1}^* - \bar{Y}_1)$, $h = C, T$, obtidas via mínimos quadrados, estão apresentados na Figura 1.7.

Figura 1.7: Diagrama de dispersão da resistência homogênea do sistema respiratório, antes (pré-teste) vs depois (pós-teste) de submeter os ratos a ventilação com ar sintético, ou com Hélio e Oxigênio, e retas esperadas segundo o modelo (1.6).



Os modelos (1.2), (1.3) e (1.4) assumem erros (de medida) aleatórios tanto no pré-teste quanto no pós-teste, enquanto os modelos (1.5) e (1.6) assumem que as respostas no pré-teste não têm erro de medida. Um modelo que considera erros de medida no pré-teste e usa as respostas no pré-teste como variável independente (error-in-variables model), sob o cenário do Exemplo 1.2, é apresentado por Knoke (1991).

Em geral, o interesse recai, entre outros, na estimação dos parâmetros do modelo segundo o qual, supõe-se, as respostas foram geradas. Neste contexto, n observações estão associadas a n realizações independentes das variáveis aleatórias inseridas no respectivo modelo, por conseguinte, os estimadores são calculados assumindo amostragem de uma população infinita (conceitual). Para os modelos (1.2) a (1.6), os parâmetros são estimados via mínimos quadrados ordinários.

1.3 Objetivos

Nossos objetivos são:

1. No contexto de amostragem aleatória simples, obter e avaliar o desempenho de estimadores e/ou preditores para combinações lineares das respostas sob *o modelo misto para populações finitas* sugerido por Stanek, Singer & Lencina (2004), incorporando *informação auxiliar* de forma semelhante a Li, Stanek & Singer (2008) e levando em conta possíveis *erros de medida* da resposta.
2. No contexto de estudos do tipo pré-teste/pós-teste, obter e avaliar o desempenho de estimadores e/ou preditores para combinações lineares das respostas pré-teste e pós-teste sob cada um dos cenários apresentados na Seção 1.2, considerando *o modelo misto para populações finitas* sugerido por Stanek, Singer & Lencina (2004) e levando em conta possíveis *erros de medida* tanto no pré-teste quanto no pós-teste.

Os erros de medida mencionados nos itens 1. e 2. podem ser *endógenos* (EMEn), i.e., originado por fatores inerentes à unidade, como em Stanek & Singer (2004), ou *exógenos* (EMEx), i.e., originado por fatores estranhos à unidade (instrumento de medida, por exemplo). Ver Stanek, Lencina, Singer, González, Li & San Martino (2008) para detalhes.

Queremos obter estimadores/preditores sob suposições baseadas apenas na amostragem e nos primeiros e segundos momentos da distribuição dos erros de medida, que são suposições mais fracas que as usualmente adotadas nestes tipos de estudos.

Considerando inicialmente estudos do tipo transversal, no Capítulo 2 descrevemos o modelo misto para populações finitas inserindo informação auxiliar e erros de medida. Na Seção 2.1 descrevemos os modelos com a incorporação de erros de medida endógenos e erros de medida exógenos. Na Seção 2.2, obtemos estimadores de combinações lineares da resposta. Como caso especial, na Seção 2.3 particularizamos os resultados para obter o estimador de B em (1.1), além disso, apresentamos exemplos numéricos para comparar o desempenho do estimador de B frente ao estimador obtido via mínimos quadrados ordinários, \hat{B}_{MQO} . Também, apresentamos os resultados de um estudo do desempenho dos estimadores propostos com base em amostras independentes, de tamanho igual, selecionadas dos dados referidos na Seção 1.1, e, de um estudo de simulação em que são considerados diferentes valores para o coeficiente de correlação entre a variável auxiliar e a variável resposta. Na Seção 2.4, estendemos os resultados para o caso de mais de uma variável auxiliar. Na Seção 2.5, fazemos uma breve discussão dos resultados obtidos no Capítulo 2.

No Capítulo 3, descrevemos estudos do tipo pré-teste/pós-teste considerando os dois cenários apresentados nos Exemplos 1.1 e 1.2 sob um modelo misto para populações finitas com possíveis erros de medida. Na Seção 3.1, considerando erros de medida endógenos e erros de medida exógenos, apresentamos esse modelo para uma população finita com dois estratos cujas unidades são submetidas a uma intervenção, obtemos estimadores/preditores de combinações lineares das respostas pré-teste e pós-teste e apresentamos um exemplo numérico baseado nos dados do Exemplo 1.1. Na Seção 3.2, considerando erros de medida endógenos e erros de medida exógenos, apresentamos esse modelo para uma população finita não estratificada com dois tipos de intervenção, obtemos estimadores/preditores de combinações lineares das respostas pré-teste e pós-teste e apresentamos um exemplo numérico baseados nos dados do Exemplo 1.2, além dos resultados de um pequeno estudo de simulação. Na Seção 3.3, fazemos uma breve discussão dos resultados obtidos no Capítulo 3.

Finalmente, no Capítulo 4 discutimos as conclusões e apresentamos propostas para estudos futuros.

Capítulo 2

Estimação e previsão em populações finitas com informação auxiliar e erros de medida

Utilizando notação semelhante a Stanek, Singer & Lencina (2004), definimos uma população finita como uma coleção de N unidades distinguíveis e rotuladas por $s = 1, \dots, N$. Associados a cada unidade s , vamos supor dois valores y_s^\bullet e x_s , em que y_s^\bullet é a resposta e x_s é uma variável auxiliar. Além disso, assumimos que ao selecionar a unidade s , observamos tanto o valor do parâmetro y_s^\bullet (denominado valor latente) quanto o valor da variável auxiliar x_s , sendo que y_s^\bullet é observado possivelmente com erro enquanto x_s é observada sem erro. Como em muitas situações é conveniente trabalhar com os valores latentes ponderados, definimos $y_s = w_s y_s^\bullet$, em que w_s , $s = 1, \dots, N$, representam os pesos.

O processo da amostragem aleatória simples é formalizado por meio de um *modelo para populações finitas*, segundo o qual qualquer uma das $N!$ possíveis permutações de unidades na população pode ser selecionada com a mesma probabilidade $[(N!)^{-1}]$. Para isto, um novo rótulo, $i = 1, \dots, N$, é associado às unidades segundo sua posição em cada permutação. De forma que podemos representá-las pela matriz aleatória $(\mathbf{Y} \ \mathbf{X})$ em que

$$\mathbf{Y} = (Y_1 \ \dots \ Y_i \ \dots \ Y_N)' \quad \text{e} \quad \mathbf{X} = (X_1 \ \dots \ X_i \ \dots \ X_N)' \quad (2.1)$$

Para associar os valores

$$\mathbf{y} = (y_1 \ \dots \ y_s \ \dots \ y_N)' \quad \text{e} \quad \mathbf{x} = (x_1 \ \dots \ x_s \ \dots \ x_N)' \quad (2.2)$$

a \mathbf{Y} e \mathbf{X} , consideramos

$$(\mathbf{Y} \ \mathbf{X}) = \mathbf{U} (\mathbf{y} \ \mathbf{x}) \quad (2.3)$$

em que $\mathbf{U} = (\mathbf{U}_1 \dots \mathbf{U}_N)'$ com $\mathbf{U}_i = (U_{i1} \dots U_{iN})'$ e U_{is} é uma variável indicadora com valor 1 se a unidade s é selecionada na posição i dentro da permutação e 0 em caso contrário. Definindo $\delta_{sy} = y_s - \mu_y$ e $\delta_{sx} = x_s - \mu_x$ e, portanto, $y_s = \delta_{sy} + \mu_y$ e $x_s = \delta_{sx} + \mu_x$, para a posição i dentro de uma permutação, temos

$$\begin{aligned} Y_i &= \sum_{s=1}^N U_{is} y_s = \mu_y + d_{iy}, \\ X_i &= \sum_{s=1}^N U_{is} x_s = \mu_x + d_{ix}, \end{aligned} \quad (2.4)$$

em que $\mu_y = N^{-1} \sum_{s=1}^N y_s$, $\mu_x = N^{-1} \sum_{s=1}^N x_s$, $d_{iy} = \sum_{s=1}^N U_{is} \delta_{sy}$ e $d_{ix} = \sum_{s=1}^N U_{is} \delta_{sx}$.

Usando as propriedades da variável indicadora U_{is} [ver (A.1)], segue que

$$\mathbb{E}_S \{ \text{vec} [(\mathbf{Y} \ \mathbf{X})] \} = (\mathbf{I}_2 \otimes \mathbf{1}_N) \boldsymbol{\mu} \quad (2.5)$$

e

$$\text{Var}_S \{ \text{vec} [(\mathbf{Y} \ \mathbf{X})] \} = \boldsymbol{\Sigma} \otimes \mathbf{P}_N$$

em que o sub-índice S indica valor esperado com respeito à distribuição das permutações das unidades, \mathbf{I}_a é a matriz identidade de ordem a , $\mathbf{1}_a$ é um vetor de dimensão $a \times 1$ com todos os elementos iguais a 1, $\mathbf{P}_{a,b} = \mathbf{I}_a - b^{-1} \mathbf{J}_a$, $\mathbf{P}_a = \mathbf{P}_{a,a}$, $\mathbf{J}_a = \mathbf{1}_a \mathbf{1}_a'$, vec e \otimes denotam o operador vec e o produto Kronecker, respectivamente, [ver Harville (1997), por exemplo],

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix} \quad \text{e} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_y^2 & \sigma_{yx} \\ \sigma_{yx} & \sigma_x^2 \end{pmatrix}, \quad (2.6)$$

com μ_y e μ_x definidos em (2.4), $\sigma_y^2 = (N-1)^{-1} \sum_{s=1}^N (y_s - \mu_y)^2$, $\sigma_x^2 = (N-1)^{-1} \sum_{s=1}^N (x_s - \mu_x)^2$ e $\sigma_{yx} = (N-1)^{-1} \sum_{s=1}^N (y_s - \mu_y)(x_s - \mu_x)$.

2.1 Modelos que contemplam erros de medida

Embora erros de medida possam modificar estatísticas como a média e a variância dos estimadores de um parâmetro de interesse, neste trabalho, temos em conta erros de medida que modificam apenas a variância dos estimadores. Consideramos dois tipos de erros de medida: o primeiro originado por fatores inerentes à unidade (EMEn) e o segundo originado por fatores estranhos à unidade (EMEx). O EMEn leva em conta a variabilidade endógena, i.e., a variabilidade devida à variação natural da resposta nas unidades, como a pressão arterial. O EMEx leva em conta a variabilidade exógena, i.e., a variabilidade causada pela imprecisão do processo e/ou instrumento de medida, como no caso de experimentos sensoriais em que avaliadores classificam grãos de café (unidades) segundo um conjunto de características gustativas. Por isso, é razoável supor que o EMEx esteja associado à posição na permutação (i).

Incorporamos EMEn em (2.3) adicionando o vetor aleatório $\mathbf{E}^\bullet = (E_1^\bullet \dots E_s^\bullet \dots E_N^\bullet)'$, em que $\mathbb{E}_R(E_s^\bullet) = 0$, $\text{Var}_R(E_s^\bullet) = \sigma_s^{\bullet 2}$ e $\text{Cov}_R(E_s^\bullet, E_{s'}^\bullet) = 0$ para $s \neq s'$, ao vetor $\mathbf{y}^\bullet = (y_1^\bullet \dots y_s^\bullet \dots y_N^\bullet)'$ e considerando o modelo

$$(\mathbf{Y}^* \ \mathbf{X}) = \mathbf{U}(\mathbf{y} \ \mathbf{x}) + \mathbf{U} \begin{pmatrix} \mathbf{E} & \mathbf{0} \\ & \mathbf{I}_{N \times 1} \end{pmatrix} \quad (2.7)$$

em que $\mathbf{Y}^* = (Y_1^* \dots Y_i^* \dots Y_N^*)'$ e $\mathbf{E} = (E_1 \dots E_s \dots E_N)'$, com $E_s = w_s E_s^\bullet$, é um vetor aleatório de EMEn. Note que $\mathbf{y} = (w_1 y_1^\bullet \dots w_s y_s^\bullet \dots w_N y_N^\bullet)'$. Então, segue que $\mathbb{E}_R(E_s) = 0$, $\text{Var}_R(E_s) = \sigma_s^2 = w_s^2 \sigma_s^{\bullet 2}$ e $\text{Cov}_R(E_s, E_{s'}) = 0$ para $s \neq s'$. O sub-índice R denota esperança com respeito à distribuição dos erros de medida. Para a posição i dentro de uma permutação, temos

$$Y_i^* = \sum_{s=1}^N U_{is} (y_s + E_s) = \mu_y + d_{iy} + \sum_{s=1}^N U_{is} E_s,$$

em que d_{iy} definido em (2.4) é o efeito aleatório da unidade.

De maneira semelhante a Li, Stanek & Singer (2008), ao concatenar \mathbf{Y}^* e \mathbf{X} usando o operador vec , obtemos o seguinte modelo misto para populações finitas com EMEn e informação auxiliar

$$\mathbf{Z}^* = \text{vec}[(\mathbf{Y}^* \ \mathbf{X})] = (\mathbf{I}_2 \otimes \mathbf{1}_N) \boldsymbol{\mu} + \mathbf{d} + \mathbf{E}^* \quad (2.8)$$

em que $\boldsymbol{\mu}$, definido em (2.6), é um vetor de médias populacionais cujos componentes representam os efeitos fixos, $\mathbf{d} = (\mathbf{d}'_y \ \mathbf{d}'_x)'$ é o vetor de efeitos aleatórios com $\mathbf{d}_y = (d_{1y} \dots d_{iy} \dots d_{Ny})'$, $\mathbf{d}_x = (d_{1x} \dots d_{ix} \dots d_{Nx})'$, e $\mathbf{E}^* = \begin{pmatrix} \mathbf{E}_y^* & \mathbf{0}' \\ & \mathbf{I}_{N \times 1} \end{pmatrix}'$ e $\mathbf{E}_y^* = \mathbf{Y}^* - \mathbf{U}\mathbf{y} = \mathbf{U}\mathbf{E}$.

No Apêndice A (pág. 79), demonstramos que

$$\mathbb{E}_{SR}(\mathbf{Z}^*) = (\mathbf{I}_2 \otimes \mathbf{1}_N) \boldsymbol{\mu} \quad (2.9)$$

e

$$\text{Var}_{SR}(\mathbf{Z}^*) = \boldsymbol{\Sigma} \otimes \mathbf{P}_N + \mathbf{e}_1 \mathbf{e}_1' \otimes \bar{\sigma}_e^2 \mathbf{I}_N, \quad (2.10)$$

com \mathbf{P}_a definido em (2.5), \mathbf{e}_i denotando a coluna i -ésima da matriz \mathbf{I}_2 , e $\bar{\sigma}_e^2 = N^{-1} \sum_{s=1}^N \sigma_s^2$. Salientamos que (2.9) e (2.10) é consequência do processo de aleatorização (ou processo da amostragem) empregado.

De outro lado, incorporamos EMEx em (2.3) adicionando o vetor aleatório $\tilde{\mathbf{E}}_y^\bullet = (\tilde{E}_1^\bullet \dots \tilde{E}_i^\bullet \dots \tilde{E}_N^\bullet)'$, em que $\mathbb{E}_R(\tilde{E}_i^\bullet) = 0$, $\text{Var}_R(\tilde{E}_i^\bullet) = \tilde{\sigma}_i^{\bullet 2}$ e $\text{Cov}_R(\tilde{E}_i^\bullet, \tilde{E}_{i'}^\bullet) = 0$ para $i \neq i'$, ao vetor $(\mathbf{U}\mathbf{y}^\bullet)$. Seja $\tilde{\mathbf{E}}_y = (\tilde{E}_1 \dots \tilde{E}_i \dots \tilde{E}_N)'$, em que $\tilde{E}_i = \tilde{E}_i^\bullet \sum_{s=1}^N U_{is} w_s$, um vetor aleatório; logo $\mathbb{E}_{SR}(\tilde{E}_i) = 0$, $\text{Var}_{SR}(\tilde{E}_i) = \tilde{\sigma}_i^2 = \mu_w \tilde{\sigma}_i^{\bullet 2}$ com $\mu_w = N^{-1} \sum_{s=1}^N w_s^2$ e $\text{Cov}_{SR}(\tilde{E}_i, \tilde{E}_{i'}) = 0$ para $i \neq i'$.

Então, para a posição i numa permutação,

$$\tilde{Y}_i = \left(\sum_{s=1}^N U_{is} y_s \right) + \tilde{E}_i = \mu_y + d_{iy} + \tilde{E}_i,$$

com d_{iy} definido em (2.4). Logo, podemos escrever

$$\tilde{\mathbf{Y}} = \mu_y \mathbf{1}_N + \mathbf{d}_y + \tilde{\mathbf{E}}_y$$

com \mathbf{d}_y definido em (2.8) e $\tilde{\mathbf{Y}} = (\tilde{Y}_1 \dots \tilde{Y}_i \dots \tilde{Y}_N)'$.

Para relacionar $\tilde{\mathbf{Y}}$ e \mathbf{X} definida em (2.1) com \mathbf{y} e \mathbf{x} definidos em (2.2), consideramos

$$\begin{pmatrix} \tilde{\mathbf{Y}} & \mathbf{X} \end{pmatrix} = \mathbf{U}(\mathbf{y} \ \mathbf{x}) + \begin{pmatrix} \tilde{\mathbf{E}}_y & \mathbf{0} \\ & \mathbf{I}_{N \times 1} \end{pmatrix}. \quad (2.11)$$

O correspondente modelo misto para populações finitas com EMEx e informação auxiliar é

$$\tilde{\mathbf{Z}} = \text{vec} \left[\begin{pmatrix} \tilde{\mathbf{Y}} & \mathbf{X} \end{pmatrix} \right] = (\mathbf{I}_2 \otimes \mathbf{1}_N) \boldsymbol{\mu} + \mathbf{d} + \tilde{\mathbf{E}} \quad (2.12)$$

com $\boldsymbol{\mu}$ e \boldsymbol{d} como em (2.8), e $\tilde{\boldsymbol{E}} = \begin{pmatrix} \tilde{\boldsymbol{E}}'_y & \mathbf{0}' \\ & \mathbf{1}_{1 \times N} \end{pmatrix}'$ denotando o vetor de EMEx. Daí segue que (ver Apêndice A, pág. 80)

$$\mathbb{E}_{SR}(\tilde{\boldsymbol{Z}}) = (\boldsymbol{I}_2 \otimes \mathbf{1}_N) \boldsymbol{\mu} \quad (2.13)$$

e

$$\text{Var}_{SR}(\tilde{\boldsymbol{Z}}) = \boldsymbol{\Sigma} \otimes \boldsymbol{P}_N + \boldsymbol{e}_1 \boldsymbol{e}'_1 \otimes \bigoplus_{i=1}^N \tilde{\sigma}_i^2 \quad (2.14)$$

com $\boldsymbol{\Sigma}$ definido em (2.6) e \boldsymbol{e}_i , em (2.10).

2.2 Combinações lineares e preditores ótimos

Nosso interesse é prever ou estimar combinações lineares da forma

$$T = \sum_{i=1}^N c_i Y_i, \quad (2.15)$$

em que os c_i são constantes conhecidas associadas aos rótulos i (posições)^a. Por exemplo, tomando $w_s = 1$ para todo s , se o interesse é o total populacional, $T = \sum_{i=1}^N Y_i$, tomamos $c_i = 1$ para todo i , e se o interesse é a média populacional, $T = \mu_y$, tomamos $c_i = N^{-1}$ para todo i . Alternativamente, para prever o valor latente da unidade na j -ésima posição dentro da permutação, $T = Y_j$, adotamos $c_i = 1$ para $i = j$ e $c_i = 0$ em caso contrário. Mais à frente, mostramos que o parâmetro B definido em (1.1) pode ser escrito de maneira semelhante.

Para obter o preditor de T definido em (2.15) com base numa amostra aleatória simples selecionada sem reposição, vamos considerar, inicialmente, o modelo (2.8) e posteriormente o modelo (2.12). O procedimento baseia-se em três passos como mencionam Li, Stanek & Singer (2008).

No primeiro passo, assumindo que o valor de x_s é conhecido para cada uma das N unidades, pre-multiplicamos cada termo em (2.8) por $\boldsymbol{R} = \text{diag}\{\boldsymbol{I}_N, \boldsymbol{P}_N\}$, em que $\text{diag}\left\{\begin{matrix} \boldsymbol{A}_1 & \dots & \boldsymbol{A}_m \\ a_1 \times b_1 & & a_m \times b_m \end{matrix}\right\}$ é uma matriz diagonal em blocos com os \boldsymbol{A}_i representando os blocos, para eliminar μ_x (porque é

^a Dependendo da escolha de c_i , a quantidade T pode representar um parâmetro ou uma variável aleatória. Embora o termo estimador seja tipicamente usado para um parâmetro, e o termo preditor para uma variável aleatória, podemos usar o termo preditor para ambos, enfatizando o fato de que a inferência é realizada com respeito à predição de combinações lineares das variáveis aleatórias não observadas, ainda que o interesse seja prever um parâmetro.

conhecida). Logo, $\mathbf{RZ}^* = \mu_y [\mathbf{e}_1 \otimes \mathbf{1}_N] + \mathbf{d} + \mathbf{E}^*$. Sob este modelo, $\mathbb{E}_{SR}(\mathbf{RZ}^*) = \mu_y [\mathbf{e}_1 \otimes \mathbf{1}_N]$ e $\text{Var}_{SR}(\mathbf{RZ}^*) = \text{Var}_{SR}(\mathbf{Z}^*)$.

No segundo passo, que representa o processo da amostragem aleatória simples, decompos $(\mathbf{Y}^* \ \mathbf{X})$ em duas partes, uma que corresponde à amostra (e que pode ser tomada, sem perda de generalidade, como as n primeiras linhas) e outra a parte restante (as últimas $N - n$ linhas). Denotando

$$\mathbf{Z}_I^* = [Y_1^* \ \dots \ Y_n^* \ (X_1 - \mu_x) \ \dots \ (X_n - \mu_x)]' \quad (2.16)$$

e

$$\mathbf{Z}_{II}^* = [Y_{n+1}^* \ \dots \ Y_N^* \ (X_{n+1} - \mu_x) \ \dots \ (X_N - \mu_x)]',$$

segue que

$$\begin{aligned} \mathbb{E}_{SR}(\mathbf{Z}_I^*) &= \mu_y [\mathbf{e}_1 \otimes \mathbf{1}_n], \\ \mathbb{E}_{SR}(\mathbf{Z}_{II}^*) &= \mu_y [\mathbf{e}_1 \otimes \mathbf{1}_{N-n}], \\ \text{Var}_{SR}(\mathbf{Z}_I^*) &= \mathbf{V}_I^* = \boldsymbol{\Sigma} \otimes \mathbf{P}_{n,N} + \mathbf{e}_1 \mathbf{e}_1' \otimes \bar{\sigma}_e^2 \mathbf{I}_n, \\ \text{Var}_{SR}(\mathbf{Z}_{II}^*) &= \mathbf{V}_{II}^* = \boldsymbol{\Sigma} \otimes \mathbf{P}_{N-n,N} + \mathbf{e}_1 \mathbf{e}_1' \otimes \bar{\sigma}_e^2 \mathbf{I}_{N-n}, \\ \text{Cov}_{SR}(\mathbf{Z}_I^*, \mathbf{Z}_{II}^*) &= \mathbf{V}_{I,II}^* = -N^{-1} \boldsymbol{\Sigma} \otimes \mathbf{J}_{n \times (N-n)} \end{aligned} \quad (2.17)$$

com $\mathbf{P}_{a,b}$ definido em (2.5) e $\mathbf{J}_{a \times b} = \mathbf{1}_a \mathbf{1}_b'$.

Como ilustração, consideremos uma população composta por $N = 5$ unidades da qual selecionamos uma amostra aleatória simples de tamanho $n = 3$ sem reposição. Seja $w_s = 1$ para todo s ,

$$\mathbf{x} = (1 \ 3 \ 4 \ 10 \ 9)', \quad \mathbf{y} = (1 \ 2.5 \ 5 \ 6.5 \ 4)', \quad \mathbf{E} = (E_1 \ E_2 \ E_3 \ E_4 \ E_5)'$$

e suponhamos que o erro de medida endógeno possa tomar apenas dois valores igualmente prováveis, dados por mais ou menos σ_s , com $\sigma_1 = 3$, $\sigma_2 = 1$, $\sigma_3 = 7$, $\sigma_4 = 10$ e $\sigma_5 = 2$. Então $\mu_x = 5.4$, $\mu_y = 3.8$, $\sigma_x^2 = 15.3$, $\sigma_y^2 = 4.575$, $\sigma_{yx} = 6.725$ e $\bar{\sigma}_e^2 = 32.6$. Portanto,

$$\begin{aligned} \mathbb{E}_{SR}(\mathbf{Z}_I^*) &= 3.8 [(1 \ 0)' \otimes \mathbf{1}_3], \\ \mathbb{E}_{SR}(\mathbf{Z}_{II}^*) &= 3.8 [(1 \ 0)' \otimes \mathbf{1}_2], \\ \mathbf{V}_I^* &= \begin{pmatrix} 4.575 & 6.725 \\ 6.725 & 15.3 \end{pmatrix} \otimes \mathbf{P}_{3,5} + (1 \ 0)' (1 \ 0) \otimes 32.6 \mathbf{I}_3, \end{aligned}$$

$$\mathbf{V}_{II}^* = \begin{pmatrix} 4.575 & 6.725 \\ 6.725 & 15.3 \end{pmatrix} \otimes \mathbf{P}_{2,5} + (1 \ 0)'(1 \ 0) \otimes 32.6\mathbf{I}_2,$$

e

$$\mathbf{V}_{I,II}^* = -\frac{1}{5} \begin{pmatrix} 4.575 & 6.725 \\ 6.725 & 15.3 \end{pmatrix} \otimes \mathbf{J}_{3 \times 2}.$$

Se a realização da matriz aleatória \mathbf{U} é

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

as unidades 3, 4, 2, 5 e 1 ficam localizadas, respectivamente, na primeira, segunda, terceira, quarta e quinta posições na permutação.

Por conseguinte,

$$\begin{aligned} \mathbf{Y}_{\{\mathbf{U}=\mathbf{u}\}}^* &= (5 + E_3, 6.5 + E_4, 2.5 + E_2, 4 + E_5, 1 + E_1)', \\ \mathbf{X}_{\{\mathbf{U}=\mathbf{u}\}} &= (4 \ 10 \ 3 \ 9 \ 1)', \\ \mathbf{Z}_{I\{\mathbf{U}=\mathbf{u}\}}^* &= (5 + E_3, 6.5 + E_4, 2.5 + E_2, 4 - 5.4, 10 - 5.4, 3 - 5.4)', \\ \mathbf{Z}_{II\{\mathbf{U}=\mathbf{u}\}}^* &= (4 + E_5, 1 + E_1, 9 - 5.4, 1 - 5.4)'. \end{aligned}$$

Expressando (2.15) como

$$T = \sum_{i=1}^n c_i Y_i + \sum_{i=n+1}^N c_i Y_i,$$

basicamente temos que prever $\sum_{i=n+1}^N c_i Y_i$, pois $\sum_{i=1}^n c_i Y_i$ pode ser calculado diretamente (sem erro de medida) ou indiretamente (com erro de medida) a partir dos dados amostrais.

Finalmente, no terceiro passo, minimizamos a esperança do erro quadrático médio (EQM) do preditor linear sujeito à restrição de que seja não viciado [ver Apêndice B, ítem ii), pág. 87]. O

preditor obtido é

$$\begin{aligned} \hat{T}^* &= nD\bar{Y}_I^* + k^* \sum_{i=1}^n c_i Y_i^* \\ &+ (\sigma_{yx}/\sigma_x^2) \left[n(D + \bar{c}_{II})(\mu_x - \bar{X}_I) - (1 - k^*) \sum_{i=1}^n c_i (\mu_x - X_i) \right], \end{aligned} \quad (2.18)$$

em que μ_x , σ_y^2 , σ_x^2 e σ_{yx} estão definidas em (2.6),

- $D = (1/f - 1)\bar{c}_{II} + (1 - k^*)\bar{c}_I$ com $f = n/N$, $\bar{c}_I = n^{-1} \sum_{i=1}^n c_i$ e $\bar{c}_{II} = (N - n)^{-1} \sum_{i=n+1}^N c_i$,
- $k^* = \sigma_y^2 (1 - \rho_{yx}^2) / [\bar{\sigma}_e^2 + \sigma_y^2 (1 - \rho_{yx}^2)]$ com $\bar{\sigma}_e^2$ definida em (2.10),
- $\rho_{yx} = \sigma_{yx}/(\sigma_y \sigma_x)$ denotando o coeficiente de correlação entre y e x (assumimos que $|\rho_{yx}| < 1$),
- $\bar{Y}_I^* = n^{-1} \sum_{i=1}^n Y_i^*$ e $\bar{X}_I = n^{-1} \sum_{i=1}^n X_i$.

A variância do preditor \hat{T}^* é (ver Apêndice B, pág. 89)

$$\begin{aligned} \text{Var}_{SR}(\hat{T}^* - T) &= \sigma_y^2 (1 - \rho_{yx}^2) \left(\frac{nD^2}{k^*} + (1 - k^*) \sum_{i=1}^n c_i^2 + (N - n)\bar{c}_{II}^2 \right) \\ &+ (N - n - 1)\sigma_y^2 \sigma_{c_{II}}^2, \end{aligned} \quad (2.19)$$

com $\sigma_{c_{II}}^2 = (N - n - 1)^{-1} \sum_{i=n+1}^N (c_i - \bar{c}_{II})^2$.

O correspondente preditor sob o modelo (2.12) é obtido de maneira semelhante e é

$$\begin{aligned} \tilde{T} &= n\tilde{D}\tilde{\mu}_{Y_I} + \sum_{i=1}^n c_i \tilde{k}_i \tilde{Y}_i \\ &+ (\sigma_{yx}/\sigma_x^2) \left[n\tilde{D}(\mu_x - \tilde{\mu}_{X_I}) - \sum_{i=1}^n c_i (1 - \tilde{k}_i) (\mu_x - X_i) + n\bar{c}_{II}(\mu_x - \bar{X}_I) \right], \end{aligned} \quad (2.20)$$

em que μ_x , σ_y^2 , σ_x^2 e σ_{yx} estão definidas em (2.6); \bar{c}_{II} , ρ_{yx} e \bar{X}_I , em (2.18), com

$$\begin{aligned}\tilde{D} &= (1/f - 1)\bar{c}_{II} + n^{-1} \sum_{i=1}^n c_i (1 - \tilde{k}_i), \quad f = n/N, \\ \tilde{k}_i &= \sigma_y^2 (1 - \rho_{yx}^2) / [\bar{\sigma}_i^2 + \sigma_y^2 (1 - \rho_{yx}^2)], \\ \tilde{\mu}_{Y_I} &= \left(\sum_{i=1}^n \tilde{k}_i \right)^{-1} \sum_{i=1}^n \tilde{k}_i \tilde{Y}_i, \quad \text{e} \\ \tilde{\mu}_{X_I} &= \left(\sum_{i=1}^n \tilde{k}_i \right)^{-1} \sum_{i=1}^n \tilde{k}_i X_i.\end{aligned}$$

A variância do preditor \tilde{T} (Ver Apêndice B, pág. 89) é

$$\begin{aligned}Var_{SR}(\tilde{T} - T) &= \sigma_y^2 (1 - \rho_{yx}^2) \left(\frac{n^2 \tilde{D}^2}{\sum_{i=1}^n \tilde{k}_i} + \sum_{i=1}^n c_i^2 (1 - \tilde{k}_i) + (N - n) \bar{c}_{II}^2 \right) \\ &\quad + (N - n - 1) \sigma_y^2 \sigma_{C_{II}}^2,\end{aligned}\tag{2.21}$$

com $\sigma_{C_{II}}^2$ definida em (2.19).

Para obter um estimador da resposta média populacional (ponderada pelos w_s , $s = 1, \dots, n$), i.e.,

$$T = N^{-1} \sum_{s=1}^N y_s = N^{-1} \sum_{i=1}^N Y_i = \mu_y,\tag{2.22}$$

podemos empregar (2.18) ou (2.20) dependendo da natureza do erro de medida, usando $c_i = N^{-1}$, $i = 1, \dots, N$. Sob o modelo com EMEn (2.8), o estimador de T (2.22) é

$$\hat{T}^* = \bar{Y}_I^* + (\sigma_{yx}/\sigma_x^2) \{\mu_x - \bar{X}_I\}\tag{2.23}$$

e sua variância é

$$Var_{SR}(\hat{T}^* - T) = n^{-1} \left[\sigma_y^2 (1 - \rho_{yx}^2) (1 - f) + \bar{\sigma}_e^2 \right].$$

Note que a constante de encolhimento $k^* = \sigma_y^2 (1 - \rho_{yx}^2) / [\bar{\sigma}_e^2 + \sigma_y^2 (1 - \rho_{yx}^2)]$ não aparece em (2.23) e nem em sua variância porque é cancelada na dedução.

Sob o modelo com EMEx (2.12), o estimador de T (2.22) é

$$\tilde{T} = \tilde{\mu}_{Y_I} + (\sigma_{yx}/\sigma_x^2) \{\mu_x - \tilde{\mu}_{X_I}\} \quad (2.24)$$

e sua variância é

$$\text{Var}_{SR}(\tilde{T} - T) = \sigma_y^2(1 - \rho_{yx}^2) \left(\frac{N - \sum_{i=1}^n \tilde{k}_i}{N \sum_{i=1}^n \tilde{k}_i} \right).$$

Consideremos agora alguns casos particulares de (2.23) e de (2.24):

a) $\sigma_{yx} = 0$, i.e., quando não tivermos variáveis auxiliares, (2.23) pode ser escrito como

$$\hat{T}^* = \bar{Y}_I^* \quad (2.25)$$

com

$$\text{Var}_{SR}(\hat{T}^* - T) = n^{-1} [\sigma_y^2(1 - f) + \bar{\sigma}_e^2] = n^{-1}\sigma_y^2(1 - f) + n^{-1}\bar{\sigma}_e^2;$$

e (2.24) pode ser escrito como

$$\tilde{T} = \tilde{\mu}_{Y_I} \quad (2.26)$$

com

$$\text{Var}_{SR}(\tilde{T} - T) = \sigma_y^2 \left(\frac{N - \sum_{i=1}^n \tilde{k}_i}{N \sum_{i=1}^n \tilde{k}_i} \right).$$

Em (2.26), $\tilde{k}_i = \sigma_y^2 / \{\tilde{\sigma}_i^2 + \sigma_y^2\}$.

b) $\sigma_{yx} \neq 0$, $\bar{\sigma}_e^2 = 0$, e, $\tilde{\sigma}_i^2 = 0$ para todo i , i.e., quando tivermos uma variável auxiliar mas não tivermos erros de medida, temos que \hat{T}^* (2.18) é igual a \tilde{T} (2.20); conseqüentemente, definindo $\bar{Y}_I = n^{-1} \sum_{i=1}^n Y_i$,

$$\hat{T} = \bar{Y}_I + (\sigma_{yx}/\sigma_x^2) \{\mu_x - \bar{X}_I\} \quad (2.27)$$

com

$$\text{Var}_s \left(\hat{T} - T \right) = n^{-1} \sigma_y^2 (1 - \rho_{yx}^2) (1 - f);$$

c) $\sigma_{yx} = 0$, $\bar{\sigma}_e^2 = 0$, e, $\tilde{\sigma}_i^2 = 0$ para todo i , i.e., quando não tivermos variáveis auxiliares nem erros de medida,

$$\hat{T} = \bar{Y}_I \quad (2.28)$$

com

$$\text{Var}_s \left(\hat{T} - T \right) = n^{-1} (1 - f) \sigma_y^2.$$

O estimador da resposta média populacional apresentado em (2.25), i.e., quando não tivermos variáveis auxiliares mas tivermos erro de medida endógeno, é idêntico ao estimador dado por Bolfarine (1991) e Bolfarine & Zacks (1992, p. 151) para o modelo de localização com erro nas variáveis sob o enfoque de super-população, ou àquele dado por Särndal, Swensson & Wretman (1992, p. 605) sob o que eles denominam modelo de medidas simples “the simple measurement model”, e sua variância é a variância da média amostral sob amostragem aleatória simples sem reposição mais um fator que depende da média das variâncias dos erros de medida endógenos. O estimador apresentado em (2.26), i.e., quando não tivermos variáveis auxiliares mas tivermos erro de medida exógeno, é idêntico àquele obtido via mínimos quadrados ponderados, como parte da solução das equações de modelos mistos “mixed model equations” dadas por Henderson, Kempthorne, Searle & von Krosigk (1959) e revisado por Robinson (1991). O estimador apresentado em (2.27), i.e., quando não tivermos erro de medida mas tivermos variáveis auxiliares, corresponde ao estimador apresentado por Li, Stanek & Singer (2008), e o estimador apresentado em (2.28), i.e., quando não tivermos erro de medida nem variáveis auxiliares, corresponde ao estimador- π ($= \bar{y}_\pi$) apresentado em Särndal, Swensson & Wretman (1992, p. 68). O estimador (2.27) também coincide com aquele apresentado em Cassel et al. (1976) e em Wu & Sitter (2001). Um outro caso em que \hat{T}^* definido em (2.18) é igual a \tilde{T} definido em (2.20) é aquele no qual os erros de medida são homocedásticos.

Finalmente, sob o modelo com EME_X, e lembrando que \tilde{k}_i não é aleatório, temos

$$\mathbb{E}_{SR} \left(\tilde{\mu}_{Y_I} \right) = \left(\sum_{i=1}^n \tilde{k}_i \right)^{-1} \sum_{i=1}^n \tilde{k}_i \left[\mathbb{E}_S \left(\sum_{s=1}^N U_{is} y_s \right) + \mathbb{E}_R \left(\tilde{E}_i \right) \right] = \mu_y$$

e

$$\mathbb{E}_{SR}(\tilde{\mu}_{X_I}) = \left(\sum_{i=1}^n \tilde{k}_i \right)^{-1} \sum_{i=1}^n \tilde{k}_i \left[\mathbb{E}_S \left(\sum_{s=1}^N U_{is} x_s \right) \right] = \mu_x;$$

logo, independentemente dos valores de $\tilde{\sigma}_i^2$, \tilde{T} (2.20) é um estimador não viciado de T (2.15).

Para prever o valor latente (ponderado) da unidade na j -ésima posição dentro da permutação, i.e., $T = Y_j$, adotamos $c_i = 1$ para $i = j$ e $c_i = 0$ em caso contrário. Como esperávamos, se não há erro de medida, i.e., $\tilde{\sigma}_e^2 = 0$, e, $\tilde{\sigma}_j^2 = 0$, quando $j \leq n$, $\hat{Y}_j = Y_j$ e $\text{Var}_S(\hat{Y}_j - Y_j) = 0$. Se $j > n$, o preditor de Y_j e sua variância sob diferentes hipóteses sobre σ_{yx} são

- $\sigma_{yx} \neq 0$, i.e., quando tivermos uma variável auxiliar,

$$\hat{Y}_j = \bar{Y}_I + [N/(N-n)] (\sigma_{yx}/\sigma_x^2) (\mu_x - \bar{X}_I) \quad (2.29)$$

com

$$\text{Var}_S(\hat{Y}_j - Y_j) = n^{-1} [N/(N-n)] \sigma_y^2 (1 - \rho_{yx}^2) + [1 - (N-n)^{-1}] \sigma_y^2.$$

- $\sigma_{yx} = 0$, i.e., quando não tivermos variáveis auxiliares,

$$\hat{Y}_j = \bar{Y}_I \quad (2.30)$$

com

$$\text{Var}_S(\hat{Y}_j - Y_j) = n^{-1} [N/(N-n)] \sigma_y^2 + [1 - (N-n)^{-1}] \sigma_y^2.$$

Sob o modelo com EMEn (2.8), o preditor de Y_j e sua variância, quando $j \leq n$, é

$$\hat{Y}_j = \bar{Y}_I^* + k^* (Y_j^* - \bar{Y}_I^*) + (\sigma_{yx}/\sigma_x^2) (1 - k^*) (X_j - \bar{X}_I) \quad (2.31)$$

com

$$\text{Var}_{SR}(\hat{Y}_j - Y_j) = n^{-1} \tilde{\sigma}_e^2 [k^* (n-1) + 1],$$

e quando $j > n$,

$$\hat{Y}_j = \bar{Y}_I^* + [N/(N-n)] (\sigma_{yx}/\sigma_x^2) (\mu_x - \bar{X}_I) \quad (2.32)$$

com

$$\text{Var}_{SR} (\hat{Y}_j - Y_j) = n^{-1} \left\{ [N/(N-n)] \sigma_y^2 (1 - \rho_{yx}^2) + \bar{\sigma}_e^2 \right\} + [1 - (N-n)^{-1}] \sigma_y^2$$

com k^* definido em (2.18).

Sob o modelo com EMEx (2.12), o preditor de Y_j e sua variância, quando $j \leq n$, é

$$\hat{Y}_j = \tilde{\mu}_{Y_I} + \tilde{k}_j (\tilde{Y}_j - \tilde{\mu}_{Y_I}) + (\sigma_{yx}/\sigma_x^2) (1 - \tilde{k}_j) (X_j - \tilde{\mu}_{X_I}) \quad (2.33)$$

e

$$\text{Var}_{SR} (\hat{Y}_j - Y_j) = \tilde{\sigma}_j^2 \tilde{k}_j \left(\sum_{i=1}^n \tilde{k}_i \right)^{-1} \left(\sum_{i=1}^n \tilde{k}_i - \tilde{k}_j + 1 \right),$$

e quando $j > n$,

$$\hat{Y}_j = \tilde{\mu}_{Y_I} + (\sigma_{yx}/\sigma_x^2) \left[\frac{\mu_x - f \bar{X}_I}{1 - f} - \tilde{\mu}_{X_I} \right] \quad (2.34)$$

e

$$\text{Var}_{SR} (\hat{Y}_j - Y_j) = \sigma_y^2 \left\{ \left(\frac{1 - \rho_{yx}^2}{\sum_{i=1}^n \tilde{k}_i} \right) \left(1 + \frac{\sum_{i=1}^n \tilde{k}_i}{N - n} \right) + [1 - (N - n)^{-1}] \right\}$$

com $\tilde{\mu}_{Y_I}$, $\tilde{\mu}_{X_I}$ e \tilde{k}_i definidos em (2.20).

Se $\sigma_{yx} = 0$, i.e., quando não tivermos variáveis auxiliares, (2.31) pode ser escrito como

$$\hat{Y}_j = \bar{Y}_I^* + k^* (Y_j^* - \bar{Y}_I^*) \quad (2.35)$$

com

$$\text{Var}_{SR} (\hat{Y}_j - Y_j) = n^{-1} \bar{\sigma}_e^2 [k^* (n - 1) + 1],$$

e $k^* = \sigma_y^2 / (\bar{\sigma}_e^2 + \sigma_y^2)$, (2.32) pode ser escrito como

$$\hat{Y}_j = \bar{Y}_I^* \quad (2.36)$$

com

$$\text{Var}_{SR}(\hat{Y}_j - Y_j) = n^{-1} \left\{ [N / (N - n)] \sigma_y^2 + \bar{\sigma}_e^2 \right\} + [1 - (N - n)^{-1}] \sigma_y^2,$$

(2.33) pode ser escrito como

$$\hat{Y}_j = \tilde{\mu}_{Y_I} + \tilde{k}_j (\tilde{Y}_j - \tilde{\mu}_{Y_I}) \quad (2.37)$$

com

$$\text{Var}_{SR}(\hat{Y}_j - Y_j) = \tilde{\sigma}_j^2 \tilde{k}_j \left(\sum_{i=1}^n \tilde{k}_i \right)^{-1} \left(\sum_{i=1}^n \tilde{k}_i - \tilde{k}_j + 1 \right),$$

e $\tilde{k}_i = \sigma_y^2 / (\tilde{\sigma}_i^2 + \sigma_y^2)$, e (2.34) pode ser escrito como

$$\hat{Y}_j = \tilde{\mu}_{Y_I} \quad (2.38)$$

com

$$\text{Var}_{SR}(\hat{Y}_j - Y_j) = \sigma_y^2 \left\{ \left(\sum_{i=1}^n \tilde{k}_i \right)^{-1} \left(1 + \frac{\sum_{i=1}^n \tilde{k}_i}{N - n} \right) + [1 - (N - n)^{-1}] \right\},$$

e \tilde{k}_i definido em (2.37).

O preditor de Y_j apresentado em (2.35) é idêntico ao preditor dado por Stanek & Singer (2004) no caso particular de amostragem por conglomerados em um estágio e erro de medida. Esse estimador, (2.35), bem como o estimador dado em (2.37) também são apresentados por Stanek, Lencina, Singer, González, Li & San Martino (2008) sob o modelo misto para populações finitas.

2.3 Particularização dos resultados para estimação de uma medida de associação

Nesta Seção, além de particularizar (2.18) e (2.20) para estimar a medida de associação B (1.1), propomos os respectivos estimadores empíricos, uma vez que (2.18) e (2.20) envolvem variâncias e covariâncias populacionais que, na prática, são desconhecidas. Também examinamos o efeito de usar a informação auxiliar no cálculo das estimativas de B , dado que já foi incluída nos pesos; apresentamos exemplos numéricos para comparar o desempenho dos estimadores propostos frente ao estimador obtido via mínimos quadrados ordinários, \hat{B}_{MQO} ; apresentamos os resultados de um estudo do desempenho dos estimadores propostos assumindo erros de medida homocedásticos com base em amostras independentes, de tamanho igual, selecionadas dos dados referidos na Seção 1.1, e apresentamos um estudo de simulação em que são considerados diferentes valores para o coeficiente de correlação entre a variável auxiliar e a variável resposta, diferentes valores de n , erros de medida homocedásticos e heterocedásticos e duas formas da distribuição da variável auxiliar.

Para estimar a medida de associação B , lembramos que (1.1) pode ser escrito como (2.15) tomando $c_i = 1$, $i = 1, \dots, N$, e usando $w_s = (x_s - \mu_x) / [(N - 1)\sigma_x^2]$. Dessa forma, o estimador é

$$\hat{B}^* = N\hat{T}^* \quad \text{sob EMEn,} \quad (2.39)$$

ou

$$\tilde{B} = N\tilde{T} \quad \text{sob EMEx,} \quad (2.40)$$

em que \hat{T}^* e \tilde{T} estão definidas em (2.23) e (2.24), respectivamente. Quando σ_{yx} e σ_x^2 são conhecidas, denotamos o estimador de B por \hat{B}^* (respectivamente, \tilde{B}); quando essas quantidades são estimadas a partir da amostra denotamos o correspondente estimador por \hat{B}_e^* (respectivamente, \tilde{B}_e), i.e.,

$$\hat{B}_e^* = N \left[\bar{Y}_I^* + (\hat{\sigma}_{yx} / \hat{\sigma}_x^2) \{ \mu_x - \bar{X}_I \} \right] \quad (2.41)$$

ou

$$\tilde{B}_e = N \left[\hat{\mu}_{Y_I} + (\tilde{\sigma}_{yx} / \hat{\sigma}_x^2) \{ \mu_x - \hat{\mu}_{X_I} \} \right], \quad (2.42)$$

em que

$$\begin{aligned}\hat{\sigma}_{yx} &= (n-1)^{-1} \sum_{i=1}^n (Y_i^* - \bar{Y}_I^*) (X_i - \bar{X}_I), \\ \hat{\sigma}_x^2 &= (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X}_I)^2, \\ \tilde{\sigma}_{yx} &= (n-1)^{-1} \sum_{i=1}^n (\tilde{Y}_i - \tilde{\bar{Y}}_I) (X_i - \bar{X}_I), \quad \tilde{\bar{Y}}_I = n^{-1} \sum_{i=1}^n \tilde{Y}_i, \\ \hat{\mu}_{Y_I} &= \left(\sum_{i=1}^n \hat{k}_i \right)^{-1} \sum_{i=1}^n \hat{k}_i \tilde{Y}_i, \quad \hat{\mu}_{X_I} = \left(\sum_{i=1}^n \hat{k}_i \right)^{-1} \sum_{i=1}^n \hat{k}_i X_i, \\ \hat{k}_i &= \tilde{\sigma}_y^2 (1 - \tilde{\rho}_{yx}^2) / \left[\tilde{\sigma}_i^2 + \tilde{\sigma}_y^2 (1 - \tilde{\rho}_{yx}^2) \right], \\ \tilde{\sigma}_y^2 &= (n-1)^{-1} \sum_{i=1}^n (\tilde{Y}_i - \tilde{\bar{Y}}_I)^2, \quad \text{e} \\ \tilde{\rho}_{yx} &= \tilde{\sigma}_{yx} / (\tilde{\sigma}_y \hat{\sigma}_x).\end{aligned}$$

As expressões (2.39) e (2.40) são idênticas se os erros de medida são homocedásticos.

Para propósitos ilustrativos, consideramos a população descrita na Seção 2.2, porém suponhamos que o erro de medida endógeno possa tomar apenas dois valores igualmente prováveis, dados por mais ou menos $\sigma_s^\bullet = 3$, $s = 1, 2, 3, 4, 5$, e usamos $w_s = (x_s - \mu_x) / [(N-1)\sigma_x^2]$. Seleccionamos amostras de tamanho $n = 3$. Na Tabela 2.1, apresentamos os valores populacionais da variável auxiliar (x_s), os valores latentes não ponderados (y_s^\bullet), além das possíveis respostas e variâncias dos EMEn, E_s . O gráfico de dispersão (populacional) e a reta de regressão estimada via mínimos quadrados ordinários estão apresentados na Figura 2.1.

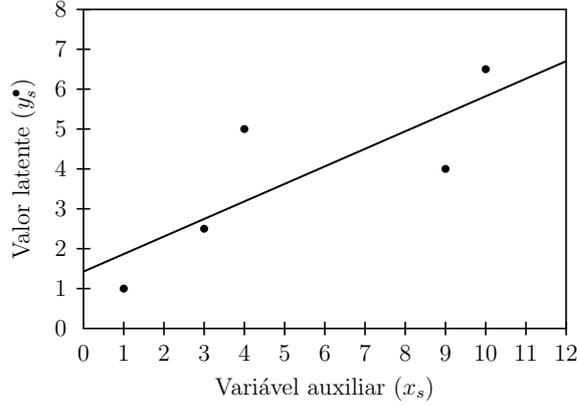
Tabela 2.1: População finita considerando variâncias iguais para os erros de medida endógenos.

Rótulos das Unidades(s)	x_s	y_s^\bullet	Possíveis Respostas	σ_s^2
1	1	1	-2 ou 4	0.0465
2	3	2.5	-0.5 ou 5.5	0.0138
3	4	5	2 ou 8	0.0047
4	10	6.5	3.5 ou 9.5	0.0508
5	9	4	1 ou 7	0.0311

$$\rho_{yx}^\bullet = 0.80$$

Neste caso, $N = 5$, $n = 3$, $\sigma_s^{\bullet 2} = 9$, $\sigma_x^2 = 15.3$, $\sigma_{yx} / \sigma_x^2 = 0.0622$, $\mu_x = 5.4$, $\sigma_y^2 = 0.0708$ e $\rho_{yx} = 0.9141$. O parâmetro populacional de interesse é $B = 0.4395$. Na Tabela 2.2 apresentamos

Figura 2.1: Gráfico de dispersão e reta de regressão calculada via mínimos quadrados ordinários ($y = 1.4265 + 0.4395x$) para o exemplo ilustrativo apresentado na Tabela 2.1.



estimativas de B definida em (1.1) denotadas por \hat{B}^* (2.39), \hat{B}_e^* (2.41) e

$$\bar{B}^* = N\bar{Y}_I^* \quad (2.43)$$

com \bar{Y}_I^* definida em (2.18).

As estimativas, para as 960 ($= 5! \times 2^3$) possíveis permutações das unidades e de suas respostas, estão esquematicamente apresentadas na Tabela 2.2 que também inclui a média das estimativas, o vício relativo [$=100 \times (\text{estimativa} - B)/B$], a contribuição relativa do vício ao erro quadrático médio (EQM) [$=100 \times (\text{estimativa} - B)^2/\text{EQM}$] e o erro quadrático médio (EQM) além das estimativas via mínimos quadrados ordinários, obtida de \hat{B}_{MQO} . Podemos perceber que o menor EQM está associado a \hat{B}^* e o maior a \hat{B}_{MQO} ; além disso, tanto \hat{B}_e^* quanto \hat{B}_{MQO} são estimadores viciados de B .

A seguir, consideramos duas variações do exemplo anterior. Na primeira, consideramos variâncias diferentes para os EMEn com $\sigma_1^2 = 9$, $\sigma_2^2 = 1$, $\sigma_3^2 = 49$, $\sigma_4^2 = 100$ e $\sigma_5^2 = 4$, e na segunda, variâncias diferentes para os EMEx, com $\tilde{\sigma}_1^2 = 100$, $\tilde{\sigma}_2^2 = 9$ e $\tilde{\sigma}_3^2 = 0.49$. Na Tabela 2.3 estão dispostos os valores populacionais da variável auxiliar (x_s), do valor latente não ponderado (y_s^*), das possíveis respostas e das variâncias dos E_s . Na Tabela 2.4, os valores das estimativas, para as 960 possíveis permutações das unidades e suas respostas, estão dispostos de forma esquemática. Observamos que as médias das estimativas, e portanto, as médias do vício relativo, mostrados nas Tabelas 2.2 e 2.4 são iguais. A relação entre os EQM na Tabela 2.4 reproduz aquela da Tabela 2.2, i.e.,

$$\text{EQM}(\hat{B}^*) < \text{EQM}(\hat{B}_e^*) < \text{EQM}(\bar{B}^*) < \text{EQM}(\hat{B}_{MQO}).$$

Tabela 2.2: Estimativas de B sob o modelo misto para populações finitas, com σ_{yx} e σ_x^2 conhecidas ou desconhecidas, estimativas de mínimos quadrados ordinários e média do EQM para $N = 5$ e $n = 3$ com variâncias iguais para os erros de medida endógenos.

Amostra	X_1	X_2	X_3	$w_s = 1$			$w_s = (x_s - \mu_x) / [(N - 1) \sigma_x^2]$			
				Y_1^*	Y_2^*	Y_3^*	\widehat{B}^*	\widehat{B}_e^*	\overline{B}^*	\widehat{B}_{MQO}
1	1	3	4	-2	-0.5	2	1.046	-0.665	0.196	1.250
2	1	3	4	-2	-0.5	8	0.817	-1.430	-0.033	2.964
...
959	4	10	3	8	9.5	-0.5	0.835	0.762	0.918	1.017
960	4	10	3	8	9.5	5.5	0.443	0.341	0.526	0.459
Média							0.4395	0.2789	0.4395	0.4722
Média do vício relativo							0.0%	-36.5%	0.0%	7.4%
$100 \times \text{vício}^2 / \text{EQM}$							0.0%	5.4%	0.0%	0.2%
Média das diferenças ao quadrado (EQM)							0.2839	0.4777	0.4811	0.5516

Tabela 2.3: População finita considerando variâncias diferentes para os erros de medida endógenos.

Rótulos das Unidades(s)	x_s	y_s^\bullet	Possíveis Respostas	σ_s^2
1	1	1	-2 ou 4	0.0465
2	3	2.5	1.5 ou 3.5	0.0015
3	4	5	-2 ou 12	0.0256
4	10	6.5	-3.5 ou 16.5	0.5650
5	9	4	2 ou 6	0.0138

Os resultados para a segunda variante estão dispostos na Tabela 2.5. Neste caso, as estimativas de B definida em (1.1) são apresentadas nas colunas rotuladas \widetilde{B} (2.40), \widetilde{B}_e (2.42),

$$\widetilde{B} = N\widetilde{\mu}_{Y_I},$$

e

$$\widetilde{B}_e = N\widehat{\mu}_{Y_I},$$

em que $\widetilde{\mu}_{Y_I}$ e $\widehat{\mu}_{Y_I}$ estão definidas em (2.20) e (2.42), respectivamente. O menor EQM está associado a \widetilde{B} e o maior a \widehat{B}_{MQO} ; além disso tanto \widetilde{B}_e quanto \widetilde{B} e \widehat{B}_{MQO} são estimadores viciados de B .

Também, com propósitos ilustrativos, consideremos um outro exemplo em que

1. Obtivemos 50000 amostras independentes de tamanho $n = 10$ via amostragem aleatória simples, da população de $N = 104$ estudantes descrita na Seção 1.1. Vamos admitir que os

Tabela 2.4: Estimativas de B sob o modelo misto para populações finitas, com σ_{yx} e σ_x^2 conhecidas ou desconhecidas, estimativas de mínimos quadrados ordinários e média do EQM para $N = 5$ e $n = 3$ com variâncias diferentes para os erros de medida endógenos.

Amostra	X_1	X_2	X_3	$w_s = 1$			$w_s = (x_s - \mu_x) / [(N - 1) \sigma_x^2]$			
				Y_1^*	Y_2^*	Y_3^*	\hat{B}^*	\hat{B}_e^*	\bar{B}^*	\hat{B}_{MQO}
1	1	3	4	-2	1.5	-2	1.068	-0.363	0.218	0.250
2	1	3	4	-2	1.5	12	0.534	-2.147	-0.316	4.250
...
959	4	10	3	12	16.5	1.5	1.429	1.233	1.511	1.657
960	4	10	3	12	16.5	3.5	1.298	1.092	1.381	1.471
Média							0.4395	0.2789	0.4395	0.4722
Média do vício relativo							0.0%	-36.5%	0.0%	7.4%
$100 \times \text{vício}^2 / \text{EQM}$							0.0%	2.2%	0.0%	0.1%
Média das diferenças ao quadrado (EQM)							1.1264	1.1712	1.3236	1.6619

Tabela 2.5: Estimativas de B sob o modelo misto para populações finitas, com σ_{yx} e σ_x^2 conhecidas ou desconhecidas, estimativas de mínimos quadrados ordinários e média do EQM para $N = 5$ e $n = 3$ com variâncias diferentes para os erros de medida exógenos.

Amostra	X_1	X_2	X_3	$w_s = 1$			$w_s = (x_s - \mu_x) / [(N - 1) \sigma_x^2]$				
				\tilde{Y}_1	\tilde{Y}_2	\tilde{Y}_3	\tilde{B}	\tilde{B}_e	$\bar{\tilde{B}}$	$\bar{\tilde{B}}_e$	\hat{B}_{MQO}
1	1	3	4	-9	-0.5	4.3	0.291	-2.415	0.088	0.306	4.407
2	1	3	4	-9	-0.5	5.7	0.173	-2.603	0.004	0.249	4.807
...
959	4	10	3	15	9.5	1.8	0.759	0.844	1.005	0.871	0.397
960	4	10	3	15	9.5	3.2	0.557	0.715	0.861	0.751	0.266
Média							0.4395	0.2852	0.4395	0.4959	0.4721
Média do vício relativo							0.0%	-35.1%	0.0%	12.8%	7.4%
$100 \times \text{vício}^2 / \text{EQM}$							0.0%	2.1%	0.0%	0.5%	0.05%
Média das diferenças ao quadrado (EQM)							0.1849	1.1413	0.5965	0.6846	1.9458

valores apresentados na Tabela F.3 (pág. 114, Apêndice F) na coluna rotulada “Cálculo”, que são as notas de Cálculo I, correspondem aos valores latentes das habilidades em Cálculo (que na realidade são desconhecidas). Neste passo, os valores latentes não ponderados são obtidos, i.e., $Y_i^\bullet = \sum_{s=1}^N U_{is} y_s^\bullet, i = 1, \dots, n$.

2. A cada $Y_i^\bullet, i = 1, \dots, n$, obtido em 1., foi somado um erro de medida (EMEn), gerado a partir de uma distribuição normal com média 0 e desvio padrão $\sigma_s^\bullet = \sigma_e^\bullet = 0.25$.
3. Foram avaliados os estimadores \hat{B}^* (2.39), \hat{B}_e^* (2.41), \bar{B}^* (2.43) e \hat{B}_{MQO} para cada uma

das 50000 amostras. Neste passo, usamos $w_s = (x_s - \mu_x) / [(N - 1) \sigma_x^2]$ para construir $Y_i^* = \sum_{s=1}^N U_{is} (y_s + E_s)$ em que $y_s = w_s y_s^\bullet$ e $E_s = w_s E_s^\bullet$, portanto $Y_i^* = \left[\sum_{s=1}^N U_{is} w_s \right] \times \left[\sum_{s=1}^N U_{is} (y_s^\bullet + E_s^\bullet) \right]$.

4. A média das 50000 estimativas obtidas de \hat{B}^* , \hat{B}_e^* , \bar{B}^* e \hat{B}_{MQO} , a média da contribuição relativa do vício ao EQM e o EQM foram calculados e dispostos na Tabela 2.6.
5. Os passos 1 a 4 foram repetidos para $n = 30$ e $n = 60$.
6. Os passos 1 a 5 foram repetidos para $\sigma_e^\bullet = 0.75$ e $\sigma_e^\bullet = 1.5$; também, consideramos o caso em que $\sigma_e^\bullet = 0$, ou seja, o caso em que não se tem erro de medida.

Na Tabela 2.6, para cada σ_e^\bullet , os menores valores para EQM estão associados a \hat{B}^* quando $n \geq 30$; em particular, $\text{EQM}(\hat{B}^*)$ é menor que $\text{EQM}(\hat{B}_{MQO})$, e $\text{EQM}(\bar{B}^*)$ é muito maior em todos os casos. Por outro lado, assim como esperávamos, quando n aumenta, os valores de EQM diminuem. Além disso, para \hat{B}_e^* e \hat{B}_{MQO} , o valor da média da contribuição relativa do vício ao EQM diminui ao aumentar n . Em geral $\text{EQM}(\hat{B}_e^*)$ é menor que $\text{EQM}(\hat{B}_{MQO})$ porém a contribuição relativa do vício ao EQM é maior.

Tabela 2.6: EQM ($\times 10^{-3}$) e média da contribuição relativa do vício ao EQM (dentro do parênteses) das estimativas.

σ_e^\bullet	n	\hat{B}^*	\hat{B}_e^*	\hat{B}_{MQO}	\bar{B}^*
0	10	8.31 (0.0%)	7.77 (2.0%)	11.56 (0.2%)	183.97 (0.0%)
	30	2.19 (0.0%)	2.30 (0.6%)	2.40 (0.1%)	48.40 (0.0%)
	60	0.65 (0.0%)	0.69 (0.2%)	0.67 (0.0%)	14.35 (0.0%)
0.25	10	8.52 (0.0%)	8.00 (1.9%)	11.84 (0.2%)	182.91 (0.0%)
	30	2.28 (0.0%)	2.39 (0.6%)	2.49 (0.1%)	48.60 (0.0%)
	60	0.69 (0.0%)	0.73 (0.2%)	0.71 (0.0%)	14.33 (0.0%)
0.75	10	10.54 (0.0%)	9.62 (1.6%)	14.79 (0.1%)	186.13 (0.0%)
	30	2.92 (0.0%)	3.01 (0.5%)	3.19 (0.1%)	48.96 (0.0%)
	60	1.02 (0.0%)	1.06 (0.1%)	1.05 (0.0%)	14.83 (0.0%)
1.5	10	17.08 (0.0%)	15.37 (1.1%)	24.23 (0.1%)	194.10 (0.0%)
	30	5.13 (0.0%)	5.17 (0.4%)	5.64 (0.1%)	51.65 (0.0%)
	60	2.13 (0.0%)	2.16 (0.1%)	2.19 (0.0%)	15.76 (0.0%)

$$\rho_{yx}^\bullet = 0.20$$

Para estudar o efeito do coeficiente de correlação, entre a variável auxiliar x_s e os valores latentes não ponderados y_s^\bullet , definido como $\rho_{yx}^\bullet = \sigma_{yx}^\bullet / (\sigma_y^\bullet \sigma_x)$ em que $\sigma_y^{\bullet 2} = (N - 1)^{-1} \sum_{s=1}^N (y_s^\bullet - \mu_y^\bullet)^2$ e

$\sigma_{yx}^\bullet = (N - 1)^{-1} \sum_{s=1}^N (y_s^\bullet - \mu_y^\bullet) (x_s - \mu_x)$, sobre o desempenho (EQM) dos estimadores propostos, \widehat{B}^* (2.39) e \widetilde{B} (2.40), e suas versões empíricas, \widehat{B}_e^* (2.41) e \widetilde{B}_e (2.42), frente ao desempenho do estimador obtido via mínimos quadrados ordinários, \widehat{B}_{MQO} , sendo abrangentes, realizamos simulações usando diferentes valores de ρ_{yx}^\bullet e consideramos dados com distribuição simétrica e dados com distribuição assimétrica in x , para $N = 2000$. O procedimento para gerar os dados, i.e., as populações finitas, é descrito na Seção 2.5. Usamos apenas valores positivos para ρ_{yx}^\bullet . Também, usamos diferentes valores de n e diferentes valores de σ_e^\bullet . Os resultados estão sintetizados nas Figuras C.1, C.2, C.3, C.4, C.5 e C.6 (pág. 91 a 95, Apêndice C), e mostram que

- EQM(\widehat{B}_{MQO}) diminui quando ρ_{yx}^\bullet aumenta nas Figuras C.1 (a) e C.5 (a).
- EQM(\widehat{B}^*) e EQM(\widehat{B}_e^*) (respectivamente, EQM(\widetilde{B}) e EQM(\widetilde{B}_e)) foram menores que EQM(\widehat{B}_{MQO}) nas Figuras C.2, C.3 e C.4 (respectivamente, C.6).
- EQM(\widehat{B}^*) e EQM(\widehat{B}_e^*) (respectivamente, EQM(\widetilde{B}) e EQM(\widetilde{B}_e)) aumentam quando ρ_{yx}^\bullet aumenta na Figura C.1 (respectivamente, C.5).
- Como esperávamos, a média do vício relativo de \widehat{B}^* e \widetilde{B} , e portanto, a contribuição relativa do vício ao EQM de \widehat{B}^* e \widetilde{B} , é praticamente zero (Figura C.1 à Figura C.6).
- A média do vício relativo de \widehat{B}_e^* , e portanto, a contribuição relativa do vício ao EQM, em todos os casos, diminui quando n aumenta (Figuras C.1, C.2, C.3 e C.4), e seu valor corresponde à média, sobre todas as permutações das unidades na população finita, dos valores

$$N [(\hat{\sigma}_{yx} / \hat{\sigma}_x^2) \{\mu_x - \bar{X}_I\}].$$

- A média do vício relativo de \widehat{B}_{MQO} é maior que a média do vício relativo de \widehat{B}_e^* [respectivamente, \widetilde{B}_e] quando foram consideradas populações com distribuição assimétrica em x , Figuras C.2 (b) e C.4 (b) [respectivamente, C.6 (b)].
- A contribuição relativa do vício ao EQM de \widehat{B}_e^* aumenta quando ρ_{yx}^\bullet aumenta e diminui quando n aumenta [Figures C.1 (c), C.2 (c), C.5 (c) e C.6 (c)]. Para \widehat{B}_{MQO} , esta ainda apresenta valores grandes para $n = 60$ quando são consideradas populações finitas com distribuição assimétrica em x [Figuras C.2 (c) e C.6 (c)].

Conseqüentemente, sugerimos:

- usar \widehat{B}^* e \widetilde{B} , segundo seja o caso, quando σ_{yx} e σ_x^2 são conhecidas, dado que \widehat{B}^* e \widetilde{B} são estimadores ótimos de B , i.e., são não viciados e apresentam variância mínima, e

- quando σ_{yx} e σ_x^2 são estimadas a partir da amostra, tendo em mente as limitações de nosso estudo de simulação e sabendo que são necessárias pesquisas adicionais para obter conclusões mais definitivas, usar \widehat{B}_e^* e \widetilde{B}_e para dados com distribuição assimétrica in x , \widehat{B}_e^* para dados com distribuição simétrica quando os erros de medida endógenos são homocedásticos, $\rho_{yx}^\bullet < 0.40$ e n é pequeno, e \widetilde{B}_e para dados com distribuição simétrica quando os erros de medida exógenos são heterocedásticos e $n \geq 30$.

2.4 Extensão para casos com múltiplas variáveis auxiliares

Em geral, várias variáveis auxiliares, x_{1s}, \dots, x_{qs} , podem estar associadas a cada unidade s juntamente com a resposta y_s . Neste caso, para estender (2.18) para obtenção do estimador de uma combinação linear de Y_i (2.15), $i = 1, \dots, N$, com EMEn é suficiente escrever

$$\mathbf{x} = \begin{pmatrix} x_{11} & \dots & x_{q1} & \dots & x_{Q1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1s} & \dots & x_{qs} & \dots & x_{Qs} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1N} & \dots & x_{qN} & \dots & x_{QN} \end{pmatrix}, \quad \mathbf{X} = \mathbf{U}\mathbf{x} = \begin{pmatrix} X_{11} & \dots & X_{q1} & \dots & X_{Q1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{1i} & \dots & X_{qi} & \dots & X_{Qi} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{1N} & \dots & X_{qN} & \dots & X_{QN} \end{pmatrix}$$

e

$$(\mathbf{Y}^* \ \mathbf{X}) = \mathbf{U} \left[(\mathbf{y} \ \mathbf{x}) + \begin{pmatrix} \mathbf{E} & \mathbf{0} \\ & \mathbf{I}_{N \times Q} \end{pmatrix} \right],$$

respectivamente, em que \mathbf{y} está definido em (2.2) e \mathbf{U} , em (2.3), e \mathbf{Y}^* e \mathbf{E} estão definidos em (2.7).

Como resultado, o estimador/preditor de T definido em (2.15) é

$$\widehat{T} = (\mathbf{g}'_I + \widehat{\mathbf{a}}') \mathbf{Z}_I^*,$$

com $\mathbf{g}_I = \mathbf{e}_1 \otimes \mathbf{C}_I$, em que \mathbf{e}_i denota a coluna i -ésima da matriz \mathbf{I}_{Q+1} , $\mathbf{C}_I = (c_1 \ \dots \ c_i \ \dots \ c_n)'$, c_i está definido em (2.15),

$$\mathbf{Z}_I^* = \left[Y_1^* \ \dots \ Y_n^* \ (X_{11} - \mu_{x_1}) \ \dots \ (X_{Qn} - \mu_{x_Q}) \right]', \quad \mu_{x_q} = N^{-1} \sum_{s=1}^N x_{qs},$$

$$\widehat{\mathbf{a}} = \begin{pmatrix} 1 \\ -\gamma \end{pmatrix} \otimes \{D\mathbf{1}_n - (1 - k^*) \mathbf{C}_I\} + \bar{c}_{II} \begin{pmatrix} 0 \\ -\gamma \end{pmatrix} \otimes \mathbf{1}_n,$$

com

$$\bullet \gamma = \sigma_{xx}^{-1} \sigma_{yx}, \Sigma = \begin{pmatrix} \sigma_y^2 & \sigma'_{yx} \\ \sigma_{yx} & \sigma_{xx} \end{pmatrix},$$

$$\sigma_{yx} = \left(\sigma_{yx_1} \quad \dots \quad \sigma_{yx_Q} \right)', \sigma_{xx} = \begin{pmatrix} \sigma_{x_1}^2 & \dots & \sigma_{x_1 x_Q} \\ \vdots & \ddots & \vdots \\ \sigma_{x_1 x_Q} & \dots & \sigma_{x_Q}^2 \end{pmatrix},$$

$$\sigma_{yx_q} = (N-1)^{-1} \sum_{s=1}^N (y_s - \mu_y)(x_{qs} - \mu_{x_q}),$$

$$\sigma_{x_q x_{q'}} = (N-1)^{-1} \sum_{s=1}^N (x_{qs} - \mu_{x_q})(x_{q's} - \mu_{x_{q'}}).$$

- $\rho_{yx}^{(Q)} = \sqrt{\sigma'_{yx} \sigma_{xx}^{-1} \sigma_{yx} / \sigma_y^2}$ é o coeficiente de correlação múltipla entre a resposta \mathbf{y} (2.2) e as variáveis auxiliares $\mathbf{x}_1 \dots \mathbf{x}_Q$ [ver Morrison (1983), por exemplo] com $\mathbf{x}_q = (x_{q1} \dots x_{qs} \dots x_{qN})'$. Assumimos $|\rho_{yx}^{(Q)}| < 1$.
- \bar{c}_{II} , D e k^* estão definidos em (2.18) com a substituição de ρ_{yx} por $\rho_{yx}^{(Q)}$.

Finalmente, $\text{Var}_{SR}(\hat{T}^* - T)$ é dada em (2.19) com a substituição de ρ_{yx} por $\rho_{yx}^{(Q)}$.

Na mesma direção, podemos obter extensões de (2.20) e (2.21).

2.5 Discussão

No Capítulo 2 mostramos como os resultados de Li, Stanek & Singer (2008) e Stanek, Lencina, Singer, González, Li & San Martino (2008) podem ser estendidos para acomodar erros de medida e incorporar informação auxiliar no modelo misto para populações finitas. Em particular, Li, Stanek & Singer (2008) desenvolveram estimadores da média populacional utilizando informação auxiliar, porém não consideraram erros de medida. Stanek, Lencina, Singer, González, Li & San Martino (2008) consideraram erros de medida tanto exógenos quanto endógenos, contudo não incorporam informação auxiliar. Por outro lado, Stanek & Singer (2004), utilizando um modelo misto para populações finitas para amostragem em dois estágios, desenvolveram preditores com erro de medida endógeno mas não inseriram informação auxiliar. As propostas (2.18) e (2.20) podem ser empregadas tanto para encontrar estimadores de uma média populacional ou de um total populacional, quanto para prever o valor latente da unidade na j -ésima posição numa permutação. Os estima-

dores/preditores so obtidos via mtodos padro para encontrar preditores lineares no viciados de varincia mnima.

No caso especfico, em que o interesse recai na mdia populacional μ_y , o estimador proposto (2.28)  igual ao estimador Horvitz-Thompson (estimador- π) se no tivermos nem erro de medida nem variveis auxiliares [Srndal, Swensson & Wretman (1992, p. 68)]; quando no tivermos erro de medida porm temos uma varivel auxiliar, o estimador (2.27)  semelhante ao estimador regresso generalizado [Cassel et al. (1976) e Wu & Sitter (2001)]; por outro lado, se no tivermos variveis auxiliares mas tivermos erros de medida endgeno, o estimador (2.25)  idntico ao estimador para o modelo de localizao com erro nas variveis “the simple location error-in-variables model” sob o enfoque de super-populao [Bolfarine & Zacks (1992, p. 151)], ou quele dado por Srndal, Swensson & Wretman (1992, p. 605) sob o que eles denominam modelo de medidas simples “the simple measurement model”, e, se no tivermos variveis auxiliares mas tivermos erros de medida exgenos, o estimador (2.26)  idntico ao estimador obtido sob o modelo misto [Henderson, Kempthorne, Searle & von Krosigk (1959)].

Os resultados foram usados, tambm, para encontrar estimadores da medida de associao ponderada, B definida em (1.1). O desempenho dos estimadores \hat{B}^* (2.39) e \tilde{B} (2.40), e de suas verses empricas, \hat{B}_e^* (2.41) e \tilde{B}_e (2.42), considerando ou no o uso da informao auxiliar, foi examinado por meio de exemplos e simulaes, e comparado com aquele do estimador obtido via mnimos quadrados ordinrios, \hat{B}_{MQO} , porque este  o estimador sugerido tanto sob o enfoque baseado no planejamento, no qual o estimador- π de B  igual a \hat{B}_{MQO} [ver Srndal, Swensson & Wretman (1992, p. 196)], quanto sob o enfoque de super-populao [ver Bolfarine & Zacks (1992)]. Dado que os pesos w_s j envolvem x_s , nos questionamos a respeito do efeito da varivel auxiliar sobre os estimadores ajustados \hat{B}^* (2.39) e \tilde{B} (2.40). Para isto, geramos 200 populaes de tamanho 2000, considerando diferentes valores do coeficiente de correlao entre a varivel auxiliar e os valores latentes no ponderados, ρ_{yx}^\bullet , de:

- i) uma distribuio normal bivariada com $\mathbb{E}(Y) = \mathbb{E}(X) = 0$, $Var(Y) = Var(X) = 1$ e $Cov(X, Y) = \rho_{yx}^\bullet$,
- ii) uma populao finita bivariada assimtrica em x , usamos inicialmente uma distribuio normal bivariada com $\mathbb{E}(Y) = \mathbb{E}(X) = 0$, $Var(Y) = Var(X) = 1$ e $Cov(X, Y) = \rho_{yx}^\bullet$, e seguidamente, substitumos os valores de x_s pela transformao $x_{s0} = (x/\sigma_x^2) e^{(x/\sigma_x^2)}$, i.e., usamos x_{s0} como a varivel auxiliar, x_s , e
- iii) uma distribuio normal bivariada com $\mathbb{E}(Y) = \mathbb{E}(X) = 0$, $Var(Y) = Var(X) = 1$ e $Cov(X, Y) = \rho_{yx}^\bullet$, seguidamente, substitumos os valores de x_s pela transformao $x_{s0} = x_s^2$ e

os valores de y_s pela transformação $y_{s0} = y_s^2$. Isto para considerar populações finitas bivariadas assimétricas em x e y .

Nas Figuras 2.2 (a), 2.3 (a) e 2.4 (a) apresentamos os diagramas de dispersão para x vs y^\bullet de uma das populações geradas sob os ítems i), ii) e iii), respectivamente. Nas Figuras 2.2 (b), 2.3 (b) e 2.4 (b) apresentamos os diagramas de dispersão para ρ_{yx}^\bullet vs ρ_{yx} (usando $w_s = (x_s - \mu_x) / [(N - 1) \sigma_x^2]$, $s = 1, \dots, N$) das 200 populações geradas sob os ítems i), ii) e iii), respectivamente. Semelhantemente, nas Figuras 2.2 (c), 2.3 (c) e 2.4 (c) [respectivamente, nas Figuras 2.2 (d), 2.3 (d) e 2.4 (d)], os diagramas de dispersão de ρ_{yx}^\bullet vs $Var_s(\hat{B}^*)$ [respectivamente, ρ_{yx}^\bullet vs $Var_s(\tilde{B}^*)$] são apresentados considerando nem EMEn nem EMEx e $n = 10$, i.e., $f = 10/2000$. Dado que a variância dos estimadores de B , \hat{B}^* (2.39) e \tilde{B} (2.40), diminuem conforme $|\rho_{yx}|$ aumenta, deduzimos da Figura 2.2 (b), sob populações geradas conforme o ítem i), que a variável auxiliar obtém pouca ou nenhuma redução nas variâncias de \hat{B}^* e \tilde{B} porque $|\rho_{yx}| < 0.10$, que é observado nas Figuras 2.2 (c) versus 2.2 (d). Da Figura 2.3 (b) [respectivamente, Figura 2.4 (b)], deduzimos que a variância de \hat{B}^* e \tilde{B} diminuem quando $|\rho_{yx}|$ aumenta sob populações geradas conforme o ítem ii) [respectivamente, o ítem iii)], que é observado nas Figuras 2.3 (c) versus Figura 2.3 (d) [respectivamente, Figuras 2.4 (c) versus Figura 2.4 (d)]. O anterior sugere, sob alguns casos, que a variável auxiliar pode ser útil para reduzir a variância dos preditores \hat{B}^* e \tilde{B} .

Os exemplos e as simulações comprovam a otimalidade de \hat{B}^* e \tilde{B} na classe dos estimadores não viciados; de fato, quando consideramos dados com distribuição assimétrica em x , o EQM(\hat{B}_e^*) (respectivamente, EQM(\tilde{B}_e)) é menor que o correspondente EQM(\hat{B}_{MQO}), em especial para n pequeno. Porém, quando consideramos dados com distribuição simétrica, o estimador \hat{B}_e^* (2.41) [respectivamente, \tilde{B}_e (2.42)] apresenta maior EQM que o estimador \hat{B}_{MQO} para $\rho_{yx}^\bullet > 0.4$ (respectivamente, $\rho_{yx}^\bullet > 0.4$ e $n = 10$) (Figuras C.1 (a) [respectivamente, C.5 (a)]), e \hat{B}_e^* apresenta vício.

O vício da \hat{B}_{MQO} , especificamente para valores pequenos de n , é justificado uma vez que “uma amostra equilibrada”, na qual $\mu_x = \bar{X}_I = \bar{X}_{II}$ com $\bar{X}_{II} = (N - n)^{-1} \sum_{i=n+1}^N X_i$, não é esperada quando selecionamos amostras de populações finitas [ver Cumberland & Royall (1988) e Bolfarine & Zacks (1992)].

Finalmente, sugerimos usar os estimadores propostos \hat{B}^* e \tilde{B} , ou, suas versões empíricas, \hat{B}_e^* e \tilde{B}_e , segundo seja o caso, quando se tem dados com distribuição assimétrica em x ou amostras pequenas.

Figura 2.2: Diagrama de dispersão de uma variável auxiliar (x) vs os valores latentes (y^\bullet) para uma população com $\rho_{yx}^\bullet = 0.5$ e tamanho 2000 e diagramas de dispersão de ρ_{yx}^\bullet vs ρ_{yx} , ρ_{yx}^\bullet vs $Var_s(\hat{B}^*)$ e ρ_{yx}^\bullet vs $Var_s(\bar{B}^*)$ para 200 populações geradas em que $w_s = (x_s - \mu_x) / [(N - 1) \sigma_x^2]$. Dados simétricos.

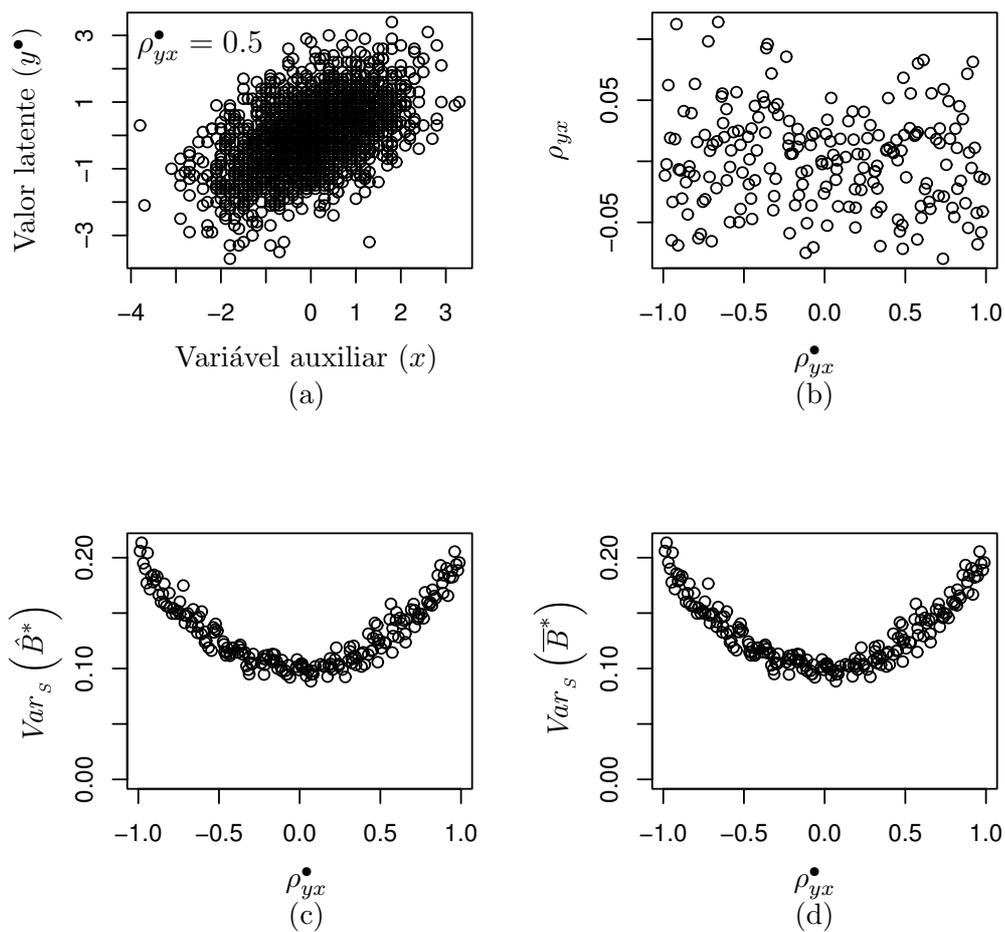


Figura 2.3: Diagrama de dispersão de uma variável auxiliar (x) vs os valores latentes (y^\bullet) para uma população com $\rho_{yx}^\bullet = 0.5$ e tamanho 2000, e diagramas de dispersão de ρ_{yx}^\bullet vs ρ_{yx} , ρ_{yx}^\bullet vs $Var_s(\hat{B}^*)$ e ρ_{yx}^\bullet vs $Var_s(\bar{B}^*)$ para 200 populações geradas em que $w_s = (x_s - \mu_x) / [(N - 1) \sigma_x^2]$. Dados assimétricos in x .

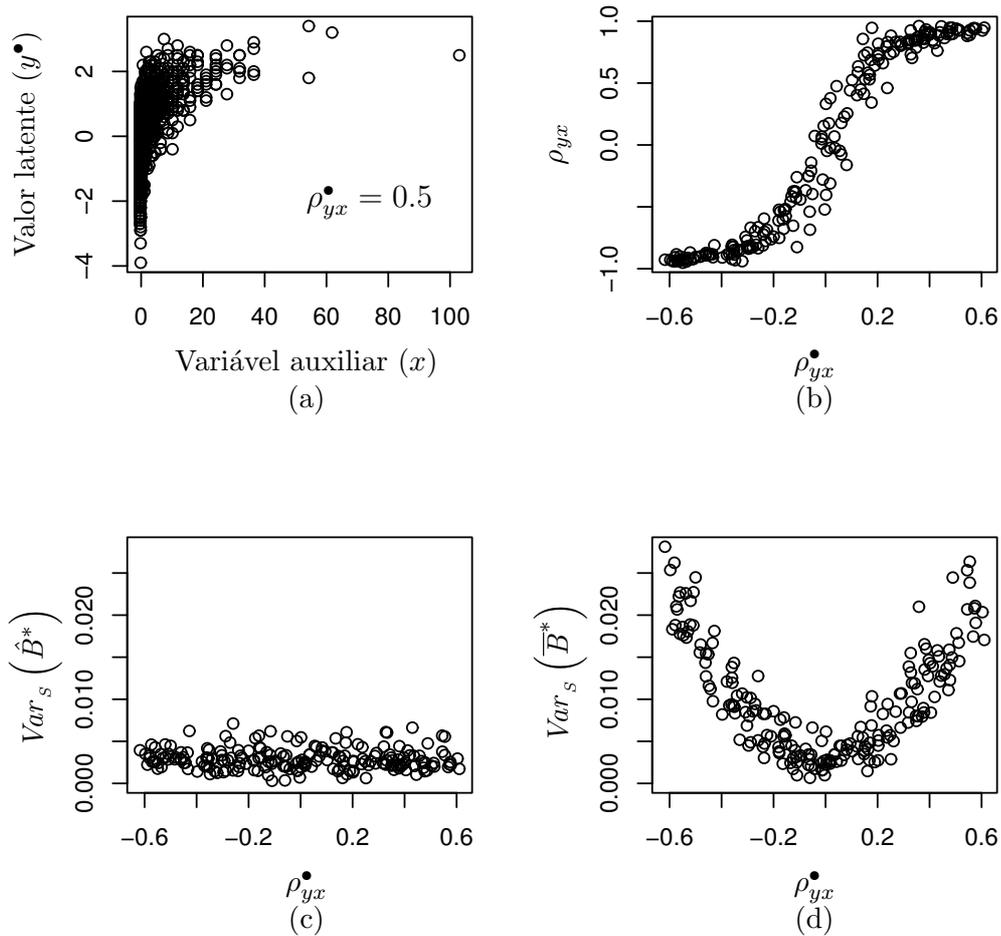
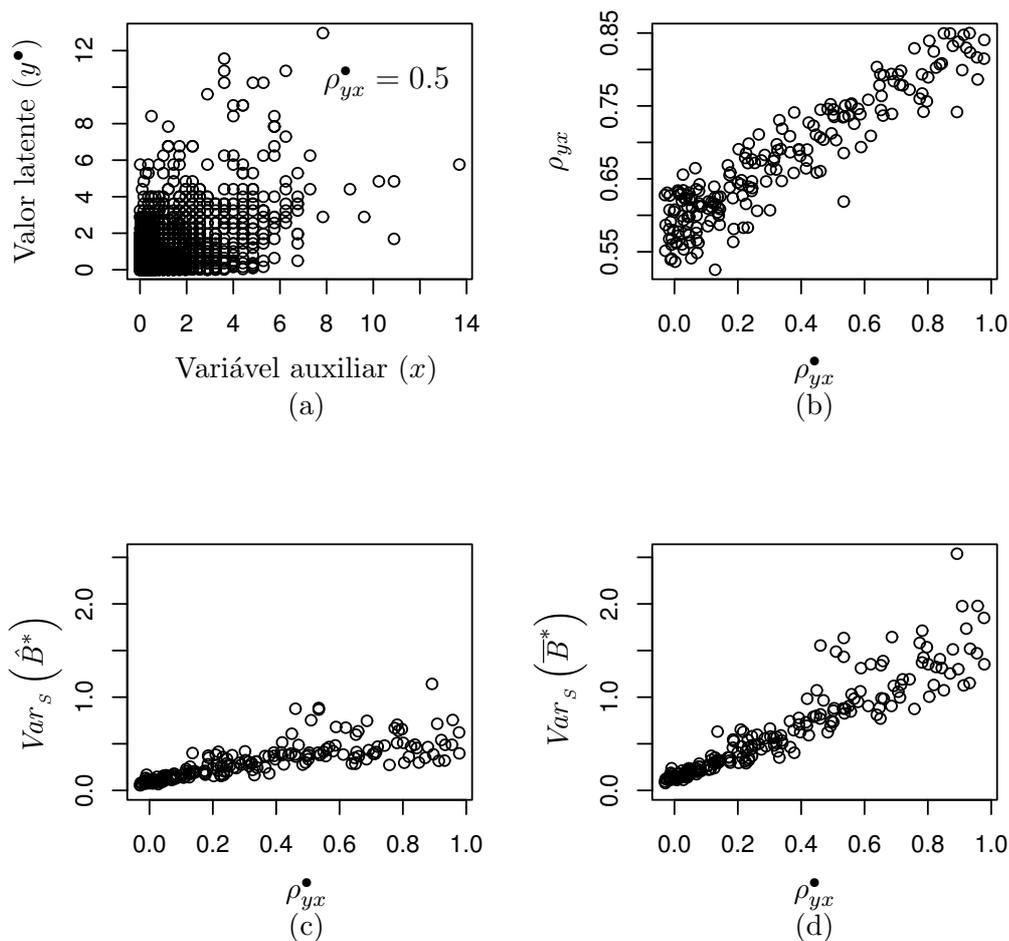


Figura 2.4: Diagrama de dispersão de uma variável auxiliar (x) vs os valores latentes (y^\bullet) para uma população com $\rho_{yx}^\bullet = 0.5$ e tamanho 2000, e diagramas de dispersão de ρ_{yx}^\bullet vs ρ_{yx} , ρ_{yx}^\bullet vs $Var_s(\hat{B}^*)$ e ρ_{yx}^\bullet vs $Var_s(\bar{B}^*)$ para 200 populações geradas em que $w_s = (x_s - \mu_x) / [(N - 1) \sigma_x^2]$. Dados assimétricos em x e y^\bullet .



Capítulo 3

Estimação e previsão para experimentos pré-teste/pós-teste com populações finitas

Enfoques baseados em modelos para análise de experimentos pré-teste/pós-teste como os discutidos na Seção 1.2, i.e., modelos lineares para medidas repetidas e modelos de análise de covariância, têm várias suposições difíceis de corroborar na prática. O primeiro enfoque assume que variâncias e covariâncias correspondentes aos indivíduos dos dois grupos (seja sob o cenário de dois estratos com uma intervenção, ou, sob o cenário de uma população não estratificada com duas intervenções) são iguais. O segundo enfoque assume respostas no pré-teste não aleatórias, observadas sem erro e com médias iguais para os dois grupos. Além disso, os modelos lineares para medidas repetidas e os modelos de análise de covariância não dependem do planejamento e podem ser usados tanto sob um planejamento que envolve a aleatorização dos tratamentos às unidades amostrais quanto em estudos observacionais, que dificulta a interpretação dos resultados dada a falta de ligação entre os parâmetros do modelo estocástico e os parâmetros populacionais.

Neste Capítulo, sob os dois cenários referidos nos Exemplos 1.1 e 1.2, propomos um modelo misto para populações finitas, levando em conta possíveis erros de medida^a tanto no pré-teste quanto no pós-teste. Esse modelo é construído diretamente das variáveis aleatórias que representam o processo de amostragem, facilitando a definição dos parâmetros de interesse e a interpretação dos resultados.

^aPara comentários sobre os erros de medida a ser considerados, ver Seção 2.1.

3.1 Modelo misto para estudos pr-teste/ps-teste em populaes finitas com dois estratos

Assumimos que a cada unidade, rotulada por $s = 1, \dots, N_h$, do estrato $h = 1, 2$, que compe a populao finita, esto associados os valores fixos y_{hs1}^\bullet e y_{hs2}^\bullet , respectivamente, a resposta antes (pr-teste) da interveno e a resposta depois (ps-teste) da interveno. Tambm, assumimos que ao selecionar a unidade s no estrato h observamos os dois valores y_{hs1}^\bullet e y_{hs2}^\bullet possivelmente com erro de medida. Como em algumas situaes  conveniente trabalhar com os valores latentes ponderados, definimos $y_{hs1} = w_{hs1}y_{hs1}^\bullet$ e $y_{hs2} = w_{hs2}y_{hs2}^\bullet$, em que w_{hs1} e w_{hs2} , $h = 1, 2$, $s = 1, \dots, N_h$, representam pesos.

O processo da amostragem aleatria simples em cada estrato  formalizado por meio de um modelo para populaes finitas. Sob este modelo, qualquer uma das $N_h!$ possveis permutaes de unidades do estrato h pode ser selecionada com a mesma probabilidade $[(N_h!)^{-1}]$, $h = 1, 2$. Neste contexto, um novo rtulo, $i = 1, \dots, N_h$,  associado a cada unidade no estrato h segundo sua posio em cada permutao. Logo, podemos representar uma permutao aleatria das respostas pr-teste e ps-teste no estrato h (dentre as $N_h!$ possveis) por meio da matriz aleatria $(\mathbf{Y}_1^{(h)} \quad \mathbf{Y}_2^{(h)})$ em que

$$\mathbf{Y}_1^{(h)} = (Y_{h11} \dots Y_{hi1} \dots Y_{hN_h1})' \quad \text{e} \quad \mathbf{Y}_2^{(h)} = (Y_{h12} \dots Y_{hi2} \dots Y_{hN_h2})', \quad h = 1, 2.$$

Para associar os valores

$$\mathbf{y}_1^{(h)} = (y_{h11} \dots y_{hs1} \dots y_{hN_h1})' \quad \text{e} \quad \mathbf{y}_2^{(h)} = (y_{h12} \dots y_{hs2} \dots y_{hN_h2})', \quad h = 1, 2,$$

a $\mathbf{Y}_1^{(h)}$ e $\mathbf{Y}_2^{(h)}$, $h = 1, 2$, consideramos

$$\begin{pmatrix} \mathbf{Y}_1^{(1)} & \mathbf{Y}_2^{(1)} \\ \mathbf{Y}_1^{(2)} & \mathbf{Y}_2^{(2)} \end{pmatrix} = \text{diag} \{ \mathbf{U}^{(1)}, \mathbf{U}^{(2)} \} \begin{pmatrix} \mathbf{y}_1^{(1)} & \mathbf{y}_2^{(1)} \\ \mathbf{y}_1^{(2)} & \mathbf{y}_2^{(2)} \end{pmatrix} \quad (3.1)$$

em que $\mathbf{U}^{(h)} = \{U_{is}^{(h)}\} \in R^{N_h \times N_h}$, $h = 1, 2$, e $U_{is}^{(h)}$  uma varivel indicadora com valor 1 se a unidade s dentro do estrato h  selecionada na posio i dentro da permutao e 0 em caso contrrio.

Incorporamos erros de medida endgenos (EMEn) tanto no pr-teste quanto no ps-teste em (3.1) adicionando a matriz aleatria

$$\begin{pmatrix} \mathbf{E}_1^{(1)\bullet} & \mathbf{E}_2^{(1)\bullet} \\ \mathbf{E}_1^{(2)\bullet} & \mathbf{E}_2^{(2)\bullet} \end{pmatrix},$$

em que $\mathbf{E}_t^{(h)\bullet} = (E_{h1t}^\bullet \dots E_{hst}^\bullet \dots E_{hN_h t}^\bullet)'$ e com $\mathbb{E}_R(E_{hst}^\bullet) = 0$, $\text{Var}_R(E_{hst}^\bullet) = \sigma_{hst}^{\bullet 2}$ e E_{hst}^\bullet , $h = 1, 2, t = 1, 2, s = 1, \dots, N_h$, independentes duas a duas, à matriz

$$\begin{pmatrix} \mathbf{y}_1^{(1)\bullet} & \mathbf{y}_2^{(1)\bullet} \\ \mathbf{y}_1^{(2)\bullet} & \mathbf{y}_2^{(2)\bullet} \end{pmatrix},$$

em que $\mathbf{y}_t^{(h)\bullet} = (y_{h1t}^\bullet \dots y_{hst}^\bullet \dots y_{hN_h t}^\bullet)'$. O modelo pode, então, ser representado como

$$\begin{pmatrix} \mathbf{Y}_1^{(1)*} & \mathbf{Y}_2^{(1)*} \\ \mathbf{Y}_1^{(2)*} & \mathbf{Y}_2^{(2)*} \end{pmatrix} = \text{diag} \{ \mathbf{U}^{(1)}, \mathbf{U}^{(2)} \} \left[\begin{pmatrix} \mathbf{y}_1^{(1)} & \mathbf{y}_2^{(1)} \\ \mathbf{y}_1^{(2)} & \mathbf{y}_2^{(2)} \end{pmatrix} + \begin{pmatrix} \mathbf{E}_1^{(1)} & \mathbf{E}_2^{(1)} \\ \mathbf{E}_1^{(2)} & \mathbf{E}_2^{(2)} \end{pmatrix} \right] \quad (3.2)$$

em que $\mathbf{Y}_t^{(h)*} = (Y_{h1t}^* \dots Y_{hit}^* \dots Y_{hN_h t}^*)'$ e $\mathbf{E}_t^{(h)} = (E_{h11} \dots E_{hs1} \dots E_{hN_h 1})'$, com $E_{hst} = w_{hst} E_{hst}^\bullet$, é o vetor aleatório de EMEn no pré-teste (se $t = 1$) ou no pós-teste (se $t = 2$). Note que $\mathbf{y}_t^{(h)} = (w_{h1t} y_{h1t}^\bullet \dots w_{hst} y_{hst}^\bullet \dots w_{hN_h t} y_{hN_h t}^\bullet)'$. Então,

$$\mathbb{E}_R(E_{hst}) = 0, \quad \text{Var}_R(E_{hst}) = \sigma_{hst}^2 = w_{hst}^2 \sigma_{hst}^{\bullet 2}$$

e as variáveis E_{hst} , $h = 1, 2, t = 1, 2, s = 1, \dots, N_h$, são independentes duas a duas. O sub-índice R denota as operações de cálculo da esperança e da variância com respeito à distribuição dos erros de medida. Para a posição i dentro de uma permutação das unidades do estrato h ,

$$Y_{hit}^* = \sum_{s=1}^{N_h} U_{is}^{(h)} (y_{hst} + E_{hst}) = \mu_t^{(h)} + d_{hit} + \sum_{s=1}^{N_h} U_{is}^{(h)} E_{hst}, \quad t = 1, 2, \quad (3.3)$$

em que $d_{hit} = \sum_{s=1}^{N_h} U_{is}^{(h)} \delta_{hst}$, $\delta_{hst} = y_{hst} - \mu_t^{(h)}$ e $\mu_t^{(h)} = N_h^{-1} \sum_{s=1}^{N_h} y_{hst}$, que se pode reparametrizar

como $\mu_t^{(h)} = \mu + \alpha_h + \gamma_t + (\alpha\gamma)_{ht}$ em que $\mu = (2N_1 + 2N_2)^{-1} \left(\sum_{h=1}^2 \sum_{s=1}^{N_h} \sum_{t=1}^2 y_{hst} \right)$, $\alpha_h = \mu - (2N_h)^{-1} \left(\sum_{s=1}^{N_h} \sum_{t=1}^2 y_{hst} \right)$, $\gamma_t = \mu - (N_1 + N_2)^{-1} \left(\sum_{h=1}^2 \sum_{s=1}^{N_h} y_{hst} \right)$ e $(\alpha\gamma)_{ht} = \mu_t^{(h)} - \mu - \alpha_h - \gamma_t$.

O modelo misto para populações finitas com EMEn é

$$\begin{aligned} \mathbf{Z}^* &= \text{vec} \left[\begin{pmatrix} \mathbf{Y}_1^{(1)*} & \mathbf{Y}_2^{(1)*} \\ \mathbf{Y}_1^{(2)*} & \mathbf{Y}_2^{(2)*} \end{pmatrix} \right] \\ &= \left[\mathbf{I}_2 \otimes \begin{pmatrix} \mathbf{1}_{N_1} \\ \mathbf{0} \end{pmatrix} \right] \boldsymbol{\mu}^{(1)} + \left[\mathbf{I}_2 \otimes \begin{pmatrix} \mathbf{0} \\ \mathbf{1}_{N_2} \end{pmatrix} \right] \boldsymbol{\mu}^{(2)} + \mathbf{d} + \mathbf{E}^* \end{aligned} \quad (3.4)$$

em que $\boldsymbol{\mu}^{(h)} = (\mu_1^{(h)} \ \mu_2^{(h)})'$, $h = 1, 2$, corresponde ao vetor de médias populacionais pré-teste, $\mu_1^{(h)}$, e pós-teste, $\mu_2^{(h)}$, no estrato h (efeitos fixos associados ao estrato h),

$$\mathbf{d} = (\mathbf{d}'_{11} \ \mathbf{d}'_{21} \ \mathbf{d}'_{12} \ \mathbf{d}'_{22})' \quad \text{com} \quad \mathbf{d}_{ht} = (d_{h1t} \ \dots \ d_{hit} \ \dots \ d_{hN_h t})'$$

é um vetor de efeitos aleatórios, e

$$\mathbf{E}^* = \left(\left[\mathbf{E}_1^{(1)*} \right]' \ \left[\mathbf{E}_1^{(2)*} \right]' \ \left[\mathbf{E}_2^{(1)*} \right]' \ \left[\mathbf{E}_2^{(2)*} \right]' \right)',$$

com $\mathbf{E}_t^{(h)*} = \mathbf{Y}_t^{(h)*} - \mathbf{Y}_t^{(h)}$.

No Apêndice A (pág. 81), demonstramos que

$$\mathbb{E}_{SR}(\mathbf{Z}^*) = \left[\mathbf{I}_2 \otimes \begin{pmatrix} \mathbf{1}_{N_1} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{1}_{N_2} \end{pmatrix} \right] \boldsymbol{\mu}^{(1)} + \left[\mathbf{I}_2 \otimes \begin{pmatrix} \mathbf{0} \\ \mathbf{1}_{N_1 \times 1} \\ \mathbf{0} \\ \mathbf{1}_{N_2} \end{pmatrix} \right] \boldsymbol{\mu}^{(2)}$$

e

$$\begin{aligned} \text{Var}_{SR}(\mathbf{Z}^*) &= \text{diag} \{ \bar{\sigma}_{11}^2 \mathbf{I}_{N_1}, \bar{\sigma}_{21}^2 \mathbf{I}_{N_2}, \bar{\sigma}_{12}^2 \mathbf{I}_{N_1}, \bar{\sigma}_{22}^2 \mathbf{I}_{N_2} \} \\ &\quad + \left[\boldsymbol{\Sigma}^{(1)} \otimes \text{diag} \left\{ \mathbf{P}_{N_1}, \mathbf{0}, \mathbf{0}, \mathbf{P}_{N_2} \right\} \right] + \left[\boldsymbol{\Sigma}^{(2)} \otimes \text{diag} \left\{ \mathbf{0}, \mathbf{P}_{N_1}, \mathbf{0}, \mathbf{P}_{N_2} \right\} \right], \end{aligned}$$

Aqui, o índice S indica valor esperado com respeito às permutações das unidades no estrato h , $h = 1, 2$, \mathbf{P}_a está definido em (2.5),

$$\bar{\sigma}_{ht}^2 = \mathbb{E}_S \left(\sum_{s=1}^{N_h} \sigma_{hst}^2 U_{is}^{(h)} \right) = N_h^{-1} \sum_{s=1}^{N_h} \sigma_{hst}^2 \quad (3.5)$$

e

$$\boldsymbol{\Sigma}^{(h)} = \begin{pmatrix} \sigma_{11}^{(h)} & \sigma_{12}^{(h)} \\ \sigma_{12}^{(h)} & \sigma_{22}^{(h)} \end{pmatrix}, \quad (3.6)$$

com

$$\sigma_{tt'}^{(h)} = (N_h - 1)^{-1} \sum_{s=1}^{N_h} \left(y_{hst} - \mu_t^{(h)} \right) \left(y_{hst'} - \mu_{t'}^{(h)} \right), \quad t, t' = 1, 2.$$

Quando os erros de medida no pré-teste e no pós-teste são exógenos (EMEx), eles são incorporados

em (3.1) por meio da adição da matriz aleatória

$$\begin{pmatrix} \tilde{\mathbf{E}}_1^{(1)\bullet} & \tilde{\mathbf{E}}_2^{(1)\bullet} \\ \tilde{\mathbf{E}}_1^{(2)\bullet} & \tilde{\mathbf{E}}_2^{(2)\bullet} \end{pmatrix},$$

em que $\tilde{\mathbf{E}}_t^{(h)\bullet} = \left(\tilde{E}_{h1t}^\bullet \dots \tilde{E}_{hit}^\bullet \dots \tilde{E}_{hN_h t}^\bullet \right)'$ com $\mathbb{E}_R \left(\tilde{E}_{hit}^\bullet \right) = 0$, $\text{Var}_R \left(\tilde{E}_{hit}^\bullet \right) = \tilde{\sigma}_{hit}^{\bullet 2}$ e \tilde{E}_{hit}^\bullet , $h = 1, 2$, $t = 1, 2$, $i = 1, \dots, N_h$, independentes duas a duas, à matriz

$$\begin{pmatrix} \mathbf{Y}_1^{(1)\bullet} & \mathbf{Y}_2^{(1)\bullet} \\ \mathbf{Y}_1^{(2)\bullet} & \mathbf{Y}_2^{(2)\bullet} \end{pmatrix},$$

em que $\mathbf{Y}_t^{(h)\bullet} = \mathbf{U}^{(h)} \mathbf{y}_t^{(h)\bullet}$. O modelo pode ser escrito como

$$\begin{pmatrix} \tilde{\mathbf{Y}}_1^{(1)} & \tilde{\mathbf{Y}}_2^{(1)} \\ \tilde{\mathbf{Y}}_1^{(2)} & \tilde{\mathbf{Y}}_2^{(2)} \end{pmatrix} = \text{diag} \left\{ \mathbf{U}^{(1)}, \mathbf{U}^{(2)} \right\} \begin{pmatrix} \mathbf{y}_1^{(1)} & \mathbf{y}_2^{(1)} \\ \mathbf{y}_1^{(2)} & \mathbf{y}_2^{(2)} \end{pmatrix} + \begin{pmatrix} \tilde{\mathbf{E}}_1^{(1)} & \tilde{\mathbf{E}}_2^{(1)} \\ \tilde{\mathbf{E}}_1^{(2)} & \tilde{\mathbf{E}}_2^{(2)} \end{pmatrix}, \quad (3.7)$$

em que $\tilde{\mathbf{Y}}_t^{(h)} = \left(\tilde{Y}_{h1t} \dots \tilde{Y}_{hit} \dots \tilde{Y}_{hN_h t} \right)'$ e $\tilde{\mathbf{E}}_t^{(h)} = \left(\tilde{E}_{h1t} \dots \tilde{E}_{hit} \dots \tilde{E}_{hN_h t} \right)'$, com $\tilde{E}_{hit} = \tilde{E}_{hit}^\bullet \sum_{s=1}^{N_h} \left(U_{is}^{(h)} w_{hst} \right)$, é um vetor aleatório. Logo $\mathbb{E}_R \left(\tilde{E}_{hit} \right) = 0$, $\text{Var}_{SR} \left(\tilde{E}_{hit} \right) = \tilde{\sigma}_{hit}^2 = \mu_{wh} \tilde{\sigma}_{hit}^{\bullet 2}$ e as variáveis \tilde{E}_{hit} , $h = 1, 2$, $t = 1, 2$, $i = 1, \dots, N_h$, são independentes duas a duas, com $\mu_{wh} = N_h^{-1} \sum_{s=1}^{N_h} w_{hst}^2$, $h = 1, 2$, $t = 1, 2$. Dessa forma, para a posição i dentro de uma permutação das unidades no estrato h ,

$$\tilde{Y}_{hit} = \mu_t^{(h)} + d_{hit} + \tilde{E}_{hit}, \quad t = 1, 2,$$

com $\mu_t^{(h)}$ e d_{hit} definidos em (3.3).

O correspondente modelo misto para populações finitas com EMEx é

$$\begin{aligned} \tilde{\mathbf{Z}} &= \text{vec} \left[\begin{pmatrix} \tilde{\mathbf{Y}}_1^{(1)} & \tilde{\mathbf{Y}}_2^{(1)} \\ \tilde{\mathbf{Y}}_1^{(2)} & \tilde{\mathbf{Y}}_2^{(2)} \end{pmatrix} \right] \\ &= \left[\mathbf{I}_2 \otimes \begin{pmatrix} \mathbf{1}_{N_1} \\ \mathbf{0} \\ \mathbf{1}_{N_2} \end{pmatrix} \right] \boldsymbol{\mu}^{(1)} + \left[\mathbf{I}_2 \otimes \begin{pmatrix} \mathbf{0} \\ \mathbf{1}_{N_1 \times 1} \\ \mathbf{1}_{N_2} \end{pmatrix} \right] \boldsymbol{\mu}^{(2)} + \mathbf{d} + \tilde{\mathbf{E}} \end{aligned} \quad (3.8)$$

em que $\boldsymbol{\mu}^{(1)}$, $\boldsymbol{\mu}^{(2)}$ e \mathbf{d} estão definidos em (3.4) e

$$\tilde{\mathbf{E}} = \left(\left[\tilde{\mathbf{E}}_1^{(1)} \right]' \left[\tilde{\mathbf{E}}_1^{(2)} \right]' \left[\tilde{\mathbf{E}}_2^{(1)} \right]' \left[\tilde{\mathbf{E}}_2^{(2)} \right]' \right)',$$

representa o vetor de EMEx.

O vetor de médias e a matriz de covariâncias de $\tilde{\mathbf{Z}}$ são dados por (ver Apêndice A, pág. 82)

$$\mathbb{E}_{SR}(\tilde{\mathbf{Z}}) = \begin{bmatrix} \mathbf{I}_2 \otimes \begin{pmatrix} \mathbf{1}_{N_1} \\ \mathbf{0} \\ \mathbf{1}_{N_2 \times 1} \end{pmatrix} \end{bmatrix} \boldsymbol{\mu}^{(1)} + \begin{bmatrix} \mathbf{I}_2 \otimes \begin{pmatrix} \mathbf{0} \\ \mathbf{1}_{N_1 \times 1} \\ \mathbf{1}_{N_2} \end{pmatrix} \end{bmatrix} \boldsymbol{\mu}^{(2)},$$

e

$$\begin{aligned} \text{Var}_{SR}(\tilde{\mathbf{Z}}) = & \text{diag} \left\{ \left(\bigoplus_{i=1}^{N_1} \tilde{\sigma}_{1i1}^2 \right), \left(\bigoplus_{i=1}^{N_2} \tilde{\sigma}_{2i1}^2 \right), \left(\bigoplus_{i=1}^{N_1} \tilde{\sigma}_{1i2}^2 \right), \left(\bigoplus_{i=1}^{N_2} \tilde{\sigma}_{2i2}^2 \right) \right\} \\ & + \left[\boldsymbol{\Sigma}^{(1)} \otimes \text{diag} \left\{ \mathbf{P}_{N_1}, \mathbf{0}_{N_2 \times N_2} \right\} \right] + \left[\boldsymbol{\Sigma}^{(2)} \otimes \text{diag} \left\{ \mathbf{0}_{N_1 \times N_1}, \mathbf{P}_{N_2} \right\} \right], \end{aligned}$$

com $\boldsymbol{\Sigma}^{(h)}$ definida em (3.6).

3.1.1 Combinações lineares de interesse e preditores ótimos

Sob os modelos (3.4) e (3.8), nosso interesse é predizer/estimar combinações lineares da forma

$$T = \sum_{i=1}^{N_1} (c_{1i1} Y_{1i1} + c_{1i2} Y_{1i2}) + \sum_{i=1}^{N_2} (c_{2i1} Y_{2i1} + c_{2i2} Y_{2i2}) \quad (3.9)$$

em que os c_{hi1} e c_{hi2} são constantes conhecidas associadas aos rótulos i (posições) numa permutação no estrato h . Por exemplo, se o interesse recai na diferença da resposta média no pré-teste entre os dois estratos, nomeadamente,

$$T_o = N_1^{-1} \sum_{i=1}^{N_1} Y_{1i1} - N_2^{-1} \sum_{i=1}^{N_2} Y_{2i1}, \quad (3.10)$$

tomamos $c_{1i1} = N_1^{-1}$, $i = 1, \dots, N_1$, $c_{2i1} = -N_2^{-1}$, $i = 1, \dots, N_2$, e $c_{hi2} = 0$. Se o interesse recai no efeito da intervenção no estrato h , ou seja,

$$T_h = N_h^{-1} \sum_{i=1}^{N_h} (Y_{hi2} - Y_{hi1}), \quad (3.11)$$

tomamos $c_{h'i2} = N_{h'}^{-1}$ e $c_{h'i1} = -N_{h'}^{-1}$, $i = 1, \dots, N_{h'}$, se $h' = h$, e $c_{h'it} = 0$ se $h' \neq h$. Desta forma, para avaliar a diferença dos efeitos da intervenção entre os dois estratos, i.e.,

$$T_d = N_1^{-1} \sum_{i=1}^{N_1} (Y_{1i2} - Y_{1i1}) - N_2^{-1} \sum_{i=1}^{N_2} (Y_{2i2} - Y_{2i1}), \quad (3.12)$$

adotamos $c_{1i1} = -N_1^{-1}$, $c_{1i2} = N_1^{-1}$, $c_{2i1} = N_2^{-1}$ e $c_{2i2} = -N_2^{-1}$ para todo i .

Para obter o estimador/preditor de T , com base numa amostra aleatória simples selecionada sem reposição dentro de cada estrato, assumimos, sem perda de generalidade, que no estrato h as primeiras n_h , $h = 1, 2$, posições na permutação aleatória correspondem às unidades selecionadas na amostra. Desta maneira, T em (3.9) pode ser decomposto como

$$T = T^{(1)} + T^{(2)} \quad (3.13)$$

com

$$T^{(1)} = \sum_{i=1}^{n_1} (c_{1i1}Y_{1i1} + c_{1i2}Y_{1i2}) + \sum_{i=1}^{n_2} (c_{2i1}Y_{2i1} + c_{2i2}Y_{2i2})$$

e

$$T^{(2)} = \sum_{i=n_1+1}^{N_1} (c_{1i1}Y_{1i1} + c_{1i2}Y_{1i2}) + \sum_{i=n_2+1}^{N_2} (c_{2i1}Y_{2i1} + c_{2i2}Y_{2i2}).$$

Como a partir dos dados amostrais observamos diretamente (sem erro de medida) ou indiretamente (com erros de medida) apenas $T^{(1)}$, temos que prever o resto, i.e., $T^{(2)}$.

O estimador/preditor de T (3.9) obtém-se utilizando métodos padrão para encontrar preditores lineares não viciados de variância mínima (segundo o paradigma apresentado na Seção 2.2). Em virtude da diferença entre o preditor sob os dois modelos (3.4) e (3.8) ser devida à variância dos EMEn e EMEx, respectivamente, vamos considerar apenas a predição sob o modelo (3.4) e posteriormente vamos particularizar os resultados para o modelo (3.8).

Como na Seção 2.2, decompomos Z^* apresentado em (3.4) em duas partes, uma que corresponde às amostras dos dois estratos e outra à parte restante (não amostrada); com essa finalidade pré-multiplicamos Z^* por

$$K = \begin{pmatrix} K_I \\ K_{II} \end{pmatrix} \quad (3.14)$$

em que

$$\mathbf{K}_I = \mathbf{I}_2 \otimes \begin{pmatrix} \mathbf{I}_{n_1} & \mathbf{0}_{n_1 \times (N_1 - n_1)} & \mathbf{0}_{n_1 \times N_2} \\ \mathbf{0}_{n_2 \times N_1} & \mathbf{I}_{n_2} & \mathbf{0}_{n_2 \times (N_2 - n_2)} \end{pmatrix}$$

e

$$\mathbf{K}_{II} = \mathbf{I}_2 \otimes \begin{pmatrix} \mathbf{0}_{(N_1 - n_1) \times n_1} & \mathbf{I}_{N_1 - n_1} & \mathbf{0}_{(N_1 - n_1) \times N_2} \\ \mathbf{0}_{(N_2 - n_2) \times N_1} & \mathbf{0}_{(N_2 - n_2) \times n_2} & \mathbf{I}_{N_2 - n_2} \end{pmatrix}.$$

Denotando

$$\begin{pmatrix} \mathbf{Z}_I^* \\ \mathbf{Z}_{II}^* \end{pmatrix} = \begin{pmatrix} \mathbf{K}_I \mathbf{Z}^* \\ \mathbf{K}_{II} \mathbf{Z}^* \end{pmatrix},$$

segue que

$$\mathbf{Z}_I^* = (Y_{111}^* \ \dots \ Y_{1n_1}^* \ Y_{211}^* \ \dots \ Y_{2n_2}^* \ Y_{112}^* \ \dots \ Y_{1n_1}^* \ Y_{212}^* \ \dots \ Y_{2n_2}^*)'$$

e

$$\mathbf{Z}_{II}^* = (Y_{1,(n_1+1),1}^* \ \dots \ Y_{1N_1}^* \ Y_{2,(n_2+1),1}^* \ \dots \ Y_{2N_2}^* \ Y_{1,(n_1+1),2}^* \ \dots \ Y_{1N_1}^* \ Y_{2,(n_2+1),2}^* \ \dots \ Y_{2N_2}^*)'. \quad (3.15)$$

Devido à complexidade das estruturas das matrizes de covariâncias de \mathbf{Z}_I^* e \mathbf{Z}_{II}^* , e de seu uso na procura do preditor linear não viciado de variância mínima, neste trabalho assumimos que tanto o tamanho dos dois estratos quanto o tamanho das amostras dentro de cada estrato é igual, i.e., $N_1 = N_2 = N$ e $n_1 = n_2 = n$, além disso, assumimos que

$$\begin{aligned} \sigma_{hs1}^2 &= \sigma_{hs2}^2, \quad h = 1, 2, s = 1, \dots, N, \quad \text{sob o modelo (3.4), e} \\ \tilde{\sigma}_{hi1}^2 &= \tilde{\sigma}_{hi2}^2, \quad h = 1, 2, i = 1, \dots, N, \quad \text{sob o modelo (3.8).} \end{aligned} \quad (3.16)$$

Para considerar o problema de tamanho diferentes dos dois estratos e das duas amostras, pode-se usar um modelo com variáveis expandidas para controlar a dimensão dos vetores e matrizes amostradas [ver Stanek & Singer (2008)].

Logo, $\mathbf{K}_I = \mathbf{I}_4 \otimes (\mathbf{I}_n \mathbf{0}_{n \times (N-n)})$, $\mathbf{K}_{II} = \mathbf{I}_4 \otimes (\mathbf{0}_{(N-n) \times n} \mathbf{I}_{N-n})$, e conseqüentemente,

$$\begin{aligned}\mathbb{E}_{SR}(\mathbf{Z}_I^*) &= \boldsymbol{\mu} \otimes \mathbf{1}_n, \\ \mathbb{E}_{SR}(\mathbf{Z}_{II}^*) &= \boldsymbol{\mu} \otimes \mathbf{1}_{N-n}, \\ \text{Var}_{SR}(\mathbf{Z}_I^*) &= \mathbf{V}_I^* = \boldsymbol{\Sigma} \otimes \mathbf{P}_{n,N} + \\ &\quad (\mathbf{I}_2 \otimes \mathbf{e}_1 \mathbf{e}_1' \otimes \bar{\sigma}_{11}^2 \mathbf{I}_n) + (\mathbf{I}_2 \otimes \mathbf{e}_2 \mathbf{e}_2' \otimes \bar{\sigma}_{21}^2 \mathbf{I}_n), \\ \text{Var}_{SR}(\mathbf{Z}_{II}^*) &= \mathbf{V}_{II}^* = \boldsymbol{\Sigma} \otimes \mathbf{P}_{(N-n),N} + \\ &\quad (\mathbf{I}_2 \otimes \mathbf{e}_1 \mathbf{e}_1' \otimes \bar{\sigma}_{11}^2 \mathbf{I}_{N-n}) + (\mathbf{I}_2 \otimes \mathbf{e}_2 \mathbf{e}_2' \otimes \bar{\sigma}_{21}^2 \mathbf{I}_{N-n}), \\ \text{Cov}_{SR}(\mathbf{Z}_I^*, \mathbf{Z}_{II}^*) &= \mathbf{V}_{I,II}^* = -N^{-1} \boldsymbol{\Sigma} \otimes \mathbf{J}_{n \times (N-n)},\end{aligned}\tag{3.17}$$

com

$$\begin{aligned}\boldsymbol{\mu} &= \left(\mu_1^{(1)} \quad \mu_1^{(2)} \quad \mu_2^{(1)} \quad \mu_2^{(2)} \right)', \\ \boldsymbol{\Sigma} &= \left(\boldsymbol{\Sigma}^{(1)} \otimes \mathbf{e}_1 \mathbf{e}_1' \right) + \left(\boldsymbol{\Sigma}^{(2)} \otimes \mathbf{e}_2 \mathbf{e}_2' \right),\end{aligned}\tag{3.18}$$

\mathbf{e}_i denotando a i -ésima coluna da matriz \mathbf{I}_2 , $\mathbf{P}_{a,b}$, $\mu_t^{(h)}$ e $\boldsymbol{\Sigma}^{(h)}$ definidos em (2.5), (3.3) e (3.6), respectivamente e $\bar{\sigma}_{h1}^2 [= \bar{\sigma}_{h2}^2]$, $h = 1, 2$, definido em (3.5).

Para ilustrar a notação, consideremos uma população composta por dois estratos de tamanho $N = 4$, dos quais selecionamos, respectivamente, uma amostra aleatória simples de tamanho $n = 3$ sem reposição. Seja $w_{hst} = 1$ para todo h, s e t ,

$$\begin{aligned}\mathbf{y}_1^{(1)} &= \begin{pmatrix} 2 \\ 5 \\ 4 \\ 7 \end{pmatrix}, & \mathbf{y}_2^{(1)} &= \begin{pmatrix} 5.5 \\ 10.5 \\ 7.5 \\ 11.5 \end{pmatrix}, & \mathbf{E}_1^{(1)} &= \begin{pmatrix} E_{111} \\ E_{121} \\ E_{131} \\ E_{141} \end{pmatrix}, & \mathbf{E}_2^{(1)} &= \begin{pmatrix} E_{112} \\ E_{122} \\ E_{132} \\ E_{142} \end{pmatrix}, \\ \mathbf{y}_1^{(2)} &= \begin{pmatrix} 1.5 \\ 6.5 \\ 5.5 \\ 2.5 \end{pmatrix}, & \mathbf{y}_2^{(2)} &= \begin{pmatrix} 6.75 \\ 13.6 \\ 13.5 \\ 9.55 \end{pmatrix}, & \mathbf{E}_1^{(2)} &= \begin{pmatrix} E_{211} \\ E_{221} \\ E_{231} \\ E_{241} \end{pmatrix}, & \mathbf{E}_2^{(2)} &= \begin{pmatrix} E_{212} \\ E_{222} \\ E_{232} \\ E_{242} \end{pmatrix},\end{aligned}$$

e suponhamos que o EMEn, tanto no pré-teste quanto no pós-teste, possa tomar apenas dois valores igualmente prováveis, dados por mais ou menos $\sigma_{1s1} [= \sigma_{1s2}]$, com $\sigma_{111}^2 = 1$, $\sigma_{121}^2 = 9$, $\sigma_{131}^2 = 0.25$ e $\sigma_{141}^2 = 100$ no estrato $h = 1$. De forma semelhante para o estrato $h = 2$, assumimos que $\sigma_{211}^2 = 9$, $\sigma_{221}^2 = 1$, $\sigma_{231}^2 = 100$ e $\sigma_{241}^2 = 49$. Então $\mu_1^{(1)} = 4.5$, $\mu_1^{(2)} = 4.0$, $\mu_2^{(1)} = 8.75$, $\mu_2^{(2)} = 10.85$,

$$\bar{\sigma}_{11}^2 = 27.5625, \bar{\sigma}_{21}^2 = 39.75,$$

$$\Sigma^{(1)} = \begin{pmatrix} 4.333 & 5.5 \\ 5.5 & 7.583 \end{pmatrix} \text{ e } \Sigma^{(2)} = \begin{pmatrix} 5.667 & 7.683 \\ 7.683 & 11.028 \end{pmatrix}.$$

Logo,

$$\mathbb{E}_{SR}(\mathbf{Z}_I^*) = (4.5 \ 4.0 \ 8.75 \ 10.85)' \otimes \mathbf{1}_3,$$

$$\mathbb{E}_{SR}(\mathbf{Z}_{II}^*) = (4.5 \ 4.0 \ 8.75 \ 10.85)',$$

$$\mathbf{V}_I^* = \begin{pmatrix} 4.333 & 0 & 5.5 & 0 \\ 0 & 5.667 & 0 & 7.683 \\ 5.5 & 0 & 7.583 & 0 \\ 0 & 7.683 & 0 & 11.028 \end{pmatrix} \otimes \mathbf{P}_{3,4} + \begin{pmatrix} 27.5625 & 0 & 0 & 0 \\ 0 & 39.75 & 0 & 0 \\ 0 & 0 & 27.5625 & 0 \\ 0 & 0 & 0 & 39.75 \end{pmatrix} \otimes \mathbf{I}_3,$$

$$\mathbf{V}_{II}^* = \left(1 - \frac{1}{4}\right) \begin{pmatrix} 4.333 & 0 & 5.5 & 0 \\ 0 & 5.667 & 0 & 7.683 \\ 5.5 & 0 & 7.583 & 0 \\ 0 & 7.683 & 0 & 11.028 \end{pmatrix} + \begin{pmatrix} 27.5625 & 0 & 0 & 0 \\ 0 & 39.75 & 0 & 0 \\ 0 & 0 & 27.5625 & 0 \\ 0 & 0 & 0 & 39.75 \end{pmatrix},$$

$$\mathbf{V}_{I,II}^* = -\frac{1}{4} \begin{pmatrix} 4.333 & 0 & 5.5 & 0 \\ 0 & 5.667 & 0 & 7.683 \\ 5.5 & 0 & 7.583 & 0 \\ 0 & 7.683 & 0 & 11.028 \end{pmatrix} \otimes \mathbf{1}_3,$$

porque $N - n = 1$.

Se as realizações das matrizes aleatórias $\mathbf{U}^{(1)}$ e $\mathbf{U}^{(2)}$ apresentadas em (3.1) são, respectivamente,

$$\mathbf{u}^{(1)} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \text{ e } \mathbf{u}^{(2)} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$

temos

$$\mathbf{Y}_{1\{\mathbf{U}^{(1)}=\mathbf{u}^{(1)}\}}^{(1)*} = (5 + E_{121}, 7 + E_{141}, 2 + E_{111}, 4 + E_{131})',$$

$$\mathbf{Y}_{1\{\mathbf{U}^{(2)}=\mathbf{u}^{(2)}\}}^{(2)*} = (2.5 + E_{241}, 5.5 + E_{231}, 6.5 + E_{221}, 1.5 + E_{211})',$$

$$\mathbf{Y}_{2\{\mathbf{U}^{(1)}=\mathbf{u}^{(1)}\}}^{(1)*} = (10.5 + E_{122}, 11.5 + E_{142}, 5.5 + E_{112}, 7.5 + E_{132})',$$

$$\begin{aligned} \mathbf{Y}_{2\{U^{(2)}=\mathbf{u}^{(2)}\}}^{(2)*} &= (9.55 + E_{242}, 13.5 + E_{232}, 13.6 + E_{222}, 6.75 + E_{212})', \\ \mathbf{Z}_{I\{U^{(1)}=\mathbf{u}^{(1)} \text{ e } U^{(2)}=\mathbf{u}^{(2)}\}}^* &= (5 + E_{121}, 7 + E_{141}, 2 + E_{111}, 2.5 + E_{241}, 5.5 + E_{231}, 6.5 + \\ &\quad E_{221}, 10.5 + E_{122}, 11.5 + E_{142}, 5.5 + E_{112}, 9.55 + E_{242}, 13.5 + E_{232}, 13.6 + E_{222})', \\ \mathbf{Z}_{II\{U^{(1)}=\mathbf{u}^{(1)} \text{ e } U^{(2)}=\mathbf{u}^{(2)}\}}^* &= (4 + E_{131}, 1.5 + E_{211}, 7.5 + E_{132}, 6.75 + E_{212})'. \end{aligned}$$

No passo seguinte minimizamos a esperança do erro quadrático médio do preditor linear sujeito à restrição de que seja não viciado [ver Apêndice D, ítem ii), pág. 98]. O preditor obtido é

$$\hat{T}^* = [\mathbf{D}'\mathbf{G}^{-1}(\mathbf{I}_4 \otimes \mathbf{1}_n)' + \mathbf{g}'_I] \mathbf{A}^{-1} \mathbf{Z}_I^* \quad (3.19)$$

em que

$$\begin{aligned} \mathbf{g}_I &= (c_{111} \ \dots \ c_{1n1} \ c_{211} \ \dots \ c_{2n1} \ c_{112} \ \dots \ c_{1n2} \ c_{212} \ \dots \ c_{2n2})', \\ \mathbf{A} &= \mathbf{I}_{4n} + \left[(\boldsymbol{\Sigma}^{(1)})^{-1} \otimes \mathbf{e}_1 \mathbf{e}'_1 \otimes \bar{\sigma}_{11}^2 \mathbf{I}_n \right] + \left[(\boldsymbol{\Sigma}^{(2)})^{-1} \otimes \mathbf{e}_2 \mathbf{e}'_2 \otimes \bar{\sigma}_{21}^2 \mathbf{I}_n \right], \\ \mathbf{G} &= (\mathbf{I}_4 \otimes \mathbf{1}_n)' \mathbf{A}^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_n), \\ \mathbf{D} &= (\mathbf{I}_4 \otimes \mathbf{1}_n)' (\mathbf{I}_{4n} - \mathbf{A}^{-1}) \mathbf{g}_I + (N - n) (\bar{c}_{II,11} \ \bar{c}_{II,21} \ \bar{c}_{II,12} \ \bar{c}_{II,22})', \\ \bar{c}_{II,ht} &= (N - n)^{-1} \sum_{i=n+1}^N c_{hit}, \end{aligned}$$

com $\boldsymbol{\Sigma}^{(h)}$, c_{hit} e \mathbf{Z}_I^* definidos em (3.6), (3.9) e (3.15), respectivamente, e $\bar{\sigma}_{h1}^2 [= \bar{\sigma}_{h2}^2]$, $h = 1, 2$.

A variância do preditor \hat{T}^* é (ver Apêndice D, pág. 99)

$$\begin{aligned} \text{Var}_{SR}(\hat{T}^* - T) &= \mathbf{D}'\boldsymbol{\Sigma}\mathbf{G}^{-1}\mathbf{D} + \mathbf{g}'_I(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)(\mathbf{I}_{4n} - \mathbf{A}^{-1})\mathbf{g}_I \\ &\quad + \mathbf{g}'_{II}(\boldsymbol{\Sigma} \otimes \mathbf{P}_{N-n})\mathbf{g}_{II} + (N - n)^{-1}\mathbf{g}'_{II}(\boldsymbol{\Sigma} \otimes \mathbf{J}_{N-n})\mathbf{g}_{II}, \end{aligned} \quad (3.20)$$

com $\mathbf{g}_{II} = (c_{1,(n+1),1} \ \dots \ c_{1N1} \ c_{2,(n+1),1} \ \dots \ c_{2N1} \ c_{1,(n+1),2} \ \dots \ c_{1N2} \ c_{2,(n+1),2} \ \dots \ c_{2N2})'$, \mathbf{P}_a definido em (2.5) e $\boldsymbol{\Sigma}$, em (3.18).

Em geral, ha interesse em avaliar a diferença da resposta média dos dois estratos no pré-teste (3.10), avaliar o efeito da intervenção dentro de cada estrato (3.11) e avaliar a diferença dos efeitos da intervenção entre os dois estratos (3.12). Para os três casos, é possível escrever (3.9) como

$$T = c_{11} \sum_{i=1}^N Y_{1i1} + c_{12} \sum_{i=1}^N Y_{1i2} + c_{21} \sum_{i=1}^N Y_{2i1} + c_{22} \sum_{i=1}^N Y_{2i2}$$

em que $c_{hit} = c_{ht}$, $h = 1, 2$, $t = 1, 2$, $i = 1, \dots, N$; portanto, (3.19) se reduz a

$$\hat{T}^* = N \left(c_{11} \bar{Y}_{I,11}^* + c_{21} \bar{Y}_{I,21}^* + c_{12} \bar{Y}_{I,12}^* + c_{22} \bar{Y}_{I,22}^* \right), \quad (3.21)$$

com $\bar{Y}_{I,ht}^* = n^{-1} \sum_{i=1}^n Y_{hit}^*$, e sua variância se reduz a

$$\text{Var}_{SR} \left(\hat{T}^* - T \right) = \frac{N^2}{n} (c_{11} \ c_{21} \ c_{12} \ c_{22}) \Sigma (\mathbf{A}_0 - f \mathbf{I}_4) (c_{11} \ c_{21} \ c_{12} \ c_{22})', \quad (3.22)$$

com

$$\mathbf{A}_0 = \mathbf{I}_4 + \bar{\sigma}_{11}^2 \left(\Sigma^{(1)} \right)^{-1} \otimes \mathbf{e}_1 \mathbf{e}_1' + \bar{\sigma}_{21}^2 \left(\Sigma^{(2)} \right)^{-1} \otimes \mathbf{e}_2 \mathbf{e}_2'$$

e $f = n/N$. Essa expressão corresponde a

$$\text{Var}_{SR} \left(\hat{T}^* - T \right) = (N^2/n) \times \begin{pmatrix} c_{11} \\ c_{21} \\ c_{12} \\ c_{22} \end{pmatrix}' \begin{pmatrix} (1-f)\sigma_{11}^{(1)} + \bar{\sigma}_{11}^2 & 0 & (1-f)\sigma_{12}^{(1)} & 0 \\ 0 & (1-f)\sigma_{11}^{(2)} + \bar{\sigma}_{21}^2 & 0 & (1-f)\sigma_{12}^{(2)} \\ (1-f)\sigma_{12}^{(1)} & 0 & (1-f)\sigma_{22}^{(1)} + \bar{\sigma}_{11}^2 & 0 \\ 0 & (1-f)\sigma_{12}^{(2)} & 0 & (1-f)\sigma_{22}^{(2)} + \bar{\sigma}_{21}^2 \end{pmatrix} \begin{pmatrix} c_{11} \\ c_{21} \\ c_{12} \\ c_{22} \end{pmatrix}.$$

Os estimadores de T_o (3.10), T_h (3.11) e T_d (3.12) e suas variâncias são, respectivamente,

- $\hat{T}_o^* = \bar{Y}_{I,11}^* - \bar{Y}_{I,21}^* \quad \mathbf{e}$

$$\text{Var} \left(\hat{T}_o^* - T_o \right) = n^{-1} \left[(1-f) \left(\sigma_{11}^{(1)} + \sigma_{11}^{(2)} \right) + \bar{\sigma}_{11}^2 + \bar{\sigma}_{21}^2 \right],$$

- $\hat{T}_h^* = \bar{Y}_{I,h2}^* - \bar{Y}_{I,h1}^*, \quad h = 1, 2, \quad \mathbf{e}$

$$\text{Var} \left(\hat{T}_h^* - T_h \right) = n^{-1} \left[(1-f) \left(\sigma_{11}^{(h)} + \sigma_{22}^{(h)} - 2\sigma_{12}^{(h)} \right) + 2\bar{\sigma}_{h1}^2 \right],$$

- $\hat{T}_d^* = \bar{Y}_{I,12}^* - \bar{Y}_{I,11}^* - \left(\bar{Y}_{I,22}^* - \bar{Y}_{I,21}^* \right) \quad \mathbf{e}$

$$\text{Var} \left(\hat{T}_d^* - T_d \right) = n^{-1} \left[(1-f) \left(\sigma_{11}^{(1)} + \sigma_{22}^{(1)} - 2\sigma_{12}^{(1)} + \sigma_{11}^{(2)} + \sigma_{22}^{(2)} - 2\sigma_{12}^{(2)} \right) + 2\bar{\sigma}_{11}^2 + 2\bar{\sigma}_{21}^2 \right].$$

Por conseguinte, os estimadores de T_o (3.10), T_h (3.11) e T_d (3.12) não dependem da fração amostral $f = n/N$ e, dado N , sua variância diminui quando n aumenta.

O correspondente preditor de T (3.9) sob o modelo (3.8) é

$$\tilde{T} = [\mathbf{D}'\mathbf{G}^{-1}(\mathbf{I}_4 \otimes \mathbf{1}_n)' + \mathbf{g}'_I] \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{Z}}_I \quad (3.23)$$

em que

$$\begin{aligned} \tilde{\mathbf{Z}}_I &= \left(\tilde{Y}_{111} \ \dots \ \tilde{Y}_{1n1} \ \tilde{Y}_{211} \ \dots \ \tilde{Y}_{2n1} \ \tilde{Y}_{112} \ \dots \ \tilde{Y}_{1n2} \ \tilde{Y}_{212} \ \dots \ \tilde{Y}_{2n2} \right)', \\ \tilde{\mathbf{A}} &= \mathbf{I}_{4n} + \left[\left(\boldsymbol{\Sigma}^{(1)} \right)^{-1} \otimes \mathbf{e}_1 \mathbf{e}'_1 \otimes \bigoplus_{i=1}^n \tilde{\sigma}_{i1}^2 \right] + \left[\left(\boldsymbol{\Sigma}^{(2)} \right)^{-1} \otimes \mathbf{e}_2 \mathbf{e}'_2 \otimes \bigoplus_{i=1}^n \tilde{\sigma}_{2i1}^2 \right], \end{aligned}$$

com $\tilde{\sigma}_{hi1}^2 [= \tilde{\sigma}_{hi2}^2]$ definido em (3.7), e \mathbf{g}_I , \mathbf{G} e \mathbf{D} definidos em (3.19) com a substituição de \mathbf{A} por $\tilde{\mathbf{A}}$.

A variância do preditor \tilde{T} (3.23) é

$$\begin{aligned} \text{Var}_{SR}(\tilde{T} - T) &= \mathbf{D}'\boldsymbol{\Sigma}\mathbf{G}^{-1}\mathbf{D} + \mathbf{g}'_I(\boldsymbol{\Sigma} \otimes \mathbf{I}_n) \left(\mathbf{I}_{4n} - \tilde{\mathbf{A}}^{-1} \right) \mathbf{g}_I \\ &\quad + \mathbf{g}'_{II}(\boldsymbol{\Sigma} \otimes \mathbf{P}_{N-n}) \mathbf{g}_{II} + (N - n)^{-1} \mathbf{g}'_{II}(\boldsymbol{\Sigma} \otimes \mathbf{J}_{N-n}) \mathbf{g}_{II}, \end{aligned} \quad (3.24)$$

com $\boldsymbol{\Sigma}$ e \mathbf{g}_{II} definidos em (3.18) e (3.20), respectivamente, e \mathbf{g}_I , \mathbf{G} e \mathbf{D} definidos em (3.19) com a substituição de \mathbf{A} por $\tilde{\mathbf{A}}$.

3.1.2 Exemplo

Nesta subsecção tomamos $w_{hst} = 1$ para todo h, s e t porque o nosso interesse recai na estimação de T_d (3.12), i.e., na diferença dos efeitos da intervenção entre os dois estratos, que é função das respostas latentes pré-teste e pós-teste não ponderadas. Consideremos os dados do Exemplo 1.1. Na Tabela 3.1 estão apresentadas algumas medidas resumo.

Tabela 3.1: Medidas resumo para os dados do Exemplo 1.1.

Estrato	Ocasão	Média Amostral	Variância Amostral	Mínimo	Máximo
Indometacina	pré-teste	117.3	271.5	92	148
	pós-teste	100.7	284.6	76	132
Nifedipina	pré-teste	94.3	108.2	82	116
	pós-teste	77	100	58	100

$$\hat{\sigma}_{12}^{(1)} = 255.0, \quad \hat{\sigma}_{12}^{(2)} = 74.5^b$$

^b $\hat{\sigma}_{12}^{(h)} = (n - 1)^{-1} \sum_{i=1}^n (Y_{hi1} - \bar{Y}_{I,h1})(Y_{hi2} - \bar{Y}_{I,h2})$ com $\bar{Y}_{I,ht} = n^{-1} \sum_{i=1}^n Y_{hit}$.

Neste caso, temos $n = 12$ e vamos supor $N = 12, 50, 100, 300, 500$ e 1000 . Para o caso com EMEn, suponhamos que $\sigma_{1s1}^2 = \sigma_{1s2}^2 = 25s \{2N/[N(N+1)]\}$ e $\sigma_{2s1}^2 = \sigma_{2s2}^2 = 100s \times \{2N/[N(N+1)]\}$, $s = 1, \dots, N$, e para o caso com EMEx, suponhamos que $\tilde{\sigma}_{1i1}^2 = \tilde{\sigma}_{1i2}^2 = 25i \{2n/[n(n+1)]\}$ e $\tilde{\sigma}_{2i1}^2 = \tilde{\sigma}_{2i2}^2 = 100i \{2n/[n(n+1)]\}$ para $i \leq n$, e $\tilde{\sigma}_{1i1}^2 = \tilde{\sigma}_{1i2}^2 = 0$ e $\tilde{\sigma}_{2i1}^2 = \tilde{\sigma}_{2i2}^2 = 0$ para $i > n$ ^c. Também é considerado o caso sem erro de medida para o qual a estimativa de T_d (3.12) é rotulada por \hat{T}_d . Sob cada um dos casos, o valor da estimativa de T_d , da variância estimada e do intervalo de confiança aproximado, baseado na desigualdade de Chebyshev com um nível de confiança de $1 - \alpha = 0.90$, são apresentados na Tabela 3.2. A estimativa obtida via modelo (1.2) para análise de medidas repetidas [ver Bonate (2000)], que corresponde à estimativa da combinação linear

$$\mu_{12} - \mu_{11} - (\mu_{22} - \mu_{21}),$$

é igual a \hat{T}_d com $\widehat{Var} \{ \hat{\mu}_{12} - \hat{\mu}_{11} - (\hat{\mu}_{22} - \hat{\mu}_{21}) \} = 4 \times 3.991^2 = 63.7$. Observamos na Tabela 3.2 que a variância estimada de \hat{T}_d^* é maior que a variância estimada de \tilde{T}_d^* . Além disso, observamos que o tamanho da população (N) gerou um pequeno efeito na variância dos estimadores para $N \geq 50$.

Tabela 3.2: Estimativas da diferença dos efeitos, na pressão arterial média em cães, da aplicação de M_gSO_4 entre o grupo submetido à droga Indometacina e o grupo submetido à droga Nifedipina, variância estimada e intervalos de confiança, considerando o caso com erros de medida endógenos, com erros de medida exógenos e sem erros de medida.

N	\hat{T}_d^*	$\widehat{Var}_{SR}(\hat{T}_d^*)$	$\hat{T}_d^* \pm a$	\tilde{T}_d^*	$\widehat{Var}_{SR}(\tilde{T}_d^*)$	$\tilde{T}_d^* \pm a$	\hat{T}_d	$\widehat{Var}_s(\hat{T}_d)$	$\hat{T}_d \pm a$
12	0.7	20.8	0.7 ± 14.4	0.3	15.9	0.3 ± 12.6	0.7	0	0
50	0.7	27.5	0.7 ± 16.6	0.3	22.6	0.3 ± 15.0	0.7	6.7	0.7 ± 8.2
100	0.7	28.5	0.7 ± 16.9	0.3	23.6	0.3 ± 15.4	0.7	7.7	0.7 ± 8.8
300	0.7	29.3	0.7 ± 17.1	0.3	24.3	0.3 ± 15.6	0.7	8.4	0.7 ± 9.2
500	0.7	29.4	0.7 ± 17.1	0.3	24.4	0.3 ± 15.6	0.7	8.6	0.7 ± 9.3
1000	0.7	29.4	0.7 ± 17.1	0.3	24.5	0.3 ± 15.6	0.7	8.8	0.7 ± 9.4

$$a = \sqrt{\widehat{Var}_{SR}(\cdot)/\alpha}, \alpha = 0.1$$

^c Note que o erro de medida nos dois casos muda de uma unidade (respectivamente, posição) a outra, mas, num caso é ligado à unidade e no outro caso é ligado à posição (na amostra no estrato) e é tal que a variância média dos EMEn sobre todas as unidades é igual à variância média dos EMEx sobre as unidades na amostra, i.e., $N^{-1} \sum_{s=1}^N \sigma_{1s1}^2 = 25$,

$N^{-1} \sum_{s=1}^N \sigma_{2s1}^2 = 100$, $n^{-1} \sum_{i=1}^n \tilde{\sigma}_{1i1}^2 = 25$ e $n^{-1} \sum_{i=1}^n \tilde{\sigma}_{2i1}^2 = 100$.

3.2 Modelo misto para estudos pré-teste/pós-teste em populações finitas com dois tipos de intervenção

Neste cenário, associamos a cada unidade, rotulada por $s = 1, \dots, N$, os valores fixos y_{s1}^\bullet , $y_{s2}^{C\bullet}$ e $y_{s2}^{T\bullet}$ que representam, respectivamente, a resposta antes (pré-teste) de qualquer intervenção, a resposta depois (pós-teste) de submeter uma unidade selecionada aleatoriamente (com mesma probabilidade) dentre as N à intervenção C , e a resposta depois (pós-teste) de submeter uma unidade selecionada aleatoriamente (com mesma probabilidade) dentre as N à intervenção T . Além disso, assumimos que o valor do pré-teste y_{s1}^\bullet e os valores pós-teste $y_{s2}^{C\bullet}$ e $y_{s2}^{T\bullet}$ são observados possivelmente com erro de medida. Na prática, a unidade s é submetida de forma aleatória apenas a uma das duas intervenções, C ou T , e é por isto que vamos chamar a $y_{s2}^{C\bullet}$ e $y_{s2}^{T\bullet}$ de *respostas potenciais* no pós-teste para o indivíduo s . Como em muitas situações é conveniente trabalhar com os valores latentes ponderados, definimos $y_{s1} = w_{s1}y_{s1}^\bullet$, $y_{s2}^C = w_{s2}^C y_{s2}^{C\bullet}$ e $y_{s2}^T = w_{s2}^T y_{s2}^{T\bullet}$, em que w_{s1} , w_{s2}^C e w_{s2}^T , $s = 1, \dots, N$, representam pesos. O enfoque de *respostas potenciais* (ou observações potenciais) tem sido abordado por diferentes autores; entre eles, citamos Leon, Tsiatis & Davidian (2003) no contexto de dados do tipo pré-teste/pós-teste, e Hinkelmann & Kempthorne (1994, p. 148) e Rubin (2005) no contexto mais geral de planejamentos experimentais. Uma excelente referência para elucidar claramente a perspectiva básica de *observações potenciais* ou *respostas potenciais* é Holland (1986).

Assim como nos casos anteriores, o processo da amostragem aleatória simples é formalizado por meio de um modelo para populações finitas. Associando a cada unidade um novo rótulo, $i = 1, \dots, N$, segundo sua posição numa permutação aleatória, podemos representar as respostas pré-teste e as respostas potenciais pós-teste pela matriz aleatória $(\mathbf{Y}_1 \ \mathbf{Y}_2^C \ \mathbf{Y}_2^T)$ em que $\mathbf{Y}_1 = (Y_{11} \ \dots \ Y_{i1} \ \dots \ Y_{N1})'$, $\mathbf{Y}_2^C = (Y_{12}^C \ \dots \ Y_{i2}^C \ \dots \ Y_{N2}^C)'$ e $\mathbf{Y}_2^T = (Y_{12}^T \ \dots \ Y_{i2}^T \ \dots \ Y_{N2}^T)'$.

Para associar os valores $\mathbf{y}_1 = (y_{11} \ \dots \ y_{s1} \ \dots \ y_{N1})'$, $\mathbf{y}_2^C = (y_{12}^C \ \dots \ y_{s2}^C \ \dots \ y_{N2}^C)'$ e $\mathbf{y}_2^T = (y_{12}^T \ \dots \ y_{s2}^T \ \dots \ y_{N2}^T)'$ a \mathbf{Y}_1 , \mathbf{Y}_2^C e \mathbf{Y}_2^T consideramos

$$(\mathbf{Y}_1 \ \mathbf{Y}_2^C \ \mathbf{Y}_2^T) = \mathbf{U} (\mathbf{y}_1 \ \mathbf{y}_2^C \ \mathbf{y}_2^T) \quad (3.25)$$

em que \mathbf{U} está definida em (2.3).

Incorporamos erro de medida endógeno (EMEn) tanto no pré-teste quanto no pós-teste em (3.25) adicionando a matriz aleatória

$$(\mathbf{E}_1^\bullet \ \mathbf{E}_2^{C\bullet} \ \mathbf{E}_2^{T\bullet}),$$

em que

$$\mathbf{E}_1^\bullet = (E_{11}^\bullet \dots E_{s1}^\bullet \dots E_{N1}^\bullet)', \mathbf{E}_2^{C\bullet} = (E_{12}^{C\bullet} \dots E_{s2}^{C\bullet} \dots E_{N2}^{C\bullet})' \text{ e } \mathbf{E}_2^{T\bullet} = (E_{12}^{T\bullet} \dots E_{s2}^{T\bullet} \dots E_{N2}^{T\bullet})',$$

com $\mathbb{E}_R(E_{s1}^\bullet) = \mathbb{E}_R(E_{s2}^{C\bullet}) = \mathbb{E}_R(E_{s2}^{T\bullet}) = 0$, $\text{Var}_R(E_{s1}^\bullet) = \sigma_{s1}^{\bullet 2}$, $\text{Var}_R(E_{s2}^{C\bullet}) = (\sigma_{s2}^{C\bullet})^2$, $\text{Var}_R(E_{s2}^{T\bullet}) = (\sigma_{s2}^{T\bullet})^2$ e E_{s1}^\bullet , $E_{s2}^{C\bullet}$ e $E_{s2}^{T\bullet}$, $s = 1, \dots, N$, independentes duas a duas, à matriz

$$(\mathbf{y}_1^\bullet \ \mathbf{y}_2^{C\bullet} \ \mathbf{y}_2^{T\bullet}),$$

em que $\mathbf{y}_1^\bullet = (y_{11}^\bullet \dots y_{s1}^\bullet \dots y_{N1}^\bullet)'$, $\mathbf{y}_2^{C\bullet} = (y_{12}^{C\bullet} \dots y_{s2}^{C\bullet} \dots y_{N2}^{C\bullet})'$ e $\mathbf{y}_2^{T\bullet} = (y_{12}^{T\bullet} \dots y_{s2}^{T\bullet} \dots y_{N2}^{T\bullet})'$. O modelo pode então ser representado como

$$(\mathbf{Y}_1^* \ \mathbf{Y}_2^{C*} \ \mathbf{Y}_2^{T*}) = \mathbf{U} [(\mathbf{y}_1 \ \mathbf{y}_2^C \ \mathbf{y}_2^T) + (\mathbf{E}_1 \ \mathbf{E}_2^C \ \mathbf{E}_2^T)] \quad (3.26)$$

em que

$$\mathbf{Y}_1^* = (Y_{11}^* \dots Y_{i1}^* \dots Y_{N1}^*)', \mathbf{Y}_2^{C*} = (Y_{12}^{C*} \dots Y_{i2}^{C*} \dots Y_{N2}^{C*})' \text{ e } \mathbf{Y}_2^{T*} = (Y_{12}^{T*} \dots Y_{i2}^{T*} \dots Y_{N2}^{T*})',$$

e

$$\mathbf{E}_1 = (E_{11} \dots E_{s1} \dots E_{N1})', \mathbf{E}_2^C = (E_{12}^C \dots E_{s2}^C \dots E_{N2}^C)' \text{ e } \mathbf{E}_2^T = (E_{12}^T \dots E_{s2}^T \dots E_{N2}^T)'$$

são vetores aleatórios, com $E_{s1} = w_{s1}E_{s1}^\bullet$, $E_{s2}^C = w_{s2}^C E_{s2}^{C\bullet}$ e $E_{s2}^T = w_{s2}^T E_{s2}^{T\bullet}$. Note que $\mathbf{y}_1 = (w_{11}y_{11}^\bullet \dots w_{s1}y_{s1}^\bullet \dots w_{N1}y_{N1}^\bullet)'$, $\mathbf{y}_2^C = (w_{12}^C y_{12}^{C\bullet} \dots w_{s2}^C y_{s2}^{C\bullet} \dots w_{N2}^C y_{N2}^{C\bullet})'$ e $\mathbf{y}_2^T = (w_{12}^T y_{12}^{T\bullet} \dots w_{s2}^T y_{s2}^{T\bullet} \dots w_{N2}^T y_{N2}^{T\bullet})'$. Então,

$$\begin{aligned} \mathbb{E}_R(E_{s1}) &= \mathbb{E}_R(E_{s2}^C) = \mathbb{E}_R(E_{s2}^T) = 0, \\ \text{Var}_R(E_{s1}) &= \sigma_{s1}^2 = w_{s1}^2 \sigma_{s1}^{\bullet 2}, \\ \text{Var}_R(E_{s2}^C) &= (\sigma_{s2}^C)^2 = (w_{s2}^C)^2 (\sigma_{s2}^{C\bullet})^2, \\ \text{Var}_R(E_{s2}^T) &= (\sigma_{s2}^T)^2 = (w_{s2}^T)^2 (\sigma_{s2}^{T\bullet})^2, \end{aligned}$$

e as variáveis E_{s1} , E_{s2}^C e E_{s2}^T , $s = 1, \dots, N$, são independentes duas a duas. Em particular, para a

posição i dentro de uma permutação,

$$\begin{aligned} Y_{i1}^* &= \sum_{s=1}^N U_{is} (y_{s1} + E_{s1}) = \mu_1 + d_{i1} + \sum_{s=1}^N U_{is} E_{s1}, \\ Y_{i2}^{C*} &= \sum_{s=1}^N U_{is} (y_{s2}^C + E_{s2}^C) = \mu_2^C + d_{i2}^C + \sum_{s=1}^N U_{is} E_{s2}^C, \\ Y_{i2}^{T*} &= \sum_{s=1}^N U_{is} (y_{s2}^T + E_{s2}^T) = \mu_2^T + d_{i2}^T + \sum_{s=1}^N U_{is} E_{s2}^T, \end{aligned} \quad (3.27)$$

em que

$$\begin{aligned} d_{i1} &= \sum_{s=1}^N U_{is} \delta_{s1}, \quad \text{com} \quad \delta_{s1} = y_{s1} - \mu_1 \quad \text{e} \quad \mu_1 = N^{-1} \sum_{s=1}^N y_{s1}, \\ d_{i2}^C &= \sum_{s=1}^N U_{is} \delta_{s2}^C, \quad \text{com} \quad \delta_{s2}^C = y_{s2}^C - \mu_2^C \quad \text{e} \quad \mu_2^C = N^{-1} \sum_{s=1}^N y_{s2}^C, \\ d_{i2}^T &= \sum_{s=1}^N U_{is} \delta_{s2}^T, \quad \text{com} \quad \delta_{s2}^T = y_{s2}^T - \mu_2^T \quad \text{e} \quad \mu_2^T = N^{-1} \sum_{s=1}^N y_{s2}^T. \end{aligned}$$

μ_2^h , $h = C, T$, se pode reparametrizar como $\mu_2^h = \mu_1 + \gamma + \alpha_h$ em que $\gamma = (2N)^{-1} \sum_{s=1}^N (y_{s2}^C + y_{s2}^T)$ e $\alpha_h = \mu_2^h - \mu_1 - \gamma$.

O modelo misto para populações finitas com EMEn é

$$\mathbf{Z}^* = \text{vec} [(\mathbf{Y}_1^* \quad \mathbf{Y}_2^{C*} \quad \mathbf{Y}_2^{T*})] = (\mathbf{I}_3 \otimes \mathbf{1}_N) \boldsymbol{\mu} + \mathbf{d} + \mathbf{E}^* \quad (3.28)$$

em que $\boldsymbol{\mu} = (\mu_1 \quad \mu_2^C \quad \mu_2^T)'$, corresponde ao vetor de médias populacionais pré-teste, μ_1 , e pós-teste, μ_2^C (respectivamente, μ_2^T), sob a intervenção C (respectivamente, T), i.e., é um vetor de efeitos fixos,

$$\mathbf{d} = \left[\mathbf{d}'_1 \quad (\mathbf{d}_2^C)' \quad (\mathbf{d}_2^T)' \right]',$$

com $\mathbf{d}'_1 = (d_{11} \dots d_{i1} \dots d_{N1})'$, $\mathbf{d}_2^C = (d_{12}^C \dots d_{i2}^C \dots d_{N2}^C)'$ e $\mathbf{d}_2^T = (d_{12}^T \dots d_{i2}^T \dots d_{N2}^T)'$, é um vetor de efeitos aleatórios, e

$$\mathbf{E}^* = \left[\mathbf{E}'_1 \quad (\mathbf{E}_2^C)' \quad (\mathbf{E}_2^T)' \right]',$$

com $\mathbf{E}_1 = \mathbf{Y}_1^* - \mathbf{Y}_1$, $\mathbf{E}_2^C = \mathbf{Y}_2^{C*} - \mathbf{Y}_2^C$ e $\mathbf{E}_2^T = \mathbf{Y}_2^{T*} - \mathbf{Y}_2^T$.

O vetor de médias e a matriz de covariâncias de \mathbf{Z}^* são dados por (ver Apêndice A, pág. 83)

$$\mathbb{E}_{SR}(\mathbf{Z}^*) = (\mathbf{I}_3 \otimes \mathbf{1}_N) \boldsymbol{\mu},$$

e

$$\text{Var}_{SR}(\mathbf{Z}^*) = \text{diag} \{ \bar{\sigma}_1^2 \mathbf{I}_N, \bar{\sigma}_2^{C2} \mathbf{I}_N, \bar{\sigma}_2^{T2} \mathbf{I}_N \} + \boldsymbol{\Sigma} \otimes \mathbf{P}_N,$$

em que \mathbf{P}_a está definida em (2.5), e

$$\begin{aligned} \bar{\sigma}_1^2 &= \mathbb{E}_S \left(\sum_{s=1}^N \sigma_{s1}^2 U_{is} \right) = N^{-1} \sum_{s=1}^N \sigma_{s1}^2, \\ \bar{\sigma}_2^{C2} &= \mathbb{E}_S \left(\sum_{s=1}^N (\sigma_{s2}^C)^2 U_{is} \right) = N^{-1} \sum_{s=1}^N (\sigma_{s2}^C)^2, \\ \bar{\sigma}_2^{T2} &= \mathbb{E}_S \left(\sum_{s=1}^N (\sigma_{s2}^T)^2 U_{is} \right) = N^{-1} \sum_{s=1}^N (\sigma_{s2}^T)^2, \end{aligned} \quad (3.29)$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{21}^C & \sigma_{21}^T \\ \sigma_{21}^C & \sigma_{22}^{CC} & \sigma_{22}^{CT} \\ \sigma_{21}^T & \sigma_{22}^{CT} & \sigma_{22}^{TT} \end{pmatrix} \quad (3.30)$$

com

$$\begin{aligned} \sigma_{11} &= (N-1)^{-1} \sum_{s=1}^N (y_{s1} - \mu_1)^2, \\ \sigma_{21}^h &= (N-1)^{-1} \sum_{s=1}^N (y_{s2}^h - \mu_2^h) (y_{s1} - \mu_1), \quad h = C, T, \\ \sigma_{22}^{hh'} &= (N-1)^{-1} \sum_{s=1}^N (y_{s2}^h - \mu_2^h) (y_{s2}^{h'} - \mu_2^{h'}), \quad h = C, T, \quad h' = C, T. \end{aligned}$$

Quando os erros de medida no pré-teste e no pós-teste são exógenos (EMEx), eles são incorporados em (3.25) por meio da adição da matriz aleatória

$$\left(\tilde{\mathbf{E}}_1^\bullet \quad \tilde{\mathbf{E}}_2^{C\bullet} \quad \tilde{\mathbf{E}}_2^{T\bullet} \right),$$

em que

$$\tilde{\mathbf{E}}_1^\bullet = \left(\tilde{E}_{11}^\bullet \dots \tilde{E}_{i1}^\bullet \dots \tilde{E}_{N1}^\bullet \right)', \quad \tilde{\mathbf{E}}_2^{C\bullet} = \left(\tilde{E}_{12}^{C\bullet} \dots \tilde{E}_{i2}^{C\bullet} \dots \tilde{E}_{N2}^{C\bullet} \right)' \text{ e } \tilde{\mathbf{E}}_2^{T\bullet} = \left(\tilde{E}_{12}^{T\bullet} \dots \tilde{E}_{i2}^{T\bullet} \dots \tilde{E}_{N2}^{T\bullet} \right)',$$

com $\mathbb{E}_R \left(\tilde{E}_{i1}^\bullet \right) = \mathbb{E}_R \left(\tilde{E}_{i2}^{C\bullet} \right) = \mathbb{E}_R \left(\tilde{E}_{i2}^{T\bullet} \right) = 0$, $\text{Var}_R \left(\tilde{E}_{i1}^\bullet \right) = \tilde{\sigma}_{i1}^2$, $\text{Var}_R \left(\tilde{E}_{i2}^{C\bullet} \right) = \left(\tilde{\sigma}_{i2}^{C\bullet} \right)^2$, $\text{Var}_R \left(\tilde{E}_{i2}^{T\bullet} \right) = \left(\tilde{\sigma}_{i2}^{T\bullet} \right)^2$ e \tilde{E}_{i1}^\bullet , $\tilde{E}_{i2}^{C\bullet}$ e $\tilde{E}_{i2}^{T\bullet}$, $i = 1, \dots, N$, independentes duas a duas, à matriz

$$\left(\mathbf{Y}_1^\bullet \quad \mathbf{Y}_2^{C\bullet} \quad \mathbf{Y}_2^{T\bullet} \right)$$

em que $\mathbf{Y}_1^\bullet = \mathbf{U} \mathbf{y}_1^\bullet$, $\mathbf{Y}_2^{C\bullet} = \mathbf{U} \mathbf{y}_2^{C\bullet}$ e $\mathbf{Y}_2^{T\bullet} = \mathbf{U} \mathbf{y}_2^{T\bullet}$. O modelo pode ser escrito como

$$\left(\tilde{\mathbf{Y}}_1 \quad \tilde{\mathbf{Y}}_2^C \quad \tilde{\mathbf{Y}}_2^T \right) = \mathbf{U} \left(\mathbf{y}_1 \quad \mathbf{y}_2^C \quad \mathbf{y}_2^T \right) + \left(\tilde{\mathbf{E}}_1 \quad \tilde{\mathbf{E}}_2^C \quad \tilde{\mathbf{E}}_2^T \right) \quad (3.31)$$

em que

$$\tilde{\mathbf{Y}}_1 = \left(\tilde{Y}_{11} \dots \tilde{Y}_{i1} \dots \tilde{Y}_{N1} \right)', \quad \tilde{\mathbf{Y}}_2^C = \left(\tilde{Y}_{12}^C \dots \tilde{Y}_{i2}^C \dots \tilde{Y}_{N2}^C \right)' \text{ e } \tilde{\mathbf{Y}}_2^T = \left(\tilde{Y}_{12}^T \dots \tilde{Y}_{i2}^T \dots \tilde{Y}_{N2}^T \right)',$$

e

$$\tilde{\mathbf{E}}_1 = \left(\tilde{E}_{11} \dots \tilde{E}_{i1} \dots \tilde{E}_{N1} \right)', \quad \tilde{\mathbf{E}}_2^C = \left(\tilde{E}_{12}^C \dots \tilde{E}_{i2}^C \dots \tilde{E}_{N2}^C \right)' \text{ e } \tilde{\mathbf{E}}_2^T = \left(\tilde{E}_{12}^T \dots \tilde{E}_{i2}^T \dots \tilde{E}_{N2}^T \right)',$$

com $\tilde{E}_{i1} = \tilde{E}_{i1}^\bullet \sum_{s=1}^N U_{is} w_{s1}$, $\tilde{E}_{i2}^C = \tilde{E}_{i2}^{C\bullet} \sum_{s=1}^N U_{is} w_{s2}^C$ e $\tilde{E}_{i2}^T = \tilde{E}_{i2}^{T\bullet} \sum_{s=1}^N U_{is} w_{s2}^T$. Logo

$$\begin{aligned} \mathbb{E}_R \left(\tilde{E}_{i1} \right) &= \mathbb{E}_R \left(\tilde{E}_{i2}^C \right) = \mathbb{E}_R \left(\tilde{E}_{i2}^T \right) = 0, \\ \text{Var}_{SR} \left(\tilde{E}_{i1} \right) &= \tilde{\sigma}_{i1}^2 = \mu_{w1} \tilde{\sigma}_{i1}^{\bullet 2}, \\ \text{Var}_{SR} \left(\tilde{E}_{i2}^C \right) &= \left(\tilde{\sigma}_{i2}^C \right)^2 = \mu_{wC} \left(\tilde{\sigma}_{i2}^{C\bullet} \right)^2, \\ \text{Var}_{SR} \left(\tilde{E}_{i2}^T \right) &= \left(\tilde{\sigma}_{i2}^T \right)^2 = \mu_{wT} \left(\tilde{\sigma}_{i2}^{T\bullet} \right)^2, \end{aligned}$$

com $\mu_{w1} = N^{-1} \sum_{s=1}^N w_{s1}^2$, $\mu_{wC} = N^{-1} \sum_{s=1}^N \left(w_{s2}^C \right)^2$ e $\mu_{wT} = N^{-1} \sum_{s=1}^N \left(w_{s2}^T \right)^2$, e \tilde{E}_{i1} , \tilde{E}_{i2}^C e \tilde{E}_{i2}^T , $i =$

$1, \dots, N$, independentes duas a duas. Neste caso, para a posição i dentro de uma permutação,

$$\begin{aligned}\tilde{Y}_{i1} &= \sum_{s=1}^N U_{is} y_{s1} + \tilde{E}_{i1} = \mu_1 + d_{i1} + \tilde{E}_{i1}, \\ \tilde{Y}_{i2}^C &= \sum_{s=1}^N U_{is} y_{s2}^C + \tilde{E}_{i2}^C = \mu_2^C + d_{i2}^C + \tilde{E}_{i2}^C, \\ \tilde{Y}_{i2}^T &= \sum_{s=1}^N U_{is} y_{s2}^T + \tilde{E}_{i2}^T = \mu_2^T + d_{i2}^T + \tilde{E}_{i2}^T,\end{aligned}$$

em que $\mu_1, \mu_2^C, \mu_2^T, d_{i1}, d_{i2}^C$ e d_{i2}^T estão definidos em (3.27).

O correspondente modelo misto para populações finitas com EMEx é

$$\tilde{\mathbf{Z}} = \text{vec} \left[\begin{pmatrix} \tilde{\mathbf{Y}}_1 & \tilde{\mathbf{Y}}_2^C & \tilde{\mathbf{Y}}_2^T \end{pmatrix} \right] = (\mathbf{I}_3 \otimes \mathbf{1}_N) \boldsymbol{\mu} + \mathbf{d} + \tilde{\mathbf{E}} \quad (3.32)$$

em que $\boldsymbol{\mu}$ e \mathbf{d} estão definidos em (3.28), e

$$\tilde{\mathbf{E}} = \left[\begin{pmatrix} \tilde{\mathbf{E}}_1 \end{pmatrix}' \quad \begin{pmatrix} \tilde{\mathbf{E}}_2^C \end{pmatrix}' \quad \begin{pmatrix} \tilde{\mathbf{E}}_2^T \end{pmatrix}' \right]'$$

representa o vetor de EMEx.

O vetor de médias e a matriz de covariâncias de $\tilde{\mathbf{Z}}$ apresentado em (3.32) são dados por (ver Apêndice A, pág. 85)

$$\mathbb{E}_{SR}(\tilde{\mathbf{Z}}) = (\mathbf{I}_3 \otimes \mathbf{1}_N) \boldsymbol{\mu}$$

e

$$\text{Var}_{SR}(\tilde{\mathbf{Z}}) = \text{diag} \left\{ \left(\bigoplus_{i=1}^N \tilde{\sigma}_{i1}^2 \right), \left(\bigoplus_{i=1}^N (\tilde{\sigma}_{i2}^C)^2 \right), \left(\bigoplus_{i=1}^N (\tilde{\sigma}_{i2}^T)^2 \right) \right\} + \boldsymbol{\Sigma} \otimes \mathbf{P}_N,$$

com $\boldsymbol{\mu}$ definida em (3.28) e $\boldsymbol{\Sigma}$, em (3.30).

3.2.1 Combinações lineares de interesse e preditores ótimos

Sob os modelos (3.28) e (3.32), i.e., quando consideramos uma população finita não estratificada com dois tipos de intervenção, nosso interesse é prever combinações lineares da forma

$$T = \sum_{i=1}^N (c_{i1}^C Y_{i1} + c_{i2}^C Y_{i2}^C) + \sum_{i=1}^N (c_{i1}^T Y_{i1} + c_{i2}^T Y_{i2}^T) \quad (3.33)$$

em que os c_{i1}^C , c_{i2}^C , c_{i1}^T , e c_{i2}^T são constantes conhecidas associadas aos rótulos i (posições). Por exemplo, se o interesse recai no efeito da intervenção C ,

$$T_C = \sum_{i=1}^N (N^{-1} Y_{i2}^C - N^{-1} Y_{i1}), \quad (3.34)$$

tomamos $c_{i2}^C = N^{-1}$, $c_{i1}^C = -N^{-1}$ e $c_{i2}^T = c_{i1}^T = 0$, $i = 1, \dots, N$. Se o interesse recai no efeito da intervenção T ,

$$T_T = \sum_{i=1}^N (N^{-1} Y_{i2}^T - N^{-1} Y_{i1}), \quad (3.35)$$

tomamos $c_{i2}^T = N^{-1}$, $c_{i1}^T = -N^{-1}$ e $c_{i2}^C = c_{i1}^C = 0$, $i = 1, \dots, N$. Para avaliar a diferença dos efeitos das intervenções C e T ,

$$T_D = \sum_{i=1}^N (N^{-1} Y_{i2}^C - N^{-1} Y_{i1}) - \sum_{i=1}^N (N^{-1} Y_{i2}^T - N^{-1} Y_{i1}) \quad (3.36)$$

adotamos $c_{i1}^T = c_{i2}^C = N^{-1}$ e $c_{i2}^T = c_{i1}^C = -N^{-1}$, $i = 1, \dots, N$.

Para obter o estimador/preditor de T , com base numa amostra aleatória simples selecionada sem reposição, assumimos, sem perda de generalidade, que

- i) as primeiras n posições na permutação aleatória, i.e., $i = 1, \dots, n$, correspondem às unidades selecionadas sem reposição numa amostra aleatória simples de tamanho n da população finita,
- ii) nas primeiras n_C unidades selecionadas, i.e., $i = 1, \dots, n_C (< n)$, observamos, direta ou indiretamente, Y_{i1} e Y_{i2}^C porém não observamos Y_{i2}^T , porquanto vamos supor, por exemplo, que unicamente as primeiras n_C unidades são submetidas à intervenção C , e
- iii) das restantes $n - n_C (= n_T)$ unidades selecionadas, i.e., $i = n_C + 1, \dots, n$, observamos, direta ou indiretamente, Y_{i1} e Y_{i2}^T porém não observamos Y_{i2}^C , porque vamos supor, por exemplo, que

somente as últimas n_T unidades, das n seleccionadas na amostra, são submetidas à intervenção T .

Desta forma, T pode ser decomposta como

$$T = T^{(1)} + T^{(2)} \quad (3.37)$$

com

$$T^{(1)} = \sum_{i=1}^{n_C} (c_{i1}^C Y_{i1} + c_{i2}^C Y_{i2}^C) + \sum_{i=n_C+1}^n (c_{i1}^T Y_{i1} + c_{i2}^T Y_{i2}^T)$$

e

$$T^{(2)} = \sum_{i=n_C+1}^N (c_{i1}^C Y_{i1} + c_{i2}^C Y_{i2}^C) + \sum_{i=1}^{n_C} (c_{i1}^T Y_{i1} + c_{i2}^T Y_{i2}^T) + \sum_{i=n+1}^N (c_{i1}^T Y_{i1} + c_{i2}^T Y_{i2}^T).$$

Portanto, observamos $T^{(1)}$, diretamente (sem erro de medida) ou indiretamente (com erros de medida) a partir dos dados amostrais, e temos que prever $T^{(2)}$.

De forma semelhante ao procedimento seguido na Seção 2.2 e na Subseção 3.1.1, o estimador/preditor de T obtém-se utilizando métodos padrão para encontrar preditores lineares não viciados de variância mínima. Como a diferença entre o preditor sob os dois modelos (3.28) e (3.32) se deve à variância dos EMEn e dos EMEx, respectivamente, vamos considerar apenas a predição sob o modelo (3.28) e, posteriormente, vamos particularizar os resultados sob o modelo (3.32).

Neste caso, a matriz utilizada para pré-multiplicar Z^* apresentada em (3.28), para decompor nas partes amostrada e não amostrada, é

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}_I \\ \mathbf{K}_{II} \end{pmatrix} \quad (3.38)$$

em que

$$\mathbf{K}_I = \begin{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \otimes (\mathbf{I}_{n_C+n_T} \mathbf{0}_{(n_C+n_T) \times (N-n_C-n_T)}) \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \otimes (\mathbf{I}_{n_C} \mathbf{0}_{n_C \times (N-n_C)}) \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \otimes (\mathbf{0}_{n_T \times n_C} \mathbf{I}_{n_T} \mathbf{0}_{n_T \times (N-n_C-n_T)}) \end{pmatrix}$$

e

$$\mathbf{K}_{II} = \begin{pmatrix} \mathbf{I}_3 & \otimes & (\mathbf{0}_{(N-n_C-n_T) \times (n_C+n_T)} \quad \mathbf{I}_{N-n_C-n_T}) \\ \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} & \otimes & \begin{pmatrix} \mathbf{0}_{n_T \times n_C} & \mathbf{I}_{n_T} & \mathbf{0}_{n_T \times (N-n_C-n_T)} \\ \mathbf{I}_{n_C} & \mathbf{0}_{n_C \times (N-n_C)} \end{pmatrix} \end{pmatrix}.$$

Observação: No caso em que $N = n_C + n_T$, as matrizes \mathbf{K}_I e \mathbf{K}_{II} se podem escrever como

$$\mathbf{K}_{I,N} = \begin{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} & \otimes & \mathbf{I}_N \\ \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} & \otimes & \begin{pmatrix} \mathbf{I}_{n_C} & \mathbf{0}_{n_C \times n_T} \\ \mathbf{0}_{n_T \times n_C} & \mathbf{I}_{n_T} \end{pmatrix} \end{pmatrix} \text{ e } \mathbf{K}_{II,N} = \begin{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} & \otimes & \begin{pmatrix} \mathbf{0}_{n_T \times n_C} & \mathbf{I}_{n_T} \\ \mathbf{I}_{n_C} & \mathbf{0}_{n_C \times n_T} \end{pmatrix} \end{pmatrix}.$$

Denotando

$$\begin{pmatrix} \mathbf{Z}_I^* \\ \mathbf{Z}_{II}^* \end{pmatrix} = \begin{pmatrix} \mathbf{K}_I \mathbf{Z}^* \\ \mathbf{K}_{II} \mathbf{Z}^* \end{pmatrix},$$

segue que

$$\mathbf{Z}_I^* = \left(Y_{11}^* \quad \cdots \quad Y_{(n_C+n_T),1}^* \quad Y_{12}^{C*} \quad \cdots \quad Y_{n_C,2}^{C*} \quad Y_{(n_C+1),2}^{T*} \quad \cdots \quad Y_{(n_C+n_T),2}^{T*} \right)'$$

e

$$\mathbf{Z}_{II}^* = \left(Y_{(n_C+n_T+1),1}^* \quad \cdots \quad Y_{N1}^* \quad Y_{(n_C+n_T+1),2}^{C*} \quad \cdots \quad Y_{N2}^{C*} \quad Y_{(n_C+n_T+1),2}^{T*} \quad \cdots \quad Y_{N,2}^{T*} \quad Y_{(n_C+1),2}^{C*} \quad \cdots \quad Y_{(n_C+n_T),2}^{C*} \quad Y_{12}^{T*} \quad \cdots \quad Y_{n_C,2}^{T*} \right)'. \quad (3.39)$$

Devido à complexidade das estruturas das matrizes de covariâncias de \mathbf{Z}_I^* e \mathbf{Z}_{II}^* , e de seu uso na procura do preditor linear não viciado de variância mínima, neste trabalho assumimos que o número de unidades submetidas à intervenção C é igual ao número de unidades submetidas à intervenção T , i.e., $n_C = n_T = n_0$, além disso, assumimos que

$$\begin{aligned} \sigma_{s1}^2 &= (\sigma_{s2}^C)^2 = (\sigma_{s2}^T)^2, \quad s = 1, \dots, N, \quad \text{sob o modelo (3.28)} \quad \text{e} \\ \tilde{\sigma}_{i1}^2 &= (\tilde{\sigma}_{i2}^C)^2 = (\tilde{\sigma}_{i2}^T)^2, \quad i = 1, \dots, N, \quad \text{sob o modelo (3.32)}. \end{aligned} \quad (3.40)$$

Para considerar o problema de número de unidades submetidas à intervenção C diferente ao número de unidades submetidas à intervenção T , pode-se usar um modelo com variáveis expandidas para controlar a dimensão dos vetores e matrizes amostradas [ver Stanek & Singer (2008)].

Portanto,

$$\mathbf{K}_I = \left(\begin{array}{c} \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right) \otimes \left(\begin{array}{cc} \mathbf{I}_{2n_0} & \mathbf{0}_{2n_0 \times (N-2n_0)} \\ \mathbf{I}_{n_0} & \mathbf{0}_{n_0 \times (N-n_0)} \end{array} \right) \\ \left(\begin{array}{ccc} 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right) \otimes \left(\begin{array}{ccc} \mathbf{0}_{n_0 \times n_0} & \mathbf{I}_{n_0} & \mathbf{0}_{n_0 \times (N-2n_0)} \end{array} \right) \end{array} \right),$$

e

$$\mathbf{K}_{II} = \left(\begin{array}{c} \mathbf{I}_3 \otimes \left(\begin{array}{cc} \mathbf{0}_{(N-2n_0) \times 2n_0} & \mathbf{I}_{N-2n_0} \end{array} \right) \\ \left(\begin{array}{ccc} 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right) \otimes \left(\begin{array}{ccc} \mathbf{0}_{n_0 \times n_0} & \mathbf{I}_{n_0} & \mathbf{0}_{n_0 \times (N-2n_0)} \end{array} \right) \\ \left(\begin{array}{ccc} 0 & 0 & 1 \end{array} \right) \otimes \left(\begin{array}{cc} \mathbf{I}_{n_0} & \mathbf{0}_{n_0 \times (N-n_0)} \end{array} \right) \end{array} \right),$$

logo

$$\begin{aligned} \mathbb{E}_{SR}(\mathbf{Z}_I^*) &= [\mathbf{H}_0 \otimes \mathbf{1}_{n_0}] \boldsymbol{\mu}, \\ \mathbb{E}_{SR}(\mathbf{Z}_{II}^*) &= \mathbf{H}_1 \boldsymbol{\mu}, \\ \text{Var}_{SR}(\mathbf{Z}_I^*) &= \mathbf{V}_I^* = \bar{\sigma}_1^2 \mathbf{I}_{4n_0} + (\boldsymbol{\Sigma}^a \otimes \mathbf{I}_{n_0}) - N^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_{n_0}) \boldsymbol{\Sigma}^b (\mathbf{I}_4 \otimes \mathbf{1}_{n_0})', \\ \text{Var}_{SR}(\mathbf{Z}_{II}^*) &= \mathbf{V}_{II}^* = \bar{\sigma}_1^2 \mathbf{I}_{3N-4n_0} + n_0^{-1} \mathbf{H}_2' (\mathbf{I}_4 \otimes \mathbf{1}_{n_0}) \boldsymbol{\Sigma}^a (\mathbf{I}_4 \otimes \mathbf{1}_{n_0})' \mathbf{H}_2 \\ &\quad + \text{diag} \left\{ \boldsymbol{\Sigma} \otimes \mathbf{P}_{N-2n_0}, \left(\begin{array}{cc} \sigma_{22}^{CC} & 0 \\ 0 & \sigma_{22}^{TT} \end{array} \right) \otimes \mathbf{P}_{n_0} \right\} - N^{-1} \mathbf{H}_1 \boldsymbol{\Sigma} \mathbf{H}_1' \\ &\quad + (N-2n_0)^{-1} \left(\begin{array}{cc} \mathbf{I}_3 \otimes \mathbf{1}_{N-2n_0} \\ \mathbf{0}_{2n_0 \times 3} \end{array} \right) \boldsymbol{\Sigma} (\mathbf{I}_3 \otimes \mathbf{1}'_{N-2n_0} | \mathbf{0}_{3 \times 2n_0}), \\ \text{Cov}_{SR}(\mathbf{Z}_I^*, \mathbf{Z}_{II}^*) &= \mathbf{V}_{I,II}^* = (\boldsymbol{\Sigma}^a \otimes \mathbf{I}_{n_0}) \mathbf{H}_3 - N^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_{n_0}) \mathbf{H}_0 \boldsymbol{\Sigma} \mathbf{H}_1' \end{aligned} \quad (3.41)$$

com

$$\begin{aligned} \boldsymbol{\Sigma}^a &= \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array} \right) \boldsymbol{\Sigma} \left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{array} \right) \otimes \mathbf{e}_1 \mathbf{e}_1' + \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 0 & 1 \end{array} \right) \boldsymbol{\Sigma} \left(\begin{array}{cc} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{array} \right) \otimes \mathbf{e}_2 \mathbf{e}_2', \\ \boldsymbol{\Sigma}^b &= \mathbf{H}_0 \boldsymbol{\Sigma} \mathbf{H}_0', \\ \mathbf{H}_0 &= \left(\begin{array}{ccc} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right), \end{aligned} \quad (3.42)$$

$$\mathbf{H}_1 = \begin{pmatrix} \mathbf{I}_3 & \otimes \mathbf{1}_{N-2n_0} \\ \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} & \otimes \mathbf{1}_{n_0} \end{pmatrix},$$

$$\mathbf{H}_2 = \left[\mathbf{0}_{4n_0 \times 3(N-2n_0)} \left| \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \mathbf{I}_{n_0} \right. \right],$$

$$\mathbf{H}_3 = \{[(\boldsymbol{\Sigma}^a)^{-1} \boldsymbol{\Sigma}^b - \mathbf{I}_4] \otimes \mathbf{I}_{n_0}\} \mathbf{H}_2,$$

\mathbf{P}_a , \mathbf{e}_i , $\boldsymbol{\mu}$, $\bar{\sigma}_1^2$ e $\boldsymbol{\Sigma}$ definidos em (2.5), (3.17), (3.28), (3.29) e (3.30), respectivamente.

Ilustramos a notação com uma população composta por $N = 7$ unidades da qual selecionamos uma amostra aleatória simples de tamanho $n = 6$ sem reposição. As primeiras $n_0 = 3$ unidades selecionadas são submetidas à intervenção C enquanto as restantes $n_0 = n - n_0 = 3$ unidades são submetidas à intervenção T . Seja $w_{s1} = w_{s2}^C = w_{s2}^T = 1$ para todo s ,

$$\begin{aligned} \mathbf{y}_1 &= (3.5 \quad 7 \quad 1.5 \quad 4 \quad 6 \quad 5.5 \quad 9)', \\ \mathbf{y}_2^C &= (12.25 \quad 13.5 \quad 7.75 \quad 10.5 \quad 15 \quad 12.25 \quad 18.5)', \\ \mathbf{y}_2^T &= (6.75 \quad 9.5 \quad 8.25 \quad 9.5 \quad 13 \quad 10.25 \quad 18.5)', \\ \mathbf{E}_1 &= (E_{11} \quad E_{21} \quad E_{31} \quad E_{41} \quad E_{51} \quad E_{61} \quad E_{71})', \\ \mathbf{E}_2^C &= (E_{12}^C \quad E_{22}^C \quad E_{32}^C \quad E_{42}^C \quad E_{52}^C \quad E_{62}^C \quad E_{72}^C)', \\ \mathbf{E}_2^T &= (E_{12}^T \quad E_{22}^T \quad E_{32}^T \quad E_{42}^T \quad E_{52}^T \quad E_{62}^T \quad E_{72}^T)' \end{aligned}$$

e suponhamos que o erro de medida endógeno, tanto no pré-teste quanto no pós-teste, possa tomar apenas dois valores igualmente prováveis, dados por mais ou menos $\sigma_{s1} [= \sigma_{s2}^C = \sigma_{s2}^T]$, com $\sigma_{11} = 1$, $\sigma_{21} = 3$, $\sigma_{31} = 10$, $\sigma_{41} = 7$, $\sigma_{51} = 6$, $\sigma_{61} = 5$ e $\sigma_{71} = 0.05$. Por conseguinte, $\mu_1 = 5.214$, $\mu_2^C = 12.821$, $\mu_2^T = 10.821$, $\bar{\sigma}_1^2 = 31.429 [= \bar{\sigma}_2^{C2} = \bar{\sigma}_2^{T2}]$ e

$$\boldsymbol{\Sigma} = \begin{pmatrix} 6.071 & 7.815 & 7.732 \\ 7.815 & 11.536 & 11.036 \\ 7.732 & 11.036 & 15.119 \end{pmatrix}.$$

Portanto,

$$\mathbb{E}_{SR}(\mathbf{Z}_I^*) = [\mathbf{H}_0 \otimes \mathbf{1}_3] \begin{pmatrix} 5.214 \\ 12.821 \\ 10.821 \end{pmatrix},$$

$$\mathbb{E}_{SR}(\mathbf{Z}_{II}^*) = \mathbf{H}_1 \begin{pmatrix} 5.214 \\ 12.821 \\ 10.821 \end{pmatrix},$$

$$\mathbf{V}_I^* = 31.429\mathbf{I}_{12} + (\boldsymbol{\Sigma}^a \otimes \mathbf{I}_3) - \frac{1}{7}(\mathbf{I}_4 \otimes \mathbf{1}_3) \boldsymbol{\Sigma}^b (\mathbf{I}_4 \otimes \mathbf{1}_3)',$$

$$\begin{aligned} \mathbf{V}_{II}^* &= 31.429\mathbf{I}_9 + \frac{1}{3}\mathbf{H}_2' (\mathbf{I}_4 \otimes \mathbf{1}_3) \boldsymbol{\Sigma}^a (\mathbf{I}_4 \otimes \mathbf{1}_3)' \mathbf{H}_2 \\ &\quad + \text{diag} \left\{ \left(1 - \frac{1}{1}\right) \begin{pmatrix} 6.071 & 7.815 & 7.732 \\ 7.815 & 11.536 & 11.036 \\ 7.732 & 11.036 & 15.119 \end{pmatrix}, \begin{pmatrix} 11.536 & 0 \\ 0 & 15.119 \end{pmatrix} \otimes \mathbf{P}_3 \right\} \\ &\quad - \frac{1}{7}\mathbf{H}_1 \begin{pmatrix} 6.071 & 7.815 & 7.732 \\ 7.815 & 11.536 & 11.036 \\ 7.732 & 11.036 & 15.119 \end{pmatrix} \mathbf{H}_1' + \begin{pmatrix} \mathbf{I}_3 \\ \mathbf{0}_{6 \times 3} \end{pmatrix} \begin{pmatrix} 6.071 & 7.815 & 7.732 \\ 7.815 & 11.536 & 11.036 \\ 7.732 & 11.036 & 15.119 \end{pmatrix} (\mathbf{I}_3 | \mathbf{0}_{3 \times 6}), \end{aligned}$$

$$\mathbf{V}_{I,II}^* = (\boldsymbol{\Sigma}^a \otimes \mathbf{I}_3) \mathbf{H}_3 - \frac{1}{7}(\mathbf{I}_4 \otimes \mathbf{1}_3) \mathbf{H}_0 \begin{pmatrix} 6.071 & 7.815 & 7.732 \\ 7.815 & 11.536 & 11.036 \\ 7.732 & 11.036 & 15.119 \end{pmatrix} \mathbf{H}_1'$$

com

$$\boldsymbol{\Sigma}^a = \begin{pmatrix} 6.071 & 0 & 7.815 & 0 \\ 0 & 6.071 & 0 & 7.732 \\ 7.815 & 0 & 11.536 & 0 \\ 0 & 7.732 & 0 & 15.119 \end{pmatrix}, \quad \boldsymbol{\Sigma}^b = \begin{pmatrix} 6.071 & 6.071 & 7.815 & 7.732 \\ 6.071 & 6.071 & 7.815 & 7.732 \\ 7.815 & 7.815 & 11.536 & 11.036 \\ 7.732 & 7.732 & 11.036 & 15.119 \end{pmatrix},$$

$$\mathbf{H}_1 = \begin{pmatrix} \mathbf{I}_3 \\ \left(\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \otimes \mathbf{1}_3 \right) \end{pmatrix}, \quad \mathbf{H}_2 = \left[\mathbf{0}_{12 \times 3} \left| \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \mathbf{I}_3 \right. \right]$$

$$\text{e } \mathbf{H}_3 = \{ [(\boldsymbol{\Sigma}^a)^{-1} \boldsymbol{\Sigma}^b - \mathbf{I}_4] \otimes \mathbf{I}_3 \} \mathbf{H}_2.$$

Se a realização da matriz aleatória U é

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

temos

$$\begin{aligned} \mathbf{Y}_1^*_{\{U=\mathbf{u}\}} &= (6 + E_{51}, 9 + E_{71}, 7 + E_{21}, 1.5 + E_{31}, 3.5 + E_{11}, 5.5 + E_{61}, 4 + E_{41})', \\ \mathbf{Y}_2^*_{\{U=\mathbf{u}\}} &= (15 + E_{52}^C, 18.5 + E_{72}^C, 13.5 + E_{22}^C, 7.75 + E_{32}^C, 12.25 + E_{12}^C, 12.25 + E_{62}^C, \\ &\quad 10.5 + E_{42}^C)', \\ \mathbf{Y}_2^{T*}_{\{U=\mathbf{u}\}} &= (13 + E_{52}^T, 18.5 + E_{72}^T, 9.5 + E_{22}^T, 8.25 + E_{32}^T, 6.75 + E_{12}^T, 10.25 + E_{62}^T, \\ &\quad 9.5 + E_{42}^T)', \\ \mathbf{Z}_{I\{U=\mathbf{u}\}}^* &= (6 + E_{51}, 9 + E_{71}, 7 + E_{21}, 1.5 + E_{31}, 3.5 + E_{11}, 5.5 + E_{61}, 15 + E_{52}^C, \\ &\quad 18.5 + E_{72}^C, 13.5 + E_{22}^C, 8.25 + E_{32}^T, 6.75 + E_{12}^T, 10.25 + E_{62}^T)', \\ \mathbf{Z}_{II\{U=\mathbf{u}\}}^* &= (4 + E_{41}, 10.5 + E_{42}^C, 9.5 + E_{42}^T, 7.75 + E_{32}^C, 12.25 + E_{12}^C, 12.25 + E_{62}^C, \\ &\quad 13 + E_{52}^T, 18.5 + E_{72}^T, 9.5 + E_{22}^T)'. \end{aligned}$$

Ao minimizar a esperança do erro quadrático médio do preditor linear sujeito à restrição de que seja não viciado [ver Apêndice D, ítem b), pág. 101], obtemos

$$\hat{T}^* = \left[\mathbf{D}' (\mathbf{H}'_4)^{-1} \mathbf{H}'_0 (\boldsymbol{\Sigma}^a)^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_{n_0})' + \mathbf{g}'_I + \mathbf{g}'_{II} \mathbf{H}'_3 \right] \mathbf{A}^{-1} \mathbf{Z}_I^* \quad (3.43)$$

em que

$$\begin{aligned} \mathbf{g}_I &= ([c_{11}^C + c_{11}^T], \dots, [c_{2n_0,1}^C + c_{2n_0,1}^T], c_{12}^C, \dots, c_{n_0,2}^C, c_{n_0+1,2}^T, \dots, c_{2n_0,2}^T)', \\ \mathbf{g}_{II} &= ([c_{2n_0+1,1}^C + c_{2n_0+1,1}^T], \dots, [c_{N1}^C + c_{N1}^T], c_{2n_0+1,2}^C, \dots, c_{N2}^C, c_{2n_0+1,2}^T, \dots, c_{N2}^T, \\ &\quad c_{n_0+1,2}^C, \dots, c_{2n_0,2}^C, c_{12}^T, \dots, c_{n_0,2}^T)', \\ \mathbf{A} &= [\mathbf{I}_4 + \bar{\sigma}_1^2 (\boldsymbol{\Sigma}^a)^{-1}] \otimes \mathbf{I}_{n_0}, \\ \mathbf{H}_4 &= \mathbf{H}'_0 (\boldsymbol{\Sigma}^a)^{-1} \mathbf{G} \mathbf{H}_0, \end{aligned}$$

$$\begin{aligned}\mathbf{G} &= (\mathbf{I}_4 \otimes \mathbf{1}_{n_0})' \mathbf{A}^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_{n_0}), \\ \mathbf{D} &= \mathbf{H}'_0 (\mathbf{I}_4 \otimes \mathbf{1}_{n_0})' [(\mathbf{I}_{4n_0} - \mathbf{A}^{-1}) \mathbf{g}_I - \mathbf{A}^{-1} \mathbf{H}_3 \mathbf{g}_{II}] + \mathbf{H}'_1 \mathbf{g}_{II},\end{aligned}$$

$\bar{\sigma}_1^2 [= \bar{\sigma}_2^{C2} = \bar{\sigma}_2^{T2}]$ está definido em (3.29) e \mathbf{Z}_I^* , em (3.39); c_{i1}^C , c_{i1}^T , c_{i2}^C e c_{i2}^T estão definidos em (3.33) e Σ^a , \mathbf{H}_0 , \mathbf{H}_1 e \mathbf{H}_3 , em (3.42).

A variância do preditor \hat{T}^* é (ver Apêndice D, pág. 102)

$$\begin{aligned}\text{Var}_{SR}(\hat{T}^* - T) &= \mathbf{D}' (\mathbf{H}'_4)^{-1} \mathbf{D} + \mathbf{g}'_I (\Sigma^a \otimes \mathbf{I}_{n_0}) (\mathbf{I}_{4n_0} - \mathbf{A}^{-1}) \mathbf{g}_I \\ &\quad - \mathbf{g}'_{II} \mathbf{H}'_3 \mathbf{A}^{-1} (\Sigma^a \otimes \mathbf{I}_{n_0}) \mathbf{H}_3 \mathbf{g}_{II} \\ &\quad + 2\mathbf{g}'_I (\Sigma^a \otimes \mathbf{I}_{n_0}) (\mathbf{I}_{4n_0} - \mathbf{A}^{-1}) \mathbf{H}_3 \mathbf{g}_{II} \\ &\quad + \mathbf{g}'_{II} \text{diag} \left\{ \Sigma \otimes \mathbf{P}_{N-2n_0}, \begin{pmatrix} \sigma_{22}^{CC} & 0 \\ 0 & \sigma_{22}^{TT} \end{pmatrix} \otimes \mathbf{P}_{n_0} \right\} \mathbf{g}_{II} \\ &\quad + (N - 2n_0)^{-1} \mathbf{g}'_{II} \begin{pmatrix} \mathbf{I}_3 \otimes \mathbf{1}_{N-2n_0} \\ \mathbf{0}_{2n_0 \times 3} \end{pmatrix} \Sigma (\mathbf{I}_3 \otimes \mathbf{1}'_{N-2n_0} | \mathbf{0}_{3 \times 2n_0}) \mathbf{g}_{II} \\ &\quad + n_0^{-1} \mathbf{g}'_{II} \mathbf{H}'_2 (\mathbf{I}_4 \otimes \mathbf{1}_{n_0}) \Sigma^a (\mathbf{I}_4 \otimes \mathbf{1}_{n_0})' \mathbf{H}_2 \mathbf{g}_{II}\end{aligned} \quad (3.44)$$

com \mathbf{P}_a e \mathbf{H}_2 definidos em (2.5) e (3.42), respectivamente, e σ_{22}^{hh} , $h = C, T$, e Σ , em (3.30).

O correspondente preditor sob o modelo (3.32) é

$$\tilde{T} = \left[\mathbf{D}' (\mathbf{H}'_4)^{-1} \mathbf{H}'_0 (\Sigma^a)^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_{n_0})' + \mathbf{g}'_I + \mathbf{g}'_{II} \mathbf{H}'_3 \right] \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{Z}}_I \quad (3.45)$$

em que

$$\begin{aligned}\tilde{\mathbf{Z}}_I &= \left(\tilde{Y}_{11} \quad \dots \quad \tilde{Y}_{2n_0,1} \quad \tilde{Y}_{12}^C \quad \dots \quad \tilde{Y}_{n_0,2}^C \quad \tilde{Y}_{(n_0+1),2}^T \quad \dots \quad \tilde{Y}_{2n_0,2}^T \right)', \\ \tilde{\mathbf{A}} &= \mathbf{I}_{4n_0} + \left[\mathbf{I}_2 \otimes \bigoplus_{i=1}^{2n_0} \tilde{\sigma}_{i1}^2 \right] [(\Sigma^a)^{-1} \otimes \mathbf{I}_{n_0}],\end{aligned}$$

$\tilde{\sigma}_{i1}^2 [= (\tilde{\sigma}_{i2}^C)^2 = (\tilde{\sigma}_{i2}^T)^2]$ definido em (3.31), \mathbf{g}_I e \mathbf{g}_{II} definidos em (3.43), Σ^a , \mathbf{H}_0 e \mathbf{H}_3 , em (3.42); e \mathbf{H}_4 , \mathbf{G} e \mathbf{D} , em (3.43) com a substituição de \mathbf{A} por $\tilde{\mathbf{A}}$.

A variância do preditor \tilde{T} definido em (3.45) é

$$\begin{aligned}
\text{Var}_{SR}(\tilde{T} - T) &= \mathbf{D}'(\mathbf{H}'_4)^{-1}\mathbf{D} + \mathbf{g}'_I(\Sigma^a \otimes \mathbf{I}_{n_0})\left(\mathbf{I}_{4n_0} - \tilde{\mathbf{A}}^{-1}\right)\mathbf{g}_I \\
&\quad - \mathbf{g}'_{II}\mathbf{H}'_3\tilde{\mathbf{A}}^{-1}(\Sigma^a \otimes \mathbf{I}_{n_0})\mathbf{H}_3\mathbf{g}_{II} \\
&\quad + 2\mathbf{g}'_I(\Sigma^a \otimes \mathbf{I}_{n_0})\left(\mathbf{I}_{4n_0} - \tilde{\mathbf{A}}^{-1}\right)\mathbf{H}_3\mathbf{g}_{II} \\
&\quad + \mathbf{g}'_{II}\text{diag}\left\{\Sigma \otimes \mathbf{P}_{N-2n_0}, \begin{pmatrix} \sigma_{22}^{CC} & 0 \\ 0 & \sigma_{22}^{TT} \end{pmatrix} \otimes \mathbf{P}_{n_0}\right\}\mathbf{g}_{II} \\
&\quad + (N - 2n_0)^{-1}\mathbf{g}'_{II}\begin{pmatrix} \mathbf{I}_3 \otimes \mathbf{1}_{N-2n_0} \\ \mathbf{0}_{2n_0 \times 3} \end{pmatrix}\Sigma(\mathbf{I}_3 \otimes \mathbf{1}'_{N-2n_0} | \mathbf{0}_{3 \times 2n_0})\mathbf{g}_{II} \\
&\quad + n_0^{-1}\mathbf{g}'_{II}\mathbf{H}'_2(\mathbf{I}_4 \otimes \mathbf{1}_{n_0})\Sigma^a(\mathbf{I}_4 \otimes \mathbf{1}_{n_0})'\mathbf{H}_2\mathbf{g}_{II}
\end{aligned} \tag{3.46}$$

com σ_{22}^{CC} e σ_{22}^{TT} definidos em (3.30) e \mathbf{H}_2 , em (3.42).

3.2.2 Exemplo

Nesta Subsecção tomamos $w_{s1} = w_{s2}^C = w_{s2}^T = 1$ para todo s porque o nosso interesse recai na estimação de T_D (3.36), i.e., na diferença dos efeitos das intervenções C e T , que é função das respostas latentes não ponderadas. Consideremos os dados do Exemplo 1.2. Como no desenvolvimento do preditor de T assumimos $n_C = n_T = n_0$ e no Exemplo 1.2, $n_C = 15$ e $n_T = 14$, desconsideramos, para efeito de ilustração, uma observação (por exemplo, a última) da intervenção com ar sintético (C). Na Tabela 3.3 estão apresentadas medidas resumo correspondentes.

Tabela 3.3: Medidas resumo ($\times 10^{-3}$) para os dados de Exemplo 1.2.

Intervenção	Ocasão	Média Amostral	Variância Amostral	Mínimo	Máximo
Ar Sintético/Ar Sintético	pré-teste	63.6	1.0	15.7	120.3
	pós-teste	50.5	1.2	2.1	117.1
Ar Sintético/Hélio-Oxigênio	pré-teste	39.2	0.9	2.5	90.9
	pós-teste	24.8	0.6	0.0	97.2

Neste caso, temos $n = 28$ e vamos supor $N = 28, 50, 100, 300, 500$ e 1000 . Para o caso com EMEn, suponhamos que $\sigma_{s1}^2 = (\sigma_{s2}^C)^2 = (\sigma_{s2}^T)^2 = 0.1s \{2N/[N(N+1)]\}$, $s = 1, \dots, N$, e para o caso com EMEx, suponhamos que $\tilde{\sigma}_{i1}^2 = (\tilde{\sigma}_{i2}^C)^2 = (\tilde{\sigma}_{i2}^T)^2 = 0.1i \{2n/[n(n+1)]\}$ para $i \leq n$ e $\tilde{\sigma}_{i1}^2 = (\tilde{\sigma}_{i2}^C)^2 = (\tilde{\sigma}_{i2}^T)^2 = 0$ para $i > n$ ^d. Também é considerado o caso sem erro de medida, para

^dNote que o erro de medida muda de uma unidade (respectivamente, posição) a outra, mas, num caso é ligado à unidade e no outro caso é ligado à posição (na amostra) e é tal que a variância média dos EMEn sobre todas as unidades

o qual a estimativa de T_D (3.36)  rotulada por \hat{T}_D . Sob cada um dos casos, o valor da estimativa de T_D , da varincia estimada e do intervalo de confiana aproximado, baseado na desigualdade de Chebyshev, com um nvel de confiana de $1 - \alpha = 0.90$, so apresentados na Tabela 3.4. A estimativa obtida sob o modelo (1.4) por meio de tcnicas usuais para anlise de medidas repetidas [ver Bonate (2000)], que corresponde  diferena entre as estimativas dos parmetros μ_C e μ_T , e a estimativa obtida sob o modelo (1.6) via anlise de covarincia [ver Bonate (2000, p. 93)], que corresponde a duas vezes a estimativa do parmetro τ_C , i.e.,

$$2\hat{\tau}_C = 2 \left[\bar{Y}_{I,2}^{C*} - (2n_0)^{-1} \left(n_0 \bar{Y}_{I,2}^{C*} + n_0 \bar{Y}_{I,2}^{T*} \right) - \hat{\beta} \left(\bar{Y}_{I,1}^{C*} - \bar{Y}_{I,1}^{T*} \right) \right] \quad (3.47)$$

com $\bar{Y}_{I,2}^{C*}$, $\bar{Y}_{I,2}^{T*}$, $\bar{Y}_{I,1}^{C*}$ e $\bar{Y}_{I,1}^{T*}$ definidos mais para a frente em (3.48) e $\hat{\beta}$ a estimativa do parmetro β sob o modelo (1.6), :

$$\hat{\mu}_C - \hat{\mu}_T = 0.0505 - 0.0248 = 25.7 \times 10^{-3} \quad \text{com} \quad \widehat{Var}(\hat{\mu}_C - \hat{\mu}_T) = 2 \times 0.008369^2 = 1.4 \times 10^{-3}$$

e

$$2\hat{\tau}_C = 10.5 \times 10^{-3} \quad \text{com} \quad \widehat{Var}(2\hat{\tau}_C) = 0.1 \times 10^{-3},$$

respectivamente. Quando temos erros de medida, observamos que varincia estimada de \hat{T}_D^*  maior que a varincia estimada de \tilde{T}_D^* . Quando no temos erros de medida, assumindo N grande, observamos que

$$\widehat{Var}_s(\hat{T}_D) < \widehat{Var}(2\hat{\tau}_C) < \widehat{Var}(\hat{\mu}_C - \hat{\mu}_T).$$

Alm disso, observamos que o tamanho da populao (N) gerou um pequeno efeito na varincia dos estimadores.

3.2.3 Estudo por simulao dos estimadores propostos

Para comparar o desempenho dos estimadores \hat{T}^* proposto em (3.43) e \tilde{T} proposto em (3.45) com o desempenho do estimador do modelo (1.6) obtido via mnimos quadrados ordinrios, aqui denotado por $2\hat{\tau}_C$ (3.47), realizamos um estudo de simulao em que so considerados diferentes valores para o coeficiente de correlao entre os valores latentes no pr-teste e os valores (*potenciais*) latentes populacionais no ps-teste (intervenes C e T), $n = 10$ e $n = 30$, e duas formas da distribuo das variveis respostas, a saber:

 igual  varincia mdia dos EMEx sobre as unidades na amostra, i.e., $N^{-1} \sum_{s=1}^N \sigma_{s1}^2 = 0.1$, e $n^{-1} \sum_{i=1}^n \tilde{\sigma}_{i1}^2 = 0.1$.

Tabela 3.4: Estimativas ($\times 10^{-3}$) da diferença dos efeitos, na resistência respiratória de ratos, da ventilação com uma mistura de Oxigênio e Hélio e da ventilação com ar sintético, variância estimada ($\times 10^{-3}$) e intervalos de confiança ($\times 10^{-3}$), considerando erros de medida endógenos, erros de medida exógenos e sem erros de medida.

N	\hat{T}_D^*	$\widehat{Var}_{SR}(\hat{T}_D^*)$	$\hat{T}_D^* \pm a$	\tilde{T}_D^*	$\widehat{Var}_{SR}(\tilde{T}_D^*)$	$\tilde{T}_D^* \pm a$	\hat{T}_d	$\widehat{Var}_s(\hat{T}_d)$	$\hat{T}_d \pm a$
28	24.4	1.5	24.4 \pm 122.0	20.9	1.3	20.9 \pm 115.8	12.2	0.02	12.2 \pm 13.8
50	24.4	1.5	24.4 \pm 123.2	20.9	1.4	20.9 \pm 117.0	12.2	0.05	12.2 \pm 21.8
100	24.4	1.5	24.4 \pm 123.9	20.9	1.4	20.9 \pm 117.8	12.2	0.07	12.2 \pm 25.7
300	24.4	1.5	24.4 \pm 124.4	20.9	1.4	20.9 \pm 118.3	12.2	0.08	12.2 \pm 27.9
500	24.4	1.6	24.4 \pm 124.5	20.9	1.4	20.9 \pm 118.4	12.2	0.08	12.2 \pm 28.4
1000	24.4	1.6	24.4 \pm 124.6	20.9	1.4	20.9 \pm 118.5	12.2	0.08	12.2 \pm 28.7

$$a = \sqrt{\widehat{Var}_{SR}(\cdot)/\alpha}, \alpha = 0.1$$

- i) uma distribuição normal trivariada com $\mathbb{E}(X) = \mathbb{E}(Y) = \mathbb{E}(Z) = 0$, $Var(X) = Var(Y) = Var(Z) = 1$, $Cov(X, Y) = \rho_{21}^C = 0.2$, $Cov(X, Z) = \rho_{21}^T = 0.2$, e, $Cov(Y, Z) = 0.2$, em que os valores gerados a partir de X , Y e Z correspondem os valores latentes populacionais no pré-teste, y_{s1} , no pós-teste sob a intervenção C , y_{s2}^C , e no pós-teste sob a intervenção T , y_{s2}^T , respectivamente, e,
- ii) uma distribuição normal trivariada como a referida no ítem i), em que seguidamente, substituímos os valores obtidos de y_{s2}^h , $h = C, T$, pela transformação $y_{s2,0}^h = [y_{s2}^h/\sigma_{22}^{hh}] e^{[y_{s2}^h/\sigma_{22}^{hh}]}$, i.e., usamos $y_{s2,0}^h$ como a resposta (potencial) pós-teste sob a intervenção h . σ_{22}^{hh} está definida em (3.30). Isto, para considerar populações assimétricas no pós-teste.

Consideramos EMEn com variâncias iguais e diferentes. Também, consideramos EMEx com variâncias diferentes. Como \hat{T}^* definido em (3.43) é igual a \tilde{T} definido em (3.45) quando os erros de medida são homocedásticos, não consideramos o caso com variâncias iguais para os EMEx. Apresentamos a seguir os passos usados para o estudo de simulação em que consideramos EMEn com variâncias iguais e dados gerados conforme o ítem i):

1. Geramos $N = 300$ vetores de tamanho 3 conforme o ítem i), produzindo desta forma y_{s1} , y_{s2}^C e y_{s2}^T , $s = 1, \dots, N$.
2. Obtivemos 10000 amostras independentes de tamanho $n = 20 (= 2n_0)$ via amostragem aleatória simples da população gerada no ítem 1, de forma que cada amostra contém $n = 20$ vetores da forma $(Y_{i1} \ Y_{i2}^C \ Y_{i2}^T)$. Tomamos Y_{i1} , $i = 1, \dots, n$, como os valores latentes no pré-teste, Y_{i2}^C , $i = 1, \dots, n_0 (= 10)$, como os valores latentes pós-teste sob a intervenção C e Y_{i2}^T , $i = n_0 + 1, \dots, n$, como os valores latentes pós-teste sob a intervenção T . Dessa forma cada

uma das 10000 amostras (de valores latentes) é formada pelo conjunto de pares (Y_{i1}, Y_{i2}^C) , $i = 1, \dots, n_0 (= 10)$ e (Y_{i1}, Y_{i2}^T) , $i = n_0 + 1, \dots, n (= 20)$.

3. A cada Y_{i1} , $i = 1, \dots, n$, Y_{i2}^C , $i = 1, \dots, n_0$, e Y_{i2}^T , $i = n_0 + 1, \dots, n$, obtido em 2, foi somado, independentemente, o EMEn, gerado a partir de uma distribuição normal com média 0 e desvio padrão $\bar{\sigma}_1 = \sqrt{0.01}$, obtendo Y_{i1}^* , $i = 1, \dots, n$, Y_{i2}^{C*} , $i = 1, \dots, n_0$ e Y_{i2}^{T*} , $i = n_0 + 1, \dots, n$.
4. Avaliamos tanto o estimador \hat{T}^* proposto em (3.43) quando Σ (3.30) é conhecida e como quando é substituída por uma estimativa amostral. No último caso rotulamos o estimador como \hat{T}_e^* (estimador empírico). A estimativa de Σ usada é

$$\hat{\Sigma} = \begin{pmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{21}^C & \hat{\sigma}_{21}^T \\ \hat{\sigma}_{21}^C & \hat{\sigma}_{22}^{CC} & 0 \\ \hat{\sigma}_{21}^T & 0 & \hat{\sigma}_{22}^{TT} \end{pmatrix} \quad (3.48)$$

em que

$$\begin{aligned} \hat{\sigma}_{11} &= (n-1)^{-1} \sum_{i=1}^n (Y_{i1}^* - \bar{Y}_{I,1}^*)^2, & \bar{Y}_{I,1}^* &= n^{-1} \sum_{i=1}^n Y_{i1}^*, \\ \hat{\sigma}_{22}^{CC} &= (n_0-1)^{-1} \sum_{i=1}^{n_0} (Y_{i2}^{C*} - \bar{Y}_{I,2}^{C*})^2, & \bar{Y}_{I,2}^{C*} &= n_0^{-1} \sum_{i=1}^{n_0} Y_{i2}^{C*}, \\ \hat{\sigma}_{22}^{TT} &= (n_0-1)^{-1} \sum_{i=n_0+1}^n (Y_{i2}^{T*} - \bar{Y}_{I,2}^{T*})^2, & \bar{Y}_{I,2}^{T*} &= n_0^{-1} \sum_{i=n_0+1}^n Y_{i2}^{T*}, \\ \hat{\sigma}_{21}^C &= (n_0-1)^{-1} \sum_{i=1}^{n_0} (Y_{i1}^* - \bar{Y}_{I,1}^*) (Y_{i2}^{C*} - \bar{Y}_{I,2}^{C*}), & \bar{Y}_{I,1}^{C*} &= n_0^{-1} \sum_{i=1}^{n_0} Y_{i1}^*, \\ \hat{\sigma}_{21}^T &= (n_0-1)^{-1} \sum_{i=n_0+1}^n (Y_{i1}^* - \bar{Y}_{I,1}^*) (Y_{i2}^{T*} - \bar{Y}_{I,2}^{T*}), & \bar{Y}_{I,1}^{T*} &= n_0^{-1} \sum_{i=n_0+1}^n Y_{i1}^*, \end{aligned}$$

e $\hat{\sigma}_{22}^{CT} = 0$. Também, avaliamos a diferença entre os efeitos médios amostrais das intervenções,

$$\bar{T} = \bar{Y}_{I,2}^{C*} - \bar{Y}_{I,1}^{C*} - (\bar{Y}_{I,2}^{T*} - \bar{Y}_{I,1}^{T*}) \quad (3.49)$$

e a estimativa obtida via mínimos quadrados ordinários obtida sob o modelo (1.6), dada em (3.47), para cada uma das 10000 amostras.

5. As médias das 10000 estimativas de \hat{T}^* , \hat{T}_e^* , \bar{T} e $2\hat{\tau}_c$ foram calculadas. Na Tabela E.1 (pág. 106) estão dispostos tanto a contribuição relativa do vício ao EQM quanto o EQM.

6. Os passos 1 a 5 foram repetidos considerando diferentes combinações de ρ_{21}^C , ρ_{21}^T e $\bar{\sigma}_1^2$. Os valores considerados foram

$$\begin{aligned} n_0 &= 10 \text{ e } 30, & \bar{\sigma}_1^2 &= 0, 0.01, 0.016 \text{ e } 0.09, \\ \rho_{21}^C &= 0.2, 0.5 \text{ e } 0.8, & \rho_{21}^T &= 0.2 \text{ e } 0.5, \end{aligned}$$

Os passos 1 a 6 foram repetidos para populações geradas conforme o ítem ii). Os resultados estão dispostos na Tabela E.2 (pág. 107). Das Tabelas E.1 e E.2, observamos que

- \hat{T}^* apresenta o menor EQM,
- quando n_0 aumenta de 10 para 30, o EQM diminui,
- sob populações geradas conforme o ítem i),

$$\text{EQM}(\hat{T}^*) < \text{EQM}(\hat{T}_e^*) \leq \text{EQM}(2\hat{\tau}_c) < \text{EQM}(\bar{T}),$$

exceto quando $0.7 < \rho_{21}^C < 0.9$ e $0.4 < \rho_{21}^T < 0.6$ em que

$$\text{EQM}(\hat{T}^*) < \text{EQM}(2\hat{\tau}_c) \leq \text{EQM}(\hat{T}_e^*) < \text{EQM}(\bar{T}),$$

- sob populações geradas conforme o ítem ii),

$$\text{EQM}(\hat{T}_e^*) \leq \text{EQM}(2\hat{\tau}_c) < \text{EQM}(\hat{T}^*) < \text{EQM}(\bar{T}),$$

quando $0.40 < \rho_{21}^C < 0.60$ e $0.25 < \rho_{21}^T < 0.50$, e para 7 das 40 situações restantes, nas outras 33

$$\text{EQM}(\hat{T}^*) < \text{EQM}(\bar{T}) < \text{EQM}(\hat{T}_e^*) \leq \text{EQM}(2\hat{\tau}_c).$$

- o vício dos estimadores \hat{T}_e^* , \bar{T} e $2\hat{\tau}_c$ é tal que sua contribuição ao EQM é menor do que 1%, enquanto o estimador \hat{T}^* é não viciado.

No segundo estudo de simulação, consideramos variâncias diferentes para os EMEn geradas como $\sigma_{s1}^2 = (0.0001s)^2$, $s = 1, \dots, N$; como consequência, $\bar{\sigma}_1^2 = 0.0003$. Para isto, no passo 2, além de selecionar os valores latentes pré-teste e pós-teste, selecionamos as variâncias dos EMEn associadas às unidades na amostra, i.e., cada uma das 10000 amostras é formada pelo conjunto de ternos $(Y_{i1}, Y_{i2}^C, \sigma_{i1}^2)$, $i = 1, \dots, n_0 (= 10)$, e $(Y_{i1}, Y_{i2}^T, \sigma_{i1}^2)$, $i = n_0 + 1, \dots, n (= 20)$, em que $\sigma_{i1}^2 = \sum_{s=1}^N U_{is} \sigma_{s1}^2$. Em seguida, no passo 3, a cada Y_{i1} , $i = 1, \dots, n$, Y_{i2}^C , $i = 1, \dots, n_0$, e Y_{i2}^T ,

$i = n_0 + 1, \dots, n$, foi somado, independentemente, o EMEn gerado a partir de uma distribuição normal com média 0 e desvio padrão σ_{i1} , obtendo Y_{i1}^* , $i = 1, \dots, n$, Y_{i2}^{C*} , $i = 1, \dots, n_0$, e Y_{i2}^{T*} , $i = n_0 + 1, \dots, n$. Nas Tabelas E.3 e E.4 (pág. 108) estão dispostos a contribuição relativa do vício ao EQM e o EQM de \hat{T}^* (3.43), de \hat{T}_e^* que utiliza a estimativa amostral, $\hat{\Sigma}$ (3.48), de \bar{T} (3.49) e de $2\hat{\tau}_c$ (3.47). Observamos que

- sob populações geradas conforme o ítem i),

$$\text{EQM}(\hat{T}^*) < \text{EQM}(\hat{T}_e^*) < \text{EQM}(2\hat{\tau}_c) < \text{EQM}(\bar{T}),$$

- sob populações geradas conforme o ítem ii),

$$\text{EQM}(\hat{T}^*) \leq \min\{\text{EQM}(\hat{T}_e^*), \text{EQM}(\bar{T})\} \leq \text{EQM}(2\hat{\tau}_c).$$

No terceiro estudo de simulação, consideramos variâncias diferentes $[\tilde{\sigma}_{i1}^2 = (0.0001i)^2, i = 1, \dots, n]$ para os EMEx. Para isto, no passo 3, a cada Y_{i1} , $i = 1, \dots, n$, Y_{i2}^C , $i = 1, \dots, n_0$, e Y_{i2}^T , $i = n_0 + 1, \dots, n$, foi somado, independentemente, o EMEx gerado a partir de uma distribuição normal com média 0 e desvio padrão $\tilde{\sigma}_{i1}$, obtendo \tilde{Y}_{i1} , $i = 1, \dots, n$, \tilde{Y}_{i2}^C , $i = 1, \dots, n_0$, e \tilde{Y}_{i2}^T , $i = n_0 + 1, \dots, n$. Nas Tabelas E.5 e E.6 (pág. 109) estão dispostos a contribuição relativa do vício ao EQM e o EQM de \tilde{T} (3.45), de \tilde{T}_e , de \bar{T} (3.49) e de $2\hat{\tau}_c$ (3.47), em que \tilde{T} é a estimativa quando Σ é conhecido e \tilde{T}_e é a estimativa quando utilizamos a estimativa amostral de Σ [ver (3.48)]. Como no segundo estudo de simulação, observamos que

- sob populações geradas conforme o ítem i),

$$\text{EQM}(\tilde{T}) < \text{EQM}(\tilde{T}_e) < \text{EQM}(2\hat{\tau}_c) < \text{EQM}(\bar{T}),$$

- sob populações geradas conforme o ítem ii),

$$\text{EQM}(\tilde{T}) \leq \min\{\text{EQM}(\tilde{T}_e), \text{EQM}(\bar{T})\} \leq \text{EQM}(2\hat{\tau}_c).$$

3.3 Discussão

No Capítulo 3, mostramos como os resultados de Stanek, Lencina, Singer, González, Li & San Martino (2008) podem ser estendidos para acomodar erros de medida endógenos e exógenos, sob o

modelo misto para populações finitas, no contexto de estudos do tipo pré-teste/pós-teste. Para isso, apresentamos dois cenários. No primeiro, consideramos uma população finita com dois estratos cujas unidades são submetidas a uma intervenção, e no segundo, consideramos uma população finita não estratificada que é submetida a dois tipos de intervenção. Os estimadores propostos (3.19) e (3.23), sob o primeiro cenário, e (3.43) e (3.45), sob o segundo cenário, podem ser empregados tanto para estimar uma combinação de totais populacionais pré-teste/pós-teste, quanto para prever combinações lineares de valores latentes (pré-teste/pós-teste) de uma unidade na j -ésima posição numa permutação aleatória da população. Como no Capítulo 2, os estimadores são obtidos por meio de métodos padrão para encontrar preditores lineares não viciados de variância mínima.

Sob o primeiro cenário, i.e., quando consideramos uma população estratificada e uma intervenção, no caso específico em que o interesse recai na diferença de efeitos da intervenção entre os dois estratos e não tivermos erro de medida, o estimador proposto em (3.19) é igual ao estimador obtido por meio de técnicas usuais para análise de medidas repetidas via mínimos quadrados ordinários ($\hat{T} = \bar{Y}_{I,12} - \bar{Y}_{I,11} + \bar{Y}_{I,22} - \bar{Y}_{I,21}$), em que $\bar{Y}_{I,ht} = n^{-1} \sum_{i=1}^n Y_{hit}$. Se o interesse é uma das médias populacionais de um dos dois estratos no pré-teste, ou no pós-teste, e tivermos erro de medida endógenos, o estimador (3.19) é igual a (2.25) e se não tivermos erro de medida, é igual a (2.28).

Não obstante as nossas suposições sejam mais fracas que as suposições usualmente adotadas em análises de medidas repetidas ou análises de covariância, e ainda que seja necessário a estimação das variâncias e das covariâncias populacionais, os estimadores propostos em (3.43) e em (3.45), e suas estimativas empíricas apresentaram melhor desempenho, i.e., menor EQM e ausência de vício, ou contribuição relativa do vício ao EQM menor do que 1%. Em particular, quando se tem uma população finita com distribuição simétrica, as simulações comprovam a otimalidade dos estimadores propostos (3.43) e (3.45), e mostram um pobre desempenho do estimador obtido por meio de técnicas usuais para análise de medidas repetidas via mínimos quadrados ordinários (3.49). Por outra parte, sob populações finitas com distribuição assimétrica no pós-teste, as simulações mostram um pobre desempenho do estimador obtido via análise de covariância.

Nos exemplos, quando o interesse recaiu na diferença de efeitos da intervenção, observamos que os estimadores propostos em (3.19), (3.23), (3.43) e (3.45), dado n , não se alteraram quando N aumentou, enquanto suas variâncias apresentaram pequenos aumentos.

Sob o segundo cenário, uma alternativa para calcular \hat{T}^* quando Σ (3.30) é desconhecida é

considerar usar as variâncias amostrais do pré-teste em separado para cada intervenção, i.e.,

$$\hat{\Sigma}^a = \begin{pmatrix} \hat{\sigma}_{11}^C & \mathbf{0} & \hat{\sigma}_{21}^C & \mathbf{0} \\ \mathbf{0} & \hat{\sigma}_{11}^T & \mathbf{0} & \hat{\sigma}_{21}^T \\ \hat{\sigma}_{21}^C & \mathbf{0} & \hat{\sigma}_{22}^{CC} & 0 \\ \mathbf{0} & \hat{\sigma}_{21}^T & 0 & \hat{\sigma}_{22}^{TT} \end{pmatrix} \quad \text{e} \quad \hat{\Sigma}^b = \mathbf{H}_0 \hat{\Sigma} \mathbf{H}_0'$$

em que $\hat{\sigma}_{11}^C = (n_0 - 1)^{-1} \sum_{i=1}^{n_0} (Y_{i1}^* - \bar{Y}_{I,1}^{C*})^2$, $\hat{\sigma}_{11}^T = (n_0 - 1)^{-1} \sum_{i=n_0+1}^n (Y_{i1}^* - \bar{Y}_{I,1}^{T*})^2$, e $\bar{Y}_{I,1}^{C*}$, $\bar{Y}_{I,1}^{T*}$, $\hat{\sigma}_{21}^C$, $\hat{\sigma}_{21}^T$, $\hat{\sigma}_{22}^{CC}$, $\hat{\sigma}_{22}^{TT}$ e $\hat{\Sigma}$ estão definidas em (3.48). Neste caso, a estimativa \hat{T}^* (3.43), quando não tivermos erro de medida, coincide com a estimativa via análise de covariância (3.47).

Capítulo 4

Considerações finais

4.1 Conclusões

Neste trabalho, discutimos a aplicação de modelos mistos para populações finitas a problemas de estimação/previsão de parâmetros da população finita quando se dispõe de informação auxiliar e quando se considera estudos do tipo pré-teste/pós-teste, levando em conta possíveis erros de medida tanto endógenos quanto exógenos.

Os estimadores propostos (2.18), (2.20) e (3.19) são particularizados para o caso em que o interesse recai na média ou combinações lineares das médias populacionais, segundo seja o caso, e os estimadores propostos (2.18), (2.20), (3.19), (3.23), (3.43) e (3.45) são comparados com alguns dos estimadores propostos na teoria não só sob o enfoque baseado no planejamento, mas também sob o enfoque de super-população e sob o enfoque de populações infinitas.

Os estudos de simulação comprovaram a otimalidade dos estimadores propostos, os pontos fortes e as debilidades de seus respectivos estimadores empíricos, assim como aqueles dos estimadores tradicionalmente usados. Entre os pontos fortes dos estimadores empíricos propostos está um erro quadrático médio menor que dos competidores considerados e uma contribuição relativa do vício ao erro quadrático médio menor do 1% para \hat{T}_e^* , o estimador empírico de \hat{T}^* (3.43), e para \tilde{T}_e , o estimador empírico de \tilde{T} (3.45), e entre suas debilidades, o vício do estimador \hat{B}_e^* (2.41) e do estimador \tilde{B}_e (2.42).

4.2 Pesquisa Futura

No trabalho, foi considerada a inclusão de erro de medida endógeno e exógeno separadamente, porém, os dois tipos de erro de medida podem-se apresentar num mesmo problema. A generalização para este caso não é trivial, assim como para o caso em que consideramos erros endógenos correlacionados, por exemplo, quando um entrevistador obtém informação de várias pessoas.

Outro tema que ainda fica aberto é a determinação de intervalos de confiança para os estimadores propostos. Apesar de que podemos considerar a desigualdade de Chebyshev para uma primeira aproximação, é de interesse estudar as distribuições assintóticas destes estimadores.

Além disso, no trabalho foi considerada amostragem aleatória simples sem reposição, porém a extensão do modelo misto para populações finitas para planos amostrais complexos é de grande interesse.

Também, no caso de estudos do tipo pré-teste/pós-teste, é de interesse considerar diferentes tamanhos dos estratos e das amostras dentro dos estratos (Seção 3.1), e diferentes tamanhos das amostras para a aplicação das intervenções (Seção 3.2), assim como considerar a inclusão de informação auxiliar. Não consideramos estes casos pela dificuldade para encontrar a inversa da matriz de covariâncias da parte amostrada, i.e., $Var_{SR}(\mathbf{Z}_I^*)$ com \mathbf{Z}_I^* definida em (3.15) e $Var_{SR}(\mathbf{Z}_I^*)$ com \mathbf{Z}_I^* definida em (3.39).

Finalmente, considerar dados longitudinais, assim como considerar dados faltantes, seria interessante.

Apêndice A

Cálculo do vetor de médias e da matriz de covariâncias dos modelos propostos nos Capítulos 2 e 3

Cálculo do vetor de médias e da matriz de covariâncias de Z^* definido em (2.8)

Para calcular o vetor de médias e a matriz de covariâncias de Z^* definido em (2.8) usamos o vetor de médias e a matriz de covariâncias de U em (2.3), nomeadamente

$$\mathbb{E}_S(U) = N^{-1}J_N \quad \text{e} \quad \text{Var}_S[\text{vec}(U)] = (N-1)^{-1}(P_N \otimes P_N), \quad (\text{A.1})$$

com J_a e P_a definidos em (2.5). Então,

$$\begin{aligned} \mathbb{E}_{SR}(Z^*) &= \mathbb{E}_S\{\text{vec}[U(\mathbf{y} \ \mathbf{x})]\} + \mathbb{E}_{SR}\left\{\text{vec}\left[U\begin{pmatrix} \mathbf{E} & \mathbf{0} \\ & \mathbf{1}_{N \times 1} \end{pmatrix}\right]\right\} \\ &= \mathbb{E}_S\{\text{vec}[U(\mathbf{y} \ \mathbf{x})]\} = (\mathbf{I}_2 \otimes \mathbf{1}_N)(\mu_y \ \mu_x)' \end{aligned}$$

porque $\mathbb{E}_{SR}(U_{is}E_s) = \mathbb{E}_S(U_{is})\mathbb{E}_R(E_s) = 0$, e

$$\begin{aligned} \text{Var}_{SR}(\mathbf{Z}^*) &= \text{Var}_{SR} \{ \text{vec} [\mathbf{U}(\mathbf{y} \ \mathbf{x})] + \text{vec} [\mathbf{U}(\mathbf{E} \ \mathbf{0})] \} \\ &= \mathbb{E}_S \left(\text{Var}_{R|S} \{ \text{vec} [\mathbf{U}(\mathbf{y} \ \mathbf{x})] + \text{vec} [\mathbf{U}(\mathbf{E} \ \mathbf{0})] \} \right) \\ &\quad + \text{Var}_S \left(\mathbb{E}_{R|S} \{ \text{vec} [\mathbf{U}(\mathbf{y} \ \mathbf{x})] + \text{vec} [\mathbf{U}(\mathbf{E} \ \mathbf{0})] \} \right), \end{aligned}$$

em que

$$\text{Var}_{R|S} \{ \text{vec} [\mathbf{U}(\mathbf{y} \ \mathbf{x})] + \text{vec} [\mathbf{U}(\mathbf{E} \ \mathbf{0})] \} = \mathbf{e}_1 \mathbf{e}'_1 \otimes \left[\mathbf{U} \left(\bigoplus_{s=1}^N \sigma_s^2 \right) \mathbf{U}' \right]$$

e

$$\mathbb{E}_{R|S} \{ \text{vec} [\mathbf{U}(\mathbf{y} \ \mathbf{x})] + \text{vec} [\mathbf{U}(\mathbf{E} \ \mathbf{0})] \} = \text{vec} [\mathbf{U}(\mathbf{y} \ \mathbf{x})],$$

com \mathbf{e}_i definida em (2.10). Logo

$$\text{Var}_{SR}(\mathbf{Z}^*) = \mathbf{e}_1 \mathbf{e}'_1 \otimes \bar{\sigma}_e^2 \mathbf{I}_N + \Sigma \otimes \mathbf{P}_N$$

com Σ definido em (2.6), e

$$\bar{\sigma}_e^2 = \mathbb{E}_S \left(\sum_{s=1}^N \sigma_s^2 U_{is} \right) = N^{-1} \sum_{s=1}^N \sigma_s^2. \quad (\text{A.2})$$

Cálculo do vetor de médias e da matriz de covariâncias de $\tilde{\mathbf{Z}}$ definido em (2.12)

Usando (A.1) temos

$$\begin{aligned} \mathbb{E}_{SR}(\tilde{\mathbf{Z}}) &= \mathbb{E}_S \{ \text{vec} [\mathbf{U}(\mathbf{y} \ \mathbf{x})] \} + \mathbb{E}_R \left\{ \text{vec} \left[\begin{pmatrix} \tilde{\mathbf{E}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{N \times 1} \end{pmatrix} \right] \right\} \\ &= (\mathbf{I}_2 \otimes \mathbf{1}_N) (\mu_y \ \mu_x)' \end{aligned}$$

e

$$\begin{aligned} \text{Var}_{SR}(\tilde{\mathbf{Z}}) &= \text{Var}_S \{ \text{vec} [\mathbf{U}(\mathbf{y} \ \mathbf{x})] \} + \text{Var}_R \left\{ \text{vec} \left[\begin{pmatrix} \tilde{\mathbf{E}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{N \times 1} \end{pmatrix} \right] \right\} \\ &= \Sigma \otimes \mathbf{P}_N + \mathbf{e}_1 \mathbf{e}'_1 \otimes \bigoplus_{i=1}^N \tilde{\sigma}_i^2, \end{aligned}$$

com \mathbf{P}_a , Σ e e_i definidos em (2.5), (2.6) e (2.10), respectivamente.

Cálculo do vetor de médias e da matriz de covariâncias de \mathbf{Z}^* definido em (3.4)

Para o cálculo do vetor de médias e a matriz de covariâncias de \mathbf{Z}^* definido em (3.4) usamos

$$\mathbb{E}_S \left(\text{diag} \left\{ \mathbf{U}^{(1)}, \mathbf{U}^{(2)} \right\} \right) = \text{diag} \left\{ N_1^{-1} \mathbf{J}_{N_1}, N_2^{-1} \mathbf{J}_{N_2} \right\} \quad (\text{A.3})$$

e

$$\text{Var}_S \left(\text{vec} \left[\text{diag} \left\{ \mathbf{U}^{(1)}, \mathbf{U}^{(2)} \right\} \right] \right) = \text{diag} \left\{ (N_1 - 1)^{-1} \mathbf{P}_{N_1} \otimes \text{diag} \left\{ \mathbf{P}_{N_1}, \mathbf{0}_{N_2 \times N_2} \right\}, \right. \\ \left. (N_2 - 1)^{-1} \mathbf{P}_{N_2} \otimes \text{diag} \left\{ \mathbf{0}_{N_1 \times N_1}, \mathbf{P}_{N_2} \right\} \right\}.$$

Então

$$\begin{aligned} \mathbb{E}_{SR} (\mathbf{Z}^*) &= E_S \left\{ \text{vec} \left[\text{diag} \left\{ \mathbf{U}^{(1)}, \mathbf{U}^{(2)} \right\} \begin{pmatrix} \mathbf{y}_1^{(1)} & \mathbf{y}_2^{(1)} \\ \mathbf{y}_1^{(2)} & \mathbf{y}_2^{(2)} \end{pmatrix} \right] \right\} \\ &\quad + \mathbb{E}_{SR} \left\{ \text{vec} \left[\text{diag} \left\{ \mathbf{U}^{(1)}, \mathbf{U}^{(2)} \right\} \begin{pmatrix} \mathbf{E}_1^{(1)} & \mathbf{E}_2^{(1)} \\ \mathbf{E}_1^{(2)} & \mathbf{E}_2^{(2)} \end{pmatrix} \right] \right\} \\ &= \text{vec} \left[\begin{pmatrix} \mu_1^{(1)} \mathbf{1}_{N_1} & \mu_2^{(1)} \mathbf{1}_{N_1} \\ \mu_1^{(2)} \mathbf{1}_{N_2} & \mu_2^{(2)} \mathbf{1}_{N_2} \end{pmatrix} \right] \\ &= \left[\mathbf{I}_2 \otimes \begin{pmatrix} \mathbf{1}_{N_1} \\ \mathbf{0}_{N_2 \times 1} \end{pmatrix} \right] \begin{pmatrix} \mu_1^{(1)} \\ \mu_2^{(1)} \end{pmatrix} + \left[\mathbf{I}_2 \otimes \begin{pmatrix} \mathbf{0}_{N_1 \times 1} \\ \mathbf{1}_{N_2} \end{pmatrix} \right] \begin{pmatrix} \mu_1^{(2)} \\ \mu_2^{(2)} \end{pmatrix}, \end{aligned}$$

com $\mu_t^{(h)}$ definido em (3.3), porquanto

$$\mathbb{E}_{SR} \left(U_{is}^{(h)} E_{hst} \right) = \mathbb{E}_S \left(U_{is}^{(h)} \right) \mathbb{E}_R \left(E_{hst} \right) = 0, \quad \text{para } t = 1, 2 \text{ e } h = 1, 2,$$

e

$$\text{Var}_{SR} (\mathbf{Z}^*) = \mathbb{E}_S \left\{ \text{Var}_{R|S} (\mathbf{Z}^*) \right\} + \text{Var}_S \left\{ \mathbb{E}_{R|S} (\mathbf{Z}^*) \right\}$$

em que

$$\mathbb{E}_{R|S}(\mathbf{Z}^*) = \text{vec} \left[\text{diag} \{ \mathbf{U}^{(1)}, \mathbf{U}^{(2)} \} \begin{pmatrix} \mathbf{y}_1^{(1)} & \mathbf{y}_2^{(1)} \\ \mathbf{y}_1^{(2)} & \mathbf{y}_2^{(2)} \end{pmatrix} \right]$$

e

$$\text{Var}_{R|S}(\mathbf{Z}^*) = \left[\mathbf{I}_2 \otimes \text{diag} \{ \mathbf{U}^{(1)}, \mathbf{U}^{(2)} \} \right] \times \text{Var}_{R|S} \left\{ \text{vec} \left[\begin{pmatrix} \mathbf{E}_1^{(1)} & \mathbf{E}_2^{(1)} \\ \mathbf{E}_1^{(2)} & \mathbf{E}_2^{(2)} \end{pmatrix} \right] \right\} \left[\mathbf{I}_2 \otimes \text{diag} \{ \mathbf{U}^{(1)}, \mathbf{U}^{(2)} \} \right]',$$

com

$$\text{Var}_{R|S} \left\{ \text{vec} \left[\begin{pmatrix} \mathbf{E}_1^{(1)} & \mathbf{E}_2^{(1)} \\ \mathbf{E}_1^{(2)} & \mathbf{E}_2^{(2)} \end{pmatrix} \right] \right\} = \text{diag} \left\{ \left(\bigoplus_{s=1}^{N_1} \sigma_{1s1}^2 \right), \left(\bigoplus_{s=1}^{N_2} \sigma_{2s1}^2 \right), \left(\bigoplus_{s=1}^{N_1} \sigma_{1s2}^2 \right), \left(\bigoplus_{s=1}^{N_2} \sigma_{2s2}^2 \right) \right\}.$$

Desta maneira,

$$\text{Var}_{SR}(\mathbf{Z}^*) = \text{diag} \{ \bar{\sigma}_{11}^2 \mathbf{I}_{N_1}, \bar{\sigma}_{21}^2 \mathbf{I}_{N_2}, \bar{\sigma}_{12}^2 \mathbf{I}_{N_1}, \bar{\sigma}_{22}^2 \mathbf{I}_{N_2} \} + \left[\boldsymbol{\Sigma}^{(1)} \otimes \text{diag} \left\{ \mathbf{P}_{N_1}, \mathbf{0}_{N_2 \times N_2} \right\} \right] + \left[\boldsymbol{\Sigma}^{(2)} \otimes \text{diag} \left\{ \mathbf{0}_{N_1 \times N_1}, \mathbf{P}_{N_2} \right\} \right]$$

com \mathbf{P}_a , $\bar{\sigma}_{ht}^2$ e $\boldsymbol{\Sigma}^{(h)}$ definidas em (2.5), (3.5) e (3.6), respectivamente.

Cálculo do vetor de médias e da matriz de covariâncias de $\tilde{\mathbf{Z}}$ definido em (3.8)

Usando (A.3) temos

$$\mathbb{E}_{SR}(\tilde{\mathbf{Z}}) = E_S \left\{ \text{vec} \left[\text{diag} \{ \mathbf{U}^{(1)}, \mathbf{U}^{(2)} \} \begin{pmatrix} \mathbf{y}_1^{(1)} & \mathbf{y}_2^{(1)} \\ \mathbf{y}_1^{(2)} & \mathbf{y}_2^{(2)} \end{pmatrix} \right] \right\} + \mathbb{E}_R \left\{ \text{vec} \left[\begin{pmatrix} \tilde{\mathbf{E}}_1^{(1)} & \tilde{\mathbf{E}}_2^{(1)} \\ \tilde{\mathbf{E}}_1^{(2)} & \tilde{\mathbf{E}}_2^{(2)} \end{pmatrix} \right] \right\}$$

$$= \begin{bmatrix} \mathbf{I}_2 \otimes \begin{pmatrix} \mathbf{1}_{N_1} \\ \mathbf{0} \\ \mathbf{1}_{N_2 \times 1} \end{pmatrix} \end{bmatrix} \begin{pmatrix} \mu_1^{(1)} \\ \mu_2^{(1)} \end{pmatrix} + \begin{bmatrix} \mathbf{I}_2 \otimes \begin{pmatrix} \mathbf{0} \\ \mathbf{1}_{N_1 \times 1} \\ \mathbf{1}_{N_2} \end{pmatrix} \end{bmatrix} \begin{pmatrix} \mu_1^{(2)} \\ \mu_2^{(2)} \end{pmatrix},$$

com $\mu_t^{(h)}$ definido em (3.3), e

$$\text{Var}_{SR}(\tilde{\mathbf{Z}}) = \mathbb{E}_S \left\{ \text{Var}_{R|S}(\tilde{\mathbf{Z}}) \right\} + \text{Var}_S \left\{ \mathbb{E}_{R|S}(\tilde{\mathbf{Z}}) \right\}$$

em que

$$\mathbb{E}_{R|S}(\tilde{\mathbf{Z}}) = \text{vec} \left[\text{diag} \left\{ \mathbf{U}^{(1)}, \mathbf{U}^{(2)} \right\} \begin{pmatrix} \mathbf{y}_1^{(1)} & \mathbf{y}_2^{(1)} \\ \mathbf{y}_1^{(2)} & \mathbf{y}_2^{(2)} \end{pmatrix} \right]$$

e

$$\text{Var}_{R|S}(\tilde{\mathbf{Z}}) = \text{diag} \left\{ \left(\bigoplus_{i=1}^{N_1} \tilde{\sigma}_{1i1}^2 \right), \left(\bigoplus_{i=1}^{N_2} \tilde{\sigma}_{2i1}^2 \right), \left(\bigoplus_{i=1}^{N_1} \tilde{\sigma}_{1i2}^2 \right), \left(\bigoplus_{i=1}^{N_2} \tilde{\sigma}_{2i2}^2 \right) \right\},$$

portanto

$$\begin{aligned} \text{Var}_{SR}(\tilde{\mathbf{Z}}) &= \text{diag} \left\{ \left(\bigoplus_{i=1}^{N_1} \tilde{\sigma}_{1i1}^2 \right), \left(\bigoplus_{i=1}^{N_2} \tilde{\sigma}_{2i1}^2 \right), \left(\bigoplus_{i=1}^{N_1} \tilde{\sigma}_{1i2}^2 \right), \left(\bigoplus_{i=1}^{N_2} \tilde{\sigma}_{2i2}^2 \right) \right\} \\ &\quad + \left[\Sigma^{(1)} \otimes \text{diag} \left\{ \mathbf{P}_{N_1}, \mathbf{0}_{N_2 \times N_2} \right\} \right] + \left[\Sigma^{(2)} \otimes \text{diag} \left\{ \mathbf{0}_{N_1 \times N_1}, \mathbf{P}_{N_2} \right\} \right], \end{aligned}$$

com \mathbf{P}_a e $\Sigma^{(h)}$ definidas em (2.5) e (3.6), respectivamente.

Cálculo do vetor de médias e da matriz de covariâncias de \mathbf{Z}^* definido em (3.28)

Usando (A.1) temos

$$\begin{aligned} \mathbb{E}_{SR}(\mathbf{Z}^*) &= \mathbb{E}_S \left\{ \text{vec} \left[\mathbf{U} \begin{pmatrix} \mathbf{y}_1 & \mathbf{y}_2^C & \mathbf{y}_2^T \end{pmatrix} \right] \right\} \\ &\quad + \mathbb{E}_{SR} \left\{ \text{vec} \left[\mathbf{U} \begin{pmatrix} \mathbf{E}_1 & \mathbf{E}_2^C & \mathbf{E}_2^T \end{pmatrix} \right] \right\} \\ &= (\mathbf{I}_3 \otimes \mathbf{1}_N) (\mu_1 \ \mu_2^C \ \mu_2^T)', \end{aligned}$$

porque

$$\begin{aligned}\mathbb{E}_{SR}(U_{is}E_{s1}) &= \mathbb{E}_S(U_{is})\mathbb{E}_R(E_{s1}) = 0, \\ \mathbb{E}_{SR}(U_{is}E_{s2}^C) &= \mathbb{E}_S(U_{is})\mathbb{E}_R(E_{s2}^C) = 0, \\ \mathbb{E}_{SR}(U_{is}E_{s2}^T) &= \mathbb{E}_S(U_{is})\mathbb{E}_R(E_{s2}^T) = 0,\end{aligned}$$

em que μ_1 , μ_2^C e μ_2^T estão definidos em (3.27), e

$$\text{Var}_{SR}(\mathbf{Z}^*) = \mathbb{E}_S \left\{ \text{Var}_{R|S}(\mathbf{Z}^*) \right\} + \text{Var}_S \left\{ \mathbb{E}_{R|S}(\mathbf{Z}^*) \right\}$$

em que

$$\mathbb{E}_{R|S}(\mathbf{Z}^*) = \text{vec} \left[\mathbf{U} \begin{pmatrix} \mathbf{y}_1 & \mathbf{y}_2^C & \mathbf{y}_2^T \end{pmatrix} \right]$$

e

$$\text{Var}_{R|S}(\mathbf{Z}^*) = (\mathbf{I}_3 \otimes \mathbf{U}) \text{Var}_{R|S} \left\{ \text{vec} \left[\begin{pmatrix} \mathbf{E}_1 & \mathbf{E}_2^C & \mathbf{E}_2^T \end{pmatrix} \right] \right\} (\mathbf{I}_3 \otimes \mathbf{U})',$$

com

$$\text{Var}_{R|S} \left\{ \text{vec} \left[\begin{pmatrix} \mathbf{E}_1 & \mathbf{E}_2^C & \mathbf{E}_2^T \end{pmatrix} \right] \right\} = \text{diag} \left\{ \left(\bigoplus_{s=1}^N \sigma_{s1}^2 \right), \left(\bigoplus_{s=1}^N (\sigma_{s2}^C)^2 \right), \left(\bigoplus_{s=1}^N (\sigma_{s2}^T)^2 \right) \right\}.$$

Portanto

$$\text{Var}_{SR}(\mathbf{Z}^*) = \text{diag} \left\{ \bar{\sigma}_1^2 \mathbf{I}_N, \bar{\sigma}_2^{C2} \mathbf{I}_N, \bar{\sigma}_2^{T2} \mathbf{I}_N \right\} + \mathbf{\Sigma} \otimes \mathbf{P}_N$$

com \mathbf{P}_a e $\mathbf{\Sigma}$ definidos em (2.5) e (3.30), respectivamente, e $\bar{\sigma}_1^2$, $\bar{\sigma}_2^{C2}$ e $\bar{\sigma}_2^{T2}$ definidos em (3.29).

Cálculo do vetor de médias e da matriz de covariâncias de $\tilde{\mathbf{Z}}$ definido em (3.32)

Usando (A.1) temos

$$\begin{aligned}\mathbb{E}_{SR}(\tilde{\mathbf{Z}}) &= \mathbb{E}_S \left\{ \text{vec} \left[\mathbf{U} \begin{pmatrix} \mathbf{y}_1 & \mathbf{y}_2^C & \mathbf{y}_2^T \end{pmatrix} \right] \right\} \\ &\quad + \mathbb{E}_R \left\{ \text{vec} \left[\begin{pmatrix} \tilde{\mathbf{E}}_1 & \tilde{\mathbf{E}}_2^C & \tilde{\mathbf{E}}_2^T \end{pmatrix} \right] \right\} \\ &= (\mathbf{I}_3 \otimes \mathbf{1}_N) (\mu_1 \ \mu_2^C \ \mu_2^T)',\end{aligned}$$

em que μ_1 , μ_2^C e μ_2^T estão definidos em (3.27), e

$$\text{Var}_{SR}(\tilde{\mathbf{Z}}) = \mathbb{E}_S \left\{ \text{Var}_{R|S}(\tilde{\mathbf{Z}}) \right\} + \text{Var}_S \left\{ \mathbb{E}_{R|S}(\tilde{\mathbf{Z}}) \right\},$$

em que

$$\mathbb{E}_{R|S}(\tilde{\mathbf{Z}}) = \text{vec} \left[\mathbf{U} \begin{pmatrix} \mathbf{y}_1 & \mathbf{y}_2^C & \mathbf{y}_2^T \end{pmatrix} \right]$$

e

$$\text{Var}_{R|S}(\tilde{\mathbf{Z}}) = \text{diag} \left\{ \left(\bigoplus_{i=1}^N \tilde{\sigma}_{i1}^2 \right), \left(\bigoplus_{i=1}^N (\tilde{\sigma}_{i2}^C)^2 \right), \left(\bigoplus_{i=1}^N (\tilde{\sigma}_{i2}^T)^2 \right) \right\}.$$

Portanto

$$\begin{aligned}\text{Var}_{SR}(\tilde{\mathbf{Z}}) &= \text{diag} \left\{ \left(\bigoplus_{i=1}^N \tilde{\sigma}_{i1}^2 \right), \left(\bigoplus_{i=1}^N (\tilde{\sigma}_{i2}^C)^2 \right), \left(\bigoplus_{i=1}^N (\tilde{\sigma}_{i2}^T)^2 \right) \right\} \\ &\quad + \Sigma \otimes \mathbf{P}_N,\end{aligned}$$

com \mathbf{P}_a definido em (2.5) e Σ , em (3.30).

Apêndice B

Obtenção do preditor de T (2.15) sob o modelo (2.8), da variância de \hat{T}^* (2.18) e da variância de \tilde{T} (2.20)

Obtenção do preditor de T (2.15) e de sua variância sob o modelo (2.8)

Para obter um estimador linear não viciado de variância mínima de $T = \sum_{i=1}^N c_i Y_i$, sob o modelo (2.8), consideramos as seguintes três condições:

- i) Deve ser uma função linear das variáveis aleatórias observadas na amostra, i.e.,

$$\hat{T}^* = (\mathbf{g}'_I + \mathbf{a}') \mathbf{Z}_I^* \quad (\text{B.1})$$

em que $\mathbf{g}'_I = \left(\mathbf{C}'_I \quad \mathbf{0}' \right)'$, com $\mathbf{C}_I = (c_1 \dots c_i \dots c_n)'$, e $\mathbf{Z}_I^* = [Y_1^* \dots Y_n^* (X_1 - \mu_x) \dots (X_n - \mu_x)]'$,

- ii) deve ser não viciado, i.e., $\mathbb{E}_{sR}([\mathbf{g}'_I + \mathbf{a}'] \mathbf{Z}_I^*) = \mathbb{E}_s(\mathbf{g}'_I \mathbf{Z}_I + \mathbf{g}'_{II} \mathbf{Z}_{II})$ em que

$$\begin{aligned} \mathbf{g}_{II} &= \left(\mathbf{C}'_{II} \quad \mathbf{0}' \right)'_{1 \times (N-n)}, \mathbf{C}_{II} = (c_{n+1} \dots c_i \dots c_N)', \\ \mathbf{Z}_I &= [Y_1 \dots Y_n (X_1 - \mu_x) \dots (X_n - \mu_x)]', \text{ e} \\ \mathbf{Z}_{II} &= [Y_{n+1} \dots Y_N (X_{n+1} - \mu_x) \dots (X_N - \mu_x)]'; \end{aligned} \quad (\text{B.2})$$

e

iii) deve minimizar a $Var_{SR}(\hat{T}^* - T)$.

Empregando multiplicadores de Lagrange, devemos encontrar a solução do sistema de equações dado por (B.3) e (B.4):

$$2[\mathbf{V}_I^* \hat{\mathbf{a}} + (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I - \mathbf{V}_{I,II} \mathbf{g}_{II}] = \lambda [\mathbf{e}_1 \otimes \mathbf{1}_n] \quad (\text{B.3})$$

e

$$[\mathbf{e}'_1 \otimes \mathbf{1}'_n] \hat{\mathbf{a}} - [\mathbf{e}'_1 \otimes \mathbf{1}'_{N-n}] \mathbf{g}_{II} = 0, \quad (\text{B.4})$$

em que

$$\begin{aligned} \mathbf{V}_I &= Var_S(\mathbf{Z}_I) = \Sigma \otimes \mathbf{P}_{n,N}, \\ \mathbf{V}_{I,II} &= Cov_S(\mathbf{Z}_I, \mathbf{Z}_{II}) = Cov_{SR}(\mathbf{Z}_I^*, \mathbf{Z}_{II}) = Cov_{SR}(\mathbf{Z}_I^*, \mathbf{Z}_{II}^*) = \mathbf{V}_{I,II}^*, \end{aligned} \quad (\text{B.5})$$

\mathbf{V}_I^* e $\mathbf{V}_{I,II}^*$ estão definidos em (2.17) e \mathbf{e}_i , em (2.10), e λ é o multiplicador de Lagrange. Simplificando, temos que

$$\hat{\mathbf{a}} = \begin{pmatrix} 1 \\ -\sigma_{yx}/\sigma_x^2 \end{pmatrix} \otimes \{D\mathbf{1}_n - (1 - k^*) \mathbf{C}_I\} + \bar{c}_{II} \begin{pmatrix} 0 \\ -\sigma_{yx}/\sigma_x^2 \end{pmatrix} \otimes \mathbf{1}_n,$$

em que σ_{yx} e σ_x^2 estão definidos em (2.6), D , k^* e \bar{c}_{II} , em (2.18), $\mathbf{C}_I = (c_1 \dots c_i \dots c_n)'$,

$$[\mathbf{V}_I^*]^{-1} = \Sigma^{-1} \otimes \mathbf{P}_{n,n-N} - \Sigma^{-1} \mathbf{e}_1 \mathbf{e}'_1 \Sigma^{-1} \otimes \mathbf{F}$$

em que

$$\Sigma^{-1} = \begin{pmatrix} \sigma_{11}^o & \sigma_{12}^o \\ \sigma_{12}^o & \sigma_{22}^o \end{pmatrix},$$

$$\mathbf{F} = (\sigma_{11}^o)^{-1} (1 - k^*) \mathbf{P}_{n,n-N} \mathbf{P}_{n,h} \mathbf{P}_{n,n-N}$$

com $h^{-1} = (1 - k^*) / (N - nk^*)$ e $\mathbf{P}_{a,b}$ definido em (2.5),

$$(\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I = \mathbf{e}'_1 \otimes \bar{\sigma}_e^2 \mathbf{C}_I$$

com $\bar{\sigma}_e^2$ definido em (A.2),

$$\mathbf{V}_{I,II}\mathbf{g}_{II} = -\left(\frac{[(N-n)\bar{c}_{II}]}{N}\right)(\boldsymbol{\Sigma}\mathbf{e}'_1 \otimes \mathbf{1}_n),$$

e $\lambda/2 = D/[\sigma_{11}^o k^*]$. A expressão final para \hat{T}^* está dada em (2.18).

$$\begin{aligned} \text{Var}_{SR}(\hat{T}^* - T) &= \hat{\boldsymbol{\alpha}}'\mathbf{V}_I^*\hat{\boldsymbol{\alpha}} + \mathbf{g}'_I(\mathbf{V}_I^* - \mathbf{V}_I)\mathbf{g}_I + 2\hat{\boldsymbol{\alpha}}'(\mathbf{V}_I^* - \mathbf{V}_I)\mathbf{g}_I \\ &\quad - 2\hat{\boldsymbol{\alpha}}'\mathbf{V}_{I,II}\mathbf{g}_{II} + \mathbf{g}'_{II}\mathbf{V}_{II}\mathbf{g}_{II} \end{aligned} \quad (\text{B.6})$$

em que

$$\begin{aligned} \mathbf{V}_{II} &= \text{Var}_S(\mathbf{Z}_{II}) = \boldsymbol{\Sigma} \otimes \mathbf{P}_{N-n,N}, \\ \hat{\boldsymbol{\alpha}}'\mathbf{V}_I^*\hat{\boldsymbol{\alpha}} &= \sigma_y^2(1 - \rho_{yx}^2) \sum_{i=1}^n \frac{1}{k^*} (D - (1 - k^*)c_i)^2 \\ &\quad + \frac{(N-n)\sigma_y^2\bar{c}_{II}^2}{N} \left(n - N(1 - \rho_{yx}^2) \right), \\ \mathbf{g}'_I(\mathbf{V}_I^* - \mathbf{V}_I)\mathbf{g}_I &= \sigma_y^2(1 - \rho_{yx}^2) \left(\frac{1 - k^*}{k^*} \right) \sum_{i=1}^n c_i^2, \\ \hat{\boldsymbol{\alpha}}'(\mathbf{V}_I^* - \mathbf{V}_I)\mathbf{g}_I &= \sigma_y^2(1 - \rho_{yx}^2) \left(\frac{1 - k^*}{k^*} \right) \sum_{i=1}^n c_i (D - (1 - k^*)c_i), \\ \hat{\boldsymbol{\alpha}}'\mathbf{V}_{I,II}\mathbf{g}_{II} &= -\frac{(N-n)\sigma_y^2\bar{c}_{II}^2}{N} \left[N(1 - \rho_{yx}^2) - n \right], \\ \mathbf{g}'_{II}\mathbf{V}_{II}\mathbf{g}_{II} &= (N - n - 1)\sigma_y^2\sigma_{c_{II}}^2 + \frac{n(N-n)}{N}\bar{c}_{II}^2\sigma_y^2, \end{aligned}$$

com ρ_{yx} e k^* definido em (2.18) e $\sigma_{c_{II}}^2$, em (2.19). A expressão final está dada em (2.19).

Obtenção da variância de \tilde{T} sob o modelo (2.12)

$$\begin{aligned} \text{Var}_{SR}(\tilde{T} - T) &= \tilde{\boldsymbol{\alpha}}'\tilde{\mathbf{V}}_I\tilde{\boldsymbol{\alpha}} + \mathbf{g}'_I(\tilde{\mathbf{V}}_I - \mathbf{V}_I)\mathbf{g}_I + 2\tilde{\boldsymbol{\alpha}}'(\tilde{\mathbf{V}}_I - \mathbf{V}_I)\mathbf{g}_I \\ &\quad - 2\tilde{\boldsymbol{\alpha}}'\mathbf{V}_{I,II}\mathbf{g}_{II} + \mathbf{g}'_{II}\mathbf{V}_{II}\mathbf{g}_{II} \end{aligned}$$

em que \mathbf{g}_I , \mathbf{g}_{II} , \mathbf{V}_I e \mathbf{V}_{II} estão definidos em (B.1), (B.2), (B.5) e (B.6), respectivamente;

$$\tilde{\mathbf{a}} = \begin{pmatrix} 1 \\ -\sigma_{yx}/\sigma_x^2 \end{pmatrix} \otimes \left\{ \frac{n\tilde{D}}{\sum_{i=1}^n \tilde{k}_i} \begin{pmatrix} \tilde{k}_1 \\ \vdots \\ \tilde{k}_n \end{pmatrix} - \begin{pmatrix} c_1(1-\tilde{k}_1) \\ \vdots \\ c_n(1-\tilde{k}_n) \end{pmatrix} \right\} + \bar{c}_{II} \begin{pmatrix} 0 \\ -\sigma_{yx}/\sigma_x^2 \end{pmatrix} \otimes \mathbf{1}_n,$$

$$\mathbf{V}_{I,II} = -N^{-1}\boldsymbol{\Sigma} \otimes \mathbf{J}_{n \times (N-n)},$$

$$\tilde{\mathbf{V}}_I = \text{Var}_{SR}(\tilde{\mathbf{Z}}_I) = \boldsymbol{\Sigma} \otimes \mathbf{P}_{n,N} + \mathbf{e}_1 \mathbf{e}_1' \otimes \left(\bigoplus_{i=1}^n \tilde{\sigma}_i^2 \right),$$

$$\begin{aligned} \tilde{\mathbf{a}}' \tilde{\mathbf{V}}_I \tilde{\mathbf{a}} &= \sigma_y^2 (1 - \rho_{yx}^2) \sum_{i=1}^n \frac{1}{\tilde{k}_i} \left(\frac{n\tilde{D}\tilde{k}_i}{\sum_{i=1}^n \tilde{k}_i} - c_i(1-\tilde{k}_i) \right)^2 \\ &\quad + \frac{(N-n)\sigma_y^2 \bar{c}_{II}^2}{N} (n - N(1 - \rho_{yx}^2)), \end{aligned}$$

$$\mathbf{g}_I' (\tilde{\mathbf{V}}_I - \mathbf{V}_I) \mathbf{g}_I = \sigma_y^2 (1 - \rho_{yx}^2) \sum_{i=1}^n c_i^2 \left(\frac{1-\tilde{k}_i}{\tilde{k}_i} \right),$$

$$\tilde{\mathbf{a}}' (\tilde{\mathbf{V}}_I - \mathbf{V}_I) \mathbf{g}_I = \sigma_y^2 (1 - \rho_{yx}^2) \sum_{i=1}^n c_i \left(\frac{1-\tilde{k}_i}{\tilde{k}_i} \right) \left(\frac{n\tilde{D}\tilde{k}_i}{\sum_{i=1}^n \tilde{k}_i} - c_i(1-\tilde{k}_i) \right),$$

$$\tilde{\mathbf{a}}' \mathbf{V}_{I,II} \mathbf{g}_{II} = \hat{\mathbf{a}}' \mathbf{V}_{I,II} \mathbf{g}_{II}, \text{ ver (B.6),}$$

e $\mathbf{g}_{II}' \mathbf{V}_{II} \mathbf{g}_{II}$ igual àquele apresentado em (B.6) com $\boldsymbol{\Sigma}$ definida em (2.6) e $\sigma_{c_{II}}^2$, em (2.19), ρ_{yx} e \bar{c}_{II} definidos em (2.18), e \tilde{D} e \tilde{k}_i , em (2.20). A expressão final está dada em (2.21).

Apêndice C

Resumo das simulações do Capítulo 2

Figura C.1: EQM ($\times 10^{-2}$), média do vício relativo e contribuição relativa do vício ao EQM de \hat{B}^* , \hat{B}_e^* e \hat{B}_{MQO} para diferentes valores de ρ_{yx}^\bullet , n (o primeiro número entre parênteses) e σ_e^\bullet (o segundo número entre parênteses). Variâncias iguais para os erros de medida endógenos. Dados com distribuição simétrica.

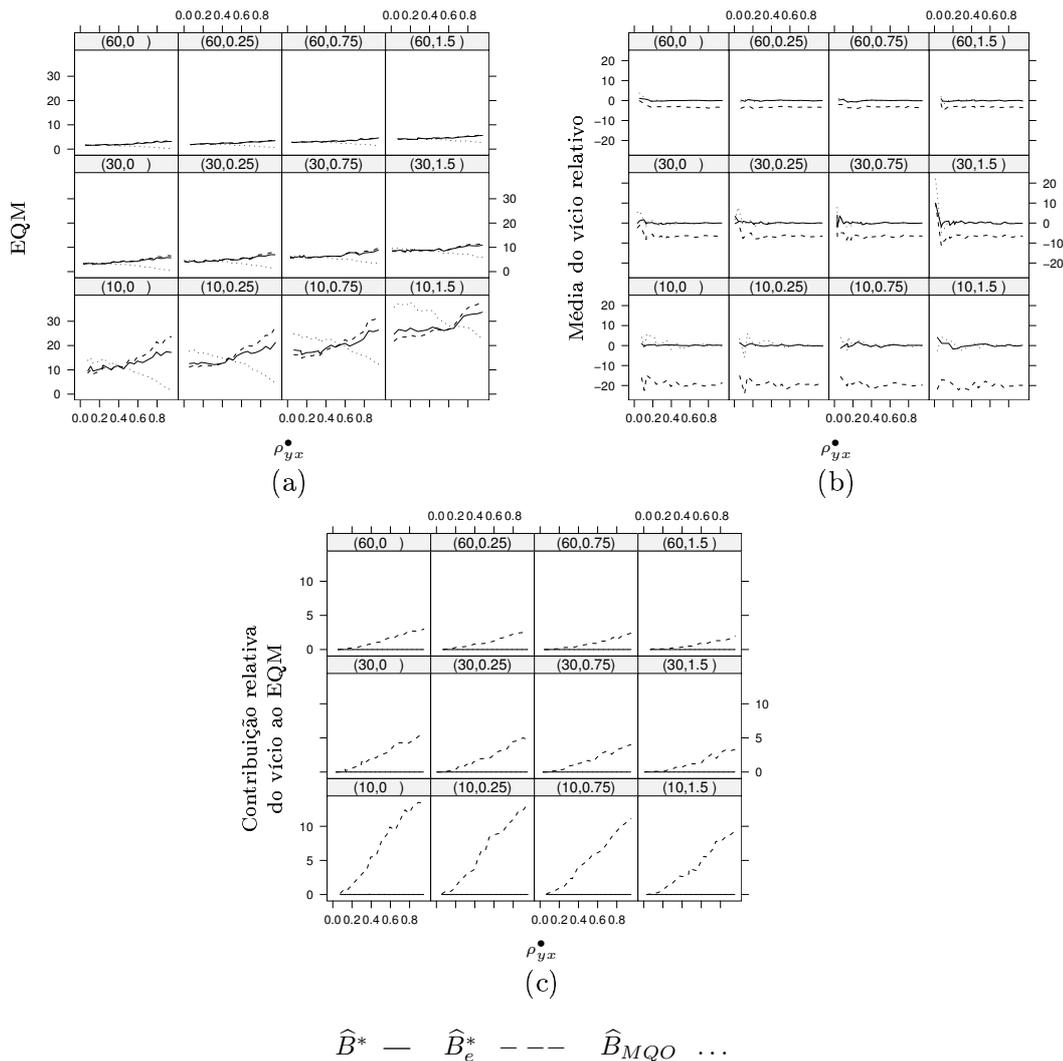
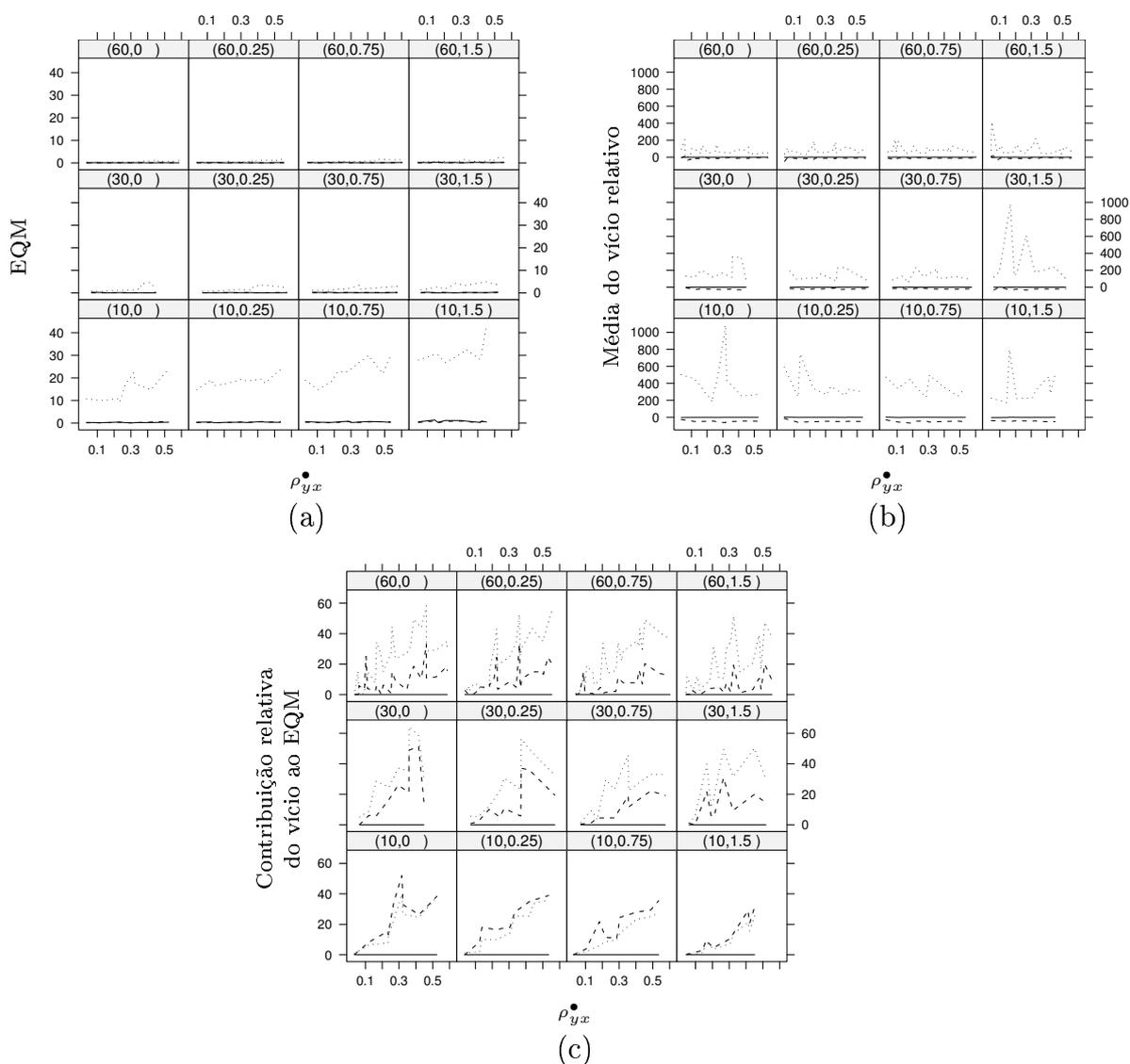


Figura C.2: EQM ($\times 10^{-2}$), média do vício relativo e contribuição relativa do vício ao EQM de \hat{B}^* , \hat{B}_e^* e \hat{B}_{MQO} para diferentes valores de ρ_{yx}^\bullet , n (o primeiro número entre parênteses) e σ_e^\bullet (o segundo número entre parênteses). Variâncias iguais para os erros de medida endógenos. Dados com distribuição assimétrica in x .



\hat{B}^* — \hat{B}_e^* - - - \hat{B}_{MQO} ...

Figura C.3: EQM ($\times 10^2$) e média do vício relativo de \hat{B}^* , \hat{B}_e^* e \hat{B}_{MQO} para diferentes valores de ρ_{yx}^\bullet e n . Variâncias diferentes para os erros de medida endógenos. Dados com distribuição simétrica.

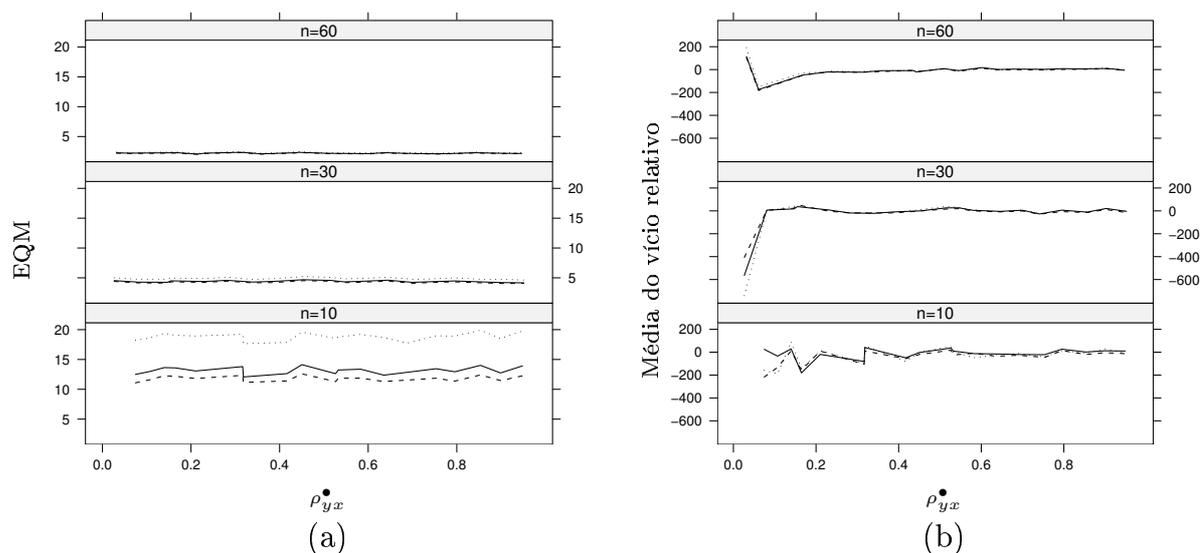
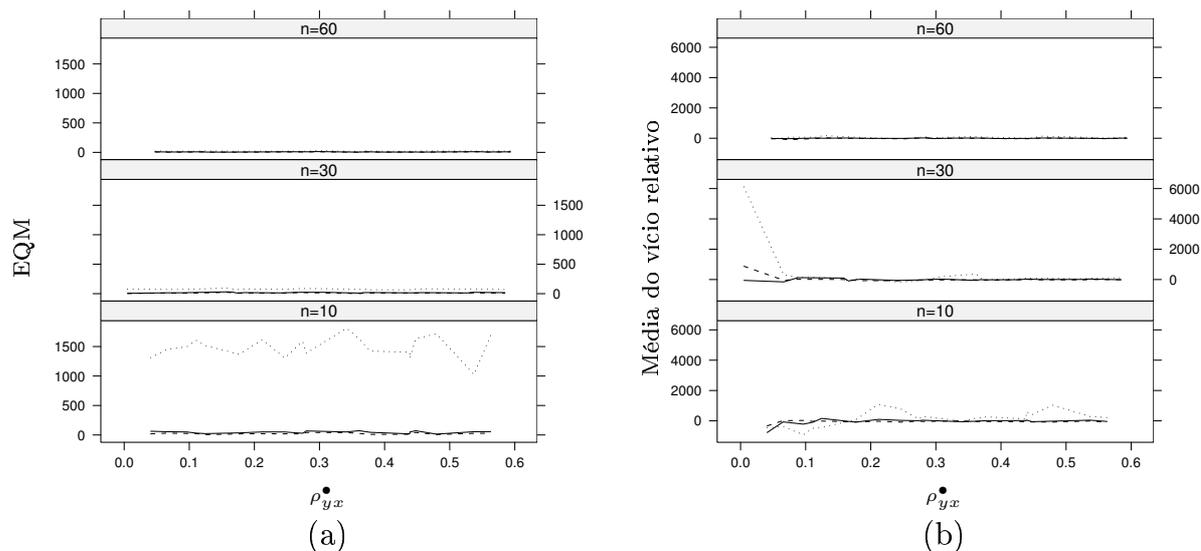


Figura C.4: EQM e média do vício relativo de \hat{B}^* , \hat{B}_e^* e \hat{B}_{MQO} para diferentes valores de ρ_{yx}^\bullet e n . Variâncias diferentes para os erros de medida endógenos. Dados com distribuição assimétrica in x .



Nas Figuras C.3 e C.4, não apresentamos o diagrama de dispersão de ρ_{yx}^\bullet versus contribuição relativa do vício ao EQM de \hat{B}^* , \hat{B}_e^* e \hat{B}_{MQO} porque é menor que 1% para os três estimadores.

Figura C.5: EQM, média do vício relativo e contribuição relativa do vício ao EQM de \tilde{B} , \tilde{B}_e e \hat{B}_{MQO} para diferentes valores de ρ_{yx}^\bullet e n . Variâncias diferentes para os erros de medida exógenos. Dados com distribuição simétrica.

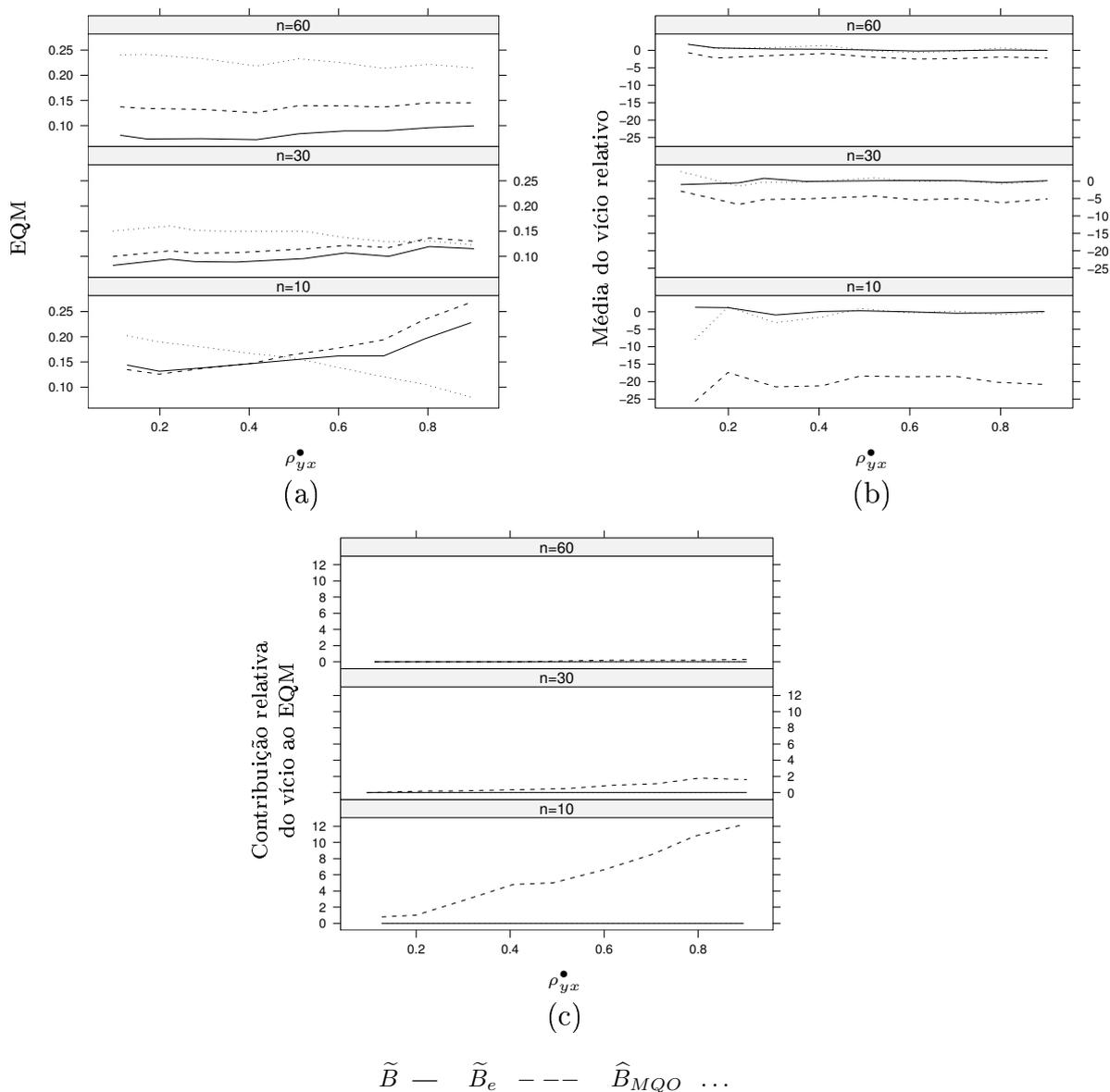
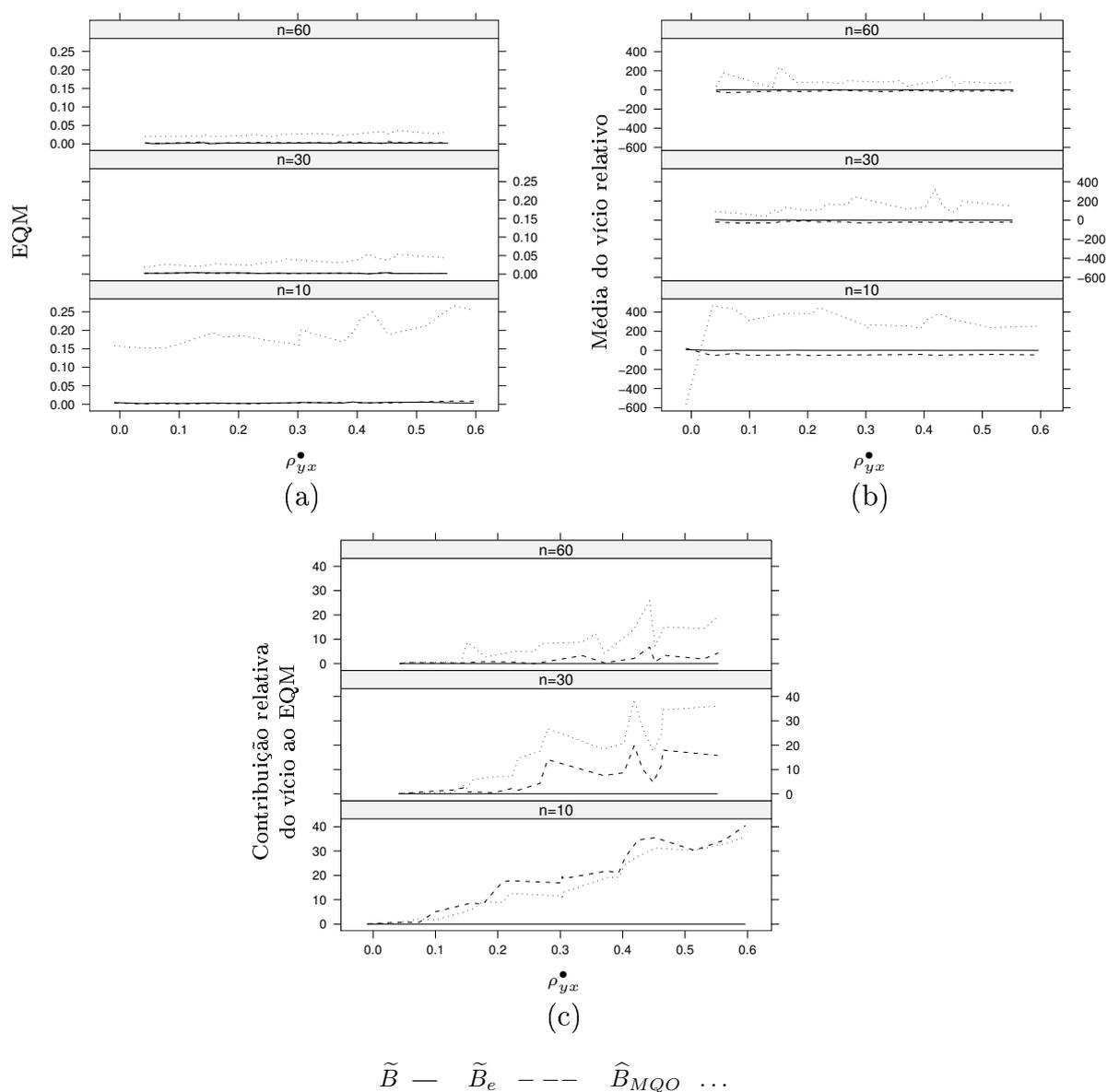


Figura C.6: EQM, média do vício relativo e contribuição relativa do vício ao EQM de \tilde{B} , \tilde{B}_e e \hat{B}_{MQO} para diferentes valores de ρ_{yx}^\bullet e n . Variâncias diferentes para os erros de medida exógenos. Dados com distribuição assimétrica in x .



Apêndice D

Obtenção do preditor de T (3.9) sob o modelo (3.4), do preditor de T (3.33) sob o modelo (3.28) e de suas respectivas variâncias

Obtenção do preditor de T (3.9) e de sua variância sob o modelo (3.4)

Semelhante ao Apêndice B, para obter um estimador linear não viciado de variância mínima de

$$T = \sum_{i=1}^N (c_{1i1}Y_{1i1} + c_{1i2}Y_{1i2}) + \sum_{i=1}^N (c_{2i1}Y_{2i1} + c_{2i2}Y_{2i2}),$$

sob o modelo (3.4), consideramos as seguintes três condições:

- i) Deve ser uma função linear das variáveis aleatórias observadas na amostra, i.e.,

$$\hat{T}^* = (\mathbf{g}'_I + \mathbf{a}') \mathbf{Z}_I^*$$

com \mathbf{Z}_I^* e \mathbf{g}_I definidos em (3.15) e (3.19), respectivamente,

ii) deve ser não viciado, i.e.,

$$\mathbb{E}_{SR}([\mathbf{g}'_I + \mathbf{a}'] \mathbf{Z}_I^*) = \mathbb{E}_S(\mathbf{g}'_I \mathbf{Z}_I + \mathbf{g}'_{II} \mathbf{Z}_{II})$$

em que

$$\begin{aligned} \mathbf{Z}_I &= (Y_{111} \ \dots \ Y_{1n1} \ Y_{211} \ \dots \ Y_{2n1} \ Y_{112} \ \dots \ Y_{1n2} \ Y_{212} \ \dots \ Y_{2n2})', \\ \mathbf{Z}_{II} &= (Y_{1,(n+1),1} \ \dots \ Y_{1N1} \ Y_{2,(n+1),1} \ \dots \ Y_{2N1} \ Y_{1,(n+1),2} \ \dots \ Y_{1N2} \ Y_{2,(n+1),2} \ \dots \\ &\quad Y_{2N2})', \end{aligned}$$

e \mathbf{g}_{II} está definido em (3.20), e

iii) deve minimizar a $\text{Var}_{SR}(\hat{T}^* - T)$.

Empregando multiplicadores de Lagrange, devemos encontrar a solução do sistema de equações dado por (D.1) e (D.2):

$$2[\mathbf{V}_I^* \hat{\mathbf{a}} + (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I - \mathbf{V}_{I,II} \mathbf{g}_{II}] = (\mathbf{I}_4 \otimes \mathbf{1}_n) \boldsymbol{\lambda}_{4 \times 1} \quad (\text{D.1})$$

e

$$(\mathbf{I}_4 \otimes \mathbf{1}_n)' \hat{\mathbf{a}} - (\mathbf{I}_4 \otimes \mathbf{1}_{N-n})' \mathbf{g}_{II} = 0, \quad (\text{D.2})$$

em que

$$\begin{aligned} \mathbf{V}_I &= \text{Var}_S(\mathbf{Z}_I) = \boldsymbol{\Sigma} \otimes \mathbf{P}_{n,N}, \\ \mathbf{V}_{I,II} &= \text{Cov}_S(\mathbf{Z}_I, \mathbf{Z}_{II}) = \text{Cov}_{SR}(\mathbf{Z}_I^*, \mathbf{Z}_{II}) \\ &= \text{Cov}_{SR}(\mathbf{Z}_I^*, \mathbf{Z}_{II}^*) = \mathbf{V}_{I,II}^*, \end{aligned} \quad (\text{D.3})$$

$\boldsymbol{\Sigma}$ e $\mathbf{P}_{a,b}$ estão definidos em (3.18) e (2.5), respectivamente, e \mathbf{V}_I^* e $\mathbf{V}_{I,II}^*$, em (3.17), e $\boldsymbol{\lambda}$ é o multiplicador de Lagrange. Então

$$\begin{aligned} \hat{\mathbf{a}} &= (\mathbf{V}_I^*)^{-1} [\mathbf{V}_{I,II} \mathbf{g}_{II} - (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I] \\ &\quad - (\mathbf{V}_I^*)^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_n) [(\mathbf{I}_4 \otimes \mathbf{1}_n)' (\mathbf{V}_I^*)^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_n)]^{-1} \times \\ &\quad \{(\mathbf{I}_4 \otimes \mathbf{1}_n)' (\mathbf{V}_I^*)^{-1} [\mathbf{V}_{I,II} \mathbf{g}_{II} - (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I] - (\mathbf{I}_4 \otimes \mathbf{1}_{N-n})' \mathbf{g}_{II}\}. \end{aligned}$$

Simplificando, temos que

•

$$(\mathbf{V}_I^*)^{-1} = [\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n] \left[\mathbf{A}^{-1} + \frac{1}{N} \mathbf{A}^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_n) \mathbf{L} (\mathbf{I}_4 \otimes \mathbf{1}_n)' \mathbf{A}^{-1} \right],$$

•

$$(\mathbf{V}_I^*)^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_n) = (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n) \mathbf{A}^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_n) \mathbf{L},$$

•

$$[(\mathbf{I}_4 \otimes \mathbf{1}_n)' (\mathbf{V}_I^*)^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_n)]^{-1} = \mathbf{L}^{-1} \mathbf{G}^{-1} \boldsymbol{\Sigma},$$

•

$$(\mathbf{V}_I^*)^{-1} (\mathbf{V}_I^* - \mathbf{V}_I) = \left[\mathbf{I}_{4n} + \frac{1}{N} \mathbf{A}^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_n) \mathbf{L} (\mathbf{I}_4 \otimes \mathbf{1}_n)' \right] (\mathbf{I}_{4n} - \mathbf{A}^{-1}),$$

•

$$(\mathbf{I}_4 \otimes \mathbf{1}_n) [(\mathbf{I}_4 \otimes \mathbf{1}_n)' (\mathbf{V}_I^*)^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_n)]^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_n)' (\mathbf{V}_I^*)^{-1} \mathbf{V}_{I,II} \mathbf{g}_{II} = \mathbf{V}_{I,II} \mathbf{g}_{II}$$

•

$$\begin{aligned} (\mathbf{V}_I^*)^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_n) [(\mathbf{I}_4 \otimes \mathbf{1}_n)' (\mathbf{V}_I^*)^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_n)]^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_n)' (\mathbf{V}_I^*)^{-1} = \\ \mathbf{A}^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_n) \mathbf{G}^{-1} \boldsymbol{\Sigma}^{-1} \mathbf{L} (\mathbf{I}_4 \otimes \mathbf{1}_n)' \mathbf{A}^{-1} \end{aligned}$$

dado que

$$\begin{aligned} (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n) \mathbf{A}^{-1} (\boldsymbol{\Sigma} \otimes \mathbf{I}_n) &= \mathbf{A}^{-1}, \\ \boldsymbol{\Sigma}^{-1} \mathbf{G}^{-1} \boldsymbol{\Sigma} &= \mathbf{G}^{-1}, \text{ e} \\ (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n) \mathbf{A}^{-1} (\mathbf{V}_I^* - \mathbf{V}_I) &= \mathbf{I}_{4n} - \mathbf{A}^{-1}, \end{aligned}$$

em que $\mathbf{L} = (\mathbf{I}_4 - \frac{1}{N} \mathbf{G})^{-1}$, $\boldsymbol{\Sigma}$ está definida em (3.18), e \mathbf{A} e \mathbf{G} , em (3.19). Daí segue

$$\hat{\mathbf{a}} = \mathbf{A}^{-1} (\mathbf{G}^{-1} \otimes \mathbf{I}_n) (\mathbf{D} \otimes \mathbf{1}_n) - (\mathbf{I}_{4n} - \mathbf{A}^{-1}) \mathbf{g}_I$$

com \mathbf{D} definido em (3.19).

$$\begin{aligned} \text{Var}_{SR} (\hat{T}^* - T) &= \hat{\mathbf{a}}' \mathbf{V}_I^* \hat{\mathbf{a}} + \mathbf{g}_I' (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I \\ &\quad + 2 \hat{\mathbf{a}}' (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I - 2 \hat{\mathbf{a}}' \mathbf{V}_{I,II} \mathbf{g}_{II} + \mathbf{g}_{II}' \mathbf{V}_{II} \mathbf{g}_{II} \end{aligned} \tag{D.4}$$

em que

•

$$\mathbf{V}_{II} = \text{Var}_s(\mathbf{Z}_{II}) = \boldsymbol{\Sigma} \otimes \mathbf{P}_{N-n, N},$$

•

$$\begin{aligned} \hat{\mathbf{a}}' \mathbf{V}_I^* \hat{\mathbf{a}} &= \mathbf{D}' \boldsymbol{\Sigma} \mathbf{G}^{-1} \mathbf{D} - 2(\mathbf{D} \otimes \mathbf{1}_n)' (\mathbf{G}^{-1} \otimes \mathbf{I}_n) \mathbf{A}^{-1} (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I \\ &\quad + \mathbf{g}'_I (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I - \mathbf{g}'_I \mathbf{A}^{-1} (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I - \frac{1}{N} \mathbf{g}'_{II} (\boldsymbol{\Sigma} \otimes \mathbf{J}_{N-n}) \mathbf{g}_{II}, \end{aligned}$$

•

$$\begin{aligned} \hat{\mathbf{a}}' (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I &= (\mathbf{D} \otimes \mathbf{1}_n)' (\mathbf{G}^{-1} \otimes \mathbf{I}_n) \mathbf{A}^{-1} (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I \\ &\quad - \mathbf{g}'_I (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I + \mathbf{g}'_I \mathbf{A}^{-1} (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I, \end{aligned}$$

•

$$\hat{\mathbf{a}}' \mathbf{V}_{I,II} \mathbf{g}_{II} = -\frac{1}{N} \mathbf{g}'_{II} (\boldsymbol{\Sigma} \otimes \mathbf{J}_{N-n}) \mathbf{g}_{II},$$

•

$$\mathbf{g}'_{II} \mathbf{V}_{II} \mathbf{g}_{II} = \mathbf{g}'_{II} (\boldsymbol{\Sigma} \otimes \mathbf{P}_{N-n}) \mathbf{g}_{II} + \frac{n}{N(N-n)} \mathbf{g}'_{II} (\boldsymbol{\Sigma} \otimes \mathbf{J}_{N-n}) \mathbf{g}_{II},$$

A expressão final está dada em (3.20).

Obtenção do preditor de T (3.33) e de sua variância sob o modelo (3.28)

Semelhante ao Apêndice B, para obter um estimador linear não viciado de variância mínima de

$$T = \sum_{i=1}^N (c_{i1}^C Y_{i1} + c_{i2}^C Y_{i2}^C) + \sum_{i=1}^N (c_{i1}^T Y_{i1} + c_{i2}^T Y_{i2}^T),$$

sob o modelo (3.28), consideramos as seguintes três condições:

a) Deve ser uma função linear das variáveis aleatórias observadas na amostra, i.e.,

$$\hat{T}^* = (\mathbf{g}'_I + \mathbf{a}') \mathbf{Z}_I^*$$

em que \mathbf{Z}_I^* e \mathbf{g}_I estão definidas em (3.39) e (3.43), respectivamente,

b) deve ser não viciado, i.e.,

$$\mathbb{E}_{SR}([\mathbf{g}'_I + \mathbf{a}'] \mathbf{Z}_I^*) = \mathbb{E}_S(\mathbf{g}'_I \mathbf{Z}_I + \mathbf{g}'_{II} \mathbf{Z}_{II})$$

em que

$$\begin{aligned} \mathbf{Z}_I &= (Y_{11} \ \dots \ Y_{2n_0,1} \ Y_{12}^C \ \dots \ Y_{n_0,2}^C \ Y_{(n_0+1),2}^T \ \dots \ Y_{2n_0,2}^T)', \\ \mathbf{Z}_{II} &= (Y_{(2n_0+1),1} \ \dots \ Y_{N1} \ Y_{(2n_0+1),2}^C \ \dots \ Y_{N2}^C \ Y_{(2n_0+1),2}^T \ \dots \ Y_{N,2}^T \ Y_{(n_0+1),2}^C \ \dots \\ &\quad Y_{2n_0,2}^C \ Y_{12}^T \ \dots \ Y_{n_0,2}^T)', \end{aligned}$$

\mathbf{g}_{II} está definido em (3.43), e

c) deve minimizar a $\text{Var}_{SR}(\hat{T}^* - T)$.

Empregando multiplicadores de Lagrange, devemos encontrar a solução do sistema de equações dado por (D.5) e (D.6):

$$2[\mathbf{V}_I^* \hat{\mathbf{a}} + (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I - \mathbf{V}_{I,II} \mathbf{g}_{II}] = [\mathbf{H}_0 \otimes \mathbf{1}_{n_0}] \boldsymbol{\lambda}_{3 \times 1} \quad (\text{D.5})$$

e

$$[\mathbf{H}_0 \otimes \mathbf{1}_{n_0}]' \tilde{\mathbf{a}} - \mathbf{H}'_1 \mathbf{g}_{II} = 0, \quad (\text{D.6})$$

em que

$$\begin{aligned} \mathbf{V}_I &= \text{Var}_S(\mathbf{Z}_I) = (\boldsymbol{\Sigma}^a \otimes \mathbf{I}_{n_0}) - N^{-1}(\mathbf{I}_4 \otimes \mathbf{1}_{n_0}) \boldsymbol{\Sigma}^b (\mathbf{I}_4 \otimes \mathbf{1}_{n_0})', \\ \mathbf{V}_{I,II} &= \text{Cov}_S(\mathbf{Z}_I, \mathbf{Z}_{II}) = \text{Cov}_{SR}(\mathbf{Z}_I^*, \mathbf{Z}_{II}) \\ &= \text{Cov}_{SR}(\mathbf{Z}_I^*, \mathbf{Z}_{II}^*) = \mathbf{V}_{I,II}^*, \end{aligned} \quad (\text{D.7})$$

\mathbf{V}_I^* e $\mathbf{V}_{I,II}^*$ estão definidos em (3.41), \mathbf{g}_I e \mathbf{g}_{II} , em (3.43); $\boldsymbol{\Sigma}^a$, $\boldsymbol{\Sigma}^b$, \mathbf{H}_0 e \mathbf{H}_1 , em (3.42) e $\boldsymbol{\lambda}$ é o multiplicador de Lagrange. Então

$$\begin{aligned} \hat{\mathbf{a}} &= (\mathbf{V}_I^*)^{-1} [\mathbf{V}_{I,II} \mathbf{g}_{II} - (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I] \\ &\quad - (\mathbf{V}_I^*)^{-1} (\mathbf{H}_0 \otimes \mathbf{1}_{n_0}) \{(\mathbf{H}_0 \otimes \mathbf{1}_{n_0})' (\mathbf{V}_I^*)^{-1} (\mathbf{H}_0 \otimes \mathbf{1}_{n_0})\}^{-1} \times \\ &\quad \{(\mathbf{H}_0 \otimes \mathbf{1}_{n_0})' (\mathbf{V}_I^*)^{-1} [\mathbf{V}_{I,II} \mathbf{g}_{II} - (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I] - \mathbf{H}'_1 \mathbf{g}_{II}\}. \end{aligned}$$

Simplificando, temos que

•

$$(\mathbf{V}_I^*)^{-1} = [(\boldsymbol{\Sigma}^a)^{-1} \otimes \mathbf{I}_{n_0}] \times \left[\mathbf{A}^{-1} + N^{-1} \mathbf{A}^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_{n_0}) \mathbf{H}_0 \boldsymbol{\Sigma} (\mathbf{I}_3 - N^{-1} \mathbf{H}_4 \boldsymbol{\Sigma})^{-1} \mathbf{H}_0' (\boldsymbol{\Sigma}^a)^{-1} (\mathbf{I}_4 \otimes \mathbf{1}'_{n_0}) \mathbf{A}^{-1} \right],$$

•

$$[(\mathbf{H}_0 \otimes \mathbf{1}_{n_0})' (\mathbf{V}_I^*)^{-1} (\mathbf{H}_0 \otimes \mathbf{1}_{n_0})]^{-1} = \mathbf{H}_4^{-1} (\mathbf{I}_3 - N^{-1} \mathbf{H}_4 \boldsymbol{\Sigma}),$$

dado que

•

$$[(\boldsymbol{\Sigma}^a)^{-1} \otimes \mathbf{I}_n] \mathbf{A}^{-1} (\mathbf{V}_I^* - \mathbf{V}_I) = (\mathbf{I}_{4n_0} - \mathbf{A}^{-1}),$$

•

$$[(\boldsymbol{\Sigma}^a)^{-1} \otimes \mathbf{I}_n] \mathbf{A}^{-1} [\boldsymbol{\Sigma}^a \otimes \mathbf{I}_n] = [\boldsymbol{\Sigma}^a \otimes \mathbf{I}_n] \mathbf{A}^{-1} [(\boldsymbol{\Sigma}^a)^{-1} \otimes \mathbf{I}_n] = \mathbf{A}^{-1},$$

em que $\boldsymbol{\Sigma}$ está definido em (3.30), $\boldsymbol{\Sigma}^a$ e \mathbf{H}_0 , em (3.42), e \mathbf{A} e \mathbf{H}_4 , em (3.43). Daí segue

$$\hat{\mathbf{a}} = \mathbf{A}^{-1} [(\boldsymbol{\Sigma}^a)^{-1} \mathbf{H}_0 \mathbf{H}_4^{-1} \mathbf{D} \otimes \mathbf{1}_{n_0}] - [\mathbf{I}_{4n_0} - \mathbf{A}^{-1}] \mathbf{g}_I + \mathbf{A}^{-1} \mathbf{H}_3 \mathbf{g}_{II}$$

com \mathbf{H}_3 definido em (3.42) e \mathbf{D} , em (3.43).

$$\text{Var}_{SR} (\hat{T}^* - T) = \hat{\mathbf{a}}' \mathbf{V}_I^* \hat{\mathbf{a}} + \mathbf{g}_I' (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I + 2 \hat{\mathbf{a}}' (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I - 2 \hat{\mathbf{a}}' \mathbf{V}_{I,II} \mathbf{g}_{II} + \mathbf{g}_{II}' \mathbf{V}_{II} \mathbf{g}_{II} \quad (\text{D.8})$$

em que

•

$$\begin{aligned} \mathbf{V}_{II} = \text{Var}_s (\mathbf{Z}_{II}) &= +n_0^{-1} \mathbf{H}_2' (\mathbf{I}_4 \otimes \mathbf{1}_{n_0}) \boldsymbol{\Sigma}^a (\mathbf{I}_4 \otimes \mathbf{1}_{n_0})' \mathbf{H}_2 \\ &+ \text{diag} \left\{ \boldsymbol{\Sigma} \otimes \mathbf{P}_{N-2n_0}, \begin{pmatrix} \sigma_{22}^{CC} & 0 \\ 0 & \sigma_{22}^{TT} \end{pmatrix} \otimes \mathbf{P}_{n_0} \right\} - N^{-1} \mathbf{H}_1 \boldsymbol{\Sigma} \mathbf{H}_1' \\ &+ (N - 2n_0)^{-1} \begin{pmatrix} \mathbf{I}_3 \otimes \mathbf{1}_{N-2n_0} \\ \mathbf{0}_{2n_0 \times 3} \end{pmatrix} \boldsymbol{\Sigma} (\mathbf{I}_3 \otimes \mathbf{1}'_{N-2n_0} | \mathbf{0}_{3 \times 2n_0}), \end{aligned}$$

•

$$\begin{aligned}
\hat{\mathbf{a}}' \mathbf{V}_I^* \hat{\mathbf{a}} &= \mathbf{D}' \mathbf{H}_4^{-1'} \mathbf{D} + \mathbf{g}'_I (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I - \mathbf{g}'_I (\boldsymbol{\Sigma}^a \otimes \mathbf{I}_{n_0}) (\mathbf{I}_{4n_0} - \mathbf{A}^{-1}) \mathbf{g}_I \\
&\quad - N^{-1} \mathbf{g}'_{II} \mathbf{H}_1 \boldsymbol{\Sigma} \mathbf{H}'_1 \mathbf{g}_{II} + \mathbf{g}'_{II} \mathbf{H}'_3 \mathbf{A}^{-1} (\boldsymbol{\Sigma}^a \otimes \mathbf{I}_{n_0}) \mathbf{H}_3 \mathbf{g}_{II} \\
&\quad - 2\mathbf{D}' (\mathbf{H}'_4)^{-1} \mathbf{H}'_0 (\boldsymbol{\Sigma}^a)^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_{n_0})' \mathbf{A}^{-1} (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I \\
&\quad + 2\mathbf{D}' (\mathbf{H}'_4)^{-1} \mathbf{H}'_0 (\mathbf{I}_4 \otimes \mathbf{1}_{n_0})' \mathbf{A}^{-1} \mathbf{H}_3 \mathbf{g}_{II} - 2\mathbf{g}'_I (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{A}^{-1} \mathbf{H}_3 \mathbf{g}_{II} \\
&\quad - 2\mathbf{g}'_I (\boldsymbol{\Sigma}^a \otimes \mathbf{I}_{n_0}) \mathbf{A}^{-1} \mathbf{H}_3 \mathbf{g}_{II} + 2\mathbf{g}'_I \mathbf{A}^{-1} (\boldsymbol{\Sigma}^a \otimes \mathbf{I}_{n_0}) \mathbf{H}_3 \mathbf{g}_{II},
\end{aligned}$$

•

$$\begin{aligned}
\hat{\mathbf{a}}' (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I &= \mathbf{D}' (\mathbf{H}'_4)^{-1} \mathbf{H}'_0 (\boldsymbol{\Sigma}^a)^{-1} (\mathbf{I}_4 \otimes \mathbf{1}_{n_0})' \mathbf{A}^{-1} (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I \\
&\quad - \mathbf{g}'_I (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I + \mathbf{g}'_I \mathbf{A}^{-1} (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I \\
&\quad + \mathbf{g}'_{II} \mathbf{H}'_3 \mathbf{A}^{-1} (\mathbf{V}_I^* - \mathbf{V}_I) \mathbf{g}_I,
\end{aligned}$$

•

$$\begin{aligned}
\hat{\mathbf{a}}' \mathbf{V}_{I,II} \mathbf{g}_{II} &= \mathbf{D}' (\mathbf{H}'_4)^{-1} \mathbf{H}'_0 (\mathbf{I}_4 \otimes \mathbf{1}_{n_0})' \mathbf{A}^{-1} \mathbf{H}_3 \mathbf{g}_{II} - N^{-1} \mathbf{g}'_{II} \mathbf{H}_1 \boldsymbol{\Sigma} \mathbf{H}'_1 \mathbf{g}_{II} \\
&\quad - \mathbf{g}'_I (\boldsymbol{\Sigma}^a \otimes \mathbf{I}_{n_0}) \mathbf{H}_3 \mathbf{g}_{II} + \mathbf{g}'_I \mathbf{A}^{-1} (\boldsymbol{\Sigma}^a \otimes \mathbf{I}_{n_0}) \mathbf{H}_3 \mathbf{g}_{II} \\
&\quad + \mathbf{g}'_{II} \mathbf{H}'_3 \mathbf{A}^{-1} (\boldsymbol{\Sigma}^a \otimes \mathbf{I}_{n_0}) \mathbf{H}_3 \mathbf{g}_{II},
\end{aligned}$$

•

$$\begin{aligned}
\mathbf{g}'_{II} \mathbf{V}_{II} \mathbf{g}_{II} &= \mathbf{g}'_{II} \text{diag} \left\{ \boldsymbol{\Sigma} \otimes \mathbf{P}_{N-2n_0}, \begin{pmatrix} \sigma_{22}^{CC} & 0 \\ 0 & \sigma_{22}^{TT} \end{pmatrix} \otimes \mathbf{P}_{n_0} \right\} \mathbf{g}_{II} - N^{-1} \mathbf{g}'_{II} \mathbf{H}_1 \boldsymbol{\Sigma} \mathbf{H}'_1 \mathbf{g}_{II} \\
&\quad + (N - 2n_0)^{-1} \mathbf{g}'_{II} \begin{pmatrix} \mathbf{I}_3 \otimes \mathbf{1}_{N-2n_0} \\ \mathbf{0}_{2n_0 \times 3} \end{pmatrix} \boldsymbol{\Sigma} (\mathbf{I}_3 \otimes \mathbf{1}'_{N-2n_0} | \mathbf{0}_{3 \times 2n_0}) \mathbf{g}_{II} \\
&\quad + n_0^{-1} \mathbf{g}'_{II} \mathbf{H}'_2 (\mathbf{I}_4 \otimes \mathbf{1}_{n_0}) \boldsymbol{\Sigma}^a (\mathbf{I}_4 \otimes \mathbf{1}_{n_0})' \mathbf{H}_2 \mathbf{g}_{II},
\end{aligned}$$

\mathbf{P}_a está definido em (2.5), σ_{22}^{CC} e σ_{22}^{TT} , em (3.30) e \mathbf{H}_2 , em (3.42). A expressão final está dada em (3.44).

Apêndice E

Resumo das simulações relativas ao Capítulo 3, Seção 3.2

Tabela E.1: EQM ($\times 10^{-2}$) e contribuição relativa do vício ao EQM (entre parênteses) das estimativas, considerando variâncias iguais para os erros de medida endógenos. Dados com distribuição simétrica.

ρ_{21}^C	ρ_{21}^T	$\bar{\sigma}_1^2$	n_0	\hat{T}^*	\hat{T}_e^*	\bar{T}	$2\hat{\tau}_c$
0.1-0.3	0.1-0.3	0.000	10	18.06 (0.0%)	18.98 (0.0%)	35.55 (0.0%)	19.18 (0.0%)
			30	5.76 (0.0%)	5.89 (0.0%)	9.74 (0.0%)	5.90 (0.0%)
		0.010	10	18.93 (0.0%)	20.19 (0.0%)	30.64 (0.0%)	20.45 (0.0%)
			30	6.19 (0.0%)	6.33 (0.0%)	10.67 (0.0%)	6.34 (0.0%)
		0.016	10	18.65 (0.0%)	19.65 (0.1%)	32.46 (0.0%)	19.87 (0.1%)
			30	6.38 (0.0%)	6.49 (0.0%)	10.62 (0.0%)	6.50 (0.0%)
		0.090	10	18.52 (0.0%)	19.43 (0.0%)	33.25 (0.0%)	19.80 (0.0%)
			30	6.95 (0.0%)	7.07 (0.0%)	11.73 (0.0%)	7.10 (0.0%)
	0.4-0.6	0.000	10	17.41 (0.0%)	18.48 (0.0%)	26.04 (0.0%)	18.68 (0.0%)
			30	5.30 (0.0%)	5.39 (0.0%)	7.73 (0.0%)	5.39 (0.0%)
		0.010	10	16.82 (0.0%)	17.75 (0.0%)	25.33 (0.0%)	17.92 (0.0%)
			30	5.07 (0.0%)	5.19 (0.0%)	7.99 (0.0%)	5.19 (0.0%)
		0.016	10	17.85 (0.0%)	18.79 (0.0%)	27.60 (0.0%)	19.01 (0.0%)
			30	5.71 (0.0%)	5.82 (0.0%)	7.91 (0.0%)	5.82 (0.0%)
		0.090	10	19.81 (0.0%)	20.76 (0.0%)	29.94 (0.1%)	21.08 (0.0%)
			30	5.66 (0.0%)	5.74 (0.0%)	9.00 (0.0%)	5.75 (0.0%)
0.3-0.6	0.1-0.3	0.000	10	15.82 (0.0%)	16.60 (0.0%)	21.67 (0.0%)	16.73 (0.0%)
			30	5.66 (0.0%)	5.75 (0.0%)	8.85 (0.0%)	5.75 (0.0%)
		0.010	10	16.50 (0.0%)	17.54 (0.0%)	24.26 (0.0%)	17.68 (0.0%)
			30	5.54 (0.0%)	5.63 (0.0%)	8.73 (0.0%)	5.64 (0.0%)
		0.016	10	18.74 (0.0%)	19.74 (0.0%)	27.65 (0.1%)	19.97 (0.0%)
			30	5.36 (0.0%)	5.51 (0.0%)	8.32 (0.0%)	5.52 (0.0%)
		0.090	10	19.25 (0.0%)	20.18 (0.0%)	28.69 (0.0%)	20.47 (0.0%)
			30	6.20 (0.0%)	6.25 (0.1%)	10.66 (0.0%)	6.27 (0.1%)
	0.4-0.6	0.000	10	14.88 (0.0%)	15.59 (0.0%)	21.60 (0.0%)	15.65 (0.0%)
			30	4.48 (0.0%)	4.57 (0.0%)	6.90 (0.0%)	4.58 (0.0%)
		0.010	10	14.02 (0.0%)	14.70 (0.0%)	18.68 (0.0%)	14.74 (0.0%)
			30	4.61 (0.0%)	4.72 (0.1%)	6.48 (0.0%)	4.72 (0.1%)
		0.016	10	15.57 (0.0%)	16.57 (0.0%)	21.45 (0.0%)	16.61 (0.0%)
			30	4.56 (0.0%)	4.65 (0.0%)	6.71 (0.0%)	4.66 (0.0%)
		0.090	10	17.29 (0.0%)	18.07 (0.0%)	23.86 (0.0%)	18.30 (0.0%)
			30	5.13 (0.0%)	5.25 (0.0%)	7.29 (0.0%)	5.25 (0.0%)
0.7-0.9	0.1-0.3	0.000	10	15.37 (0.0%)	16.30 (0.0%)	21.23 (0.0%)	16.43 (0.0%)
			30	4.37 (0.0%)	4.47 (0.0%)	6.04 (0.0%)	4.47 (0.0%)
		0.010	10	14.97 (0.0%)	15.75 (0.0%)	21.53 (0.0%)	15.88 (0.0%)
			30	4.82 (0.0%)	4.95 (0.0%)	6.76 (0.0%)	4.96 (0.0%)
		0.016	10	15.58 (0.0%)	16.36 (0.0%)	20.47 (0.0%)	16.45 (0.0%)
			30	4.66 (0.0%)	4.78 (0.0%)	6.26 (0.0%)	4.78 (0.0%)
		0.090	10	16.62 (0.0%)	17.68 (0.0%)	22.45 (0.0%)	17.90 (0.0%)
			30	4.83 (0.0%)	4.94 (0.0%)	7.40 (0.0%)	4.95 (0.0%)
	0.4-0.6	0.000	10	12.21 (0.0%)	13.21 (0.0%)	14.92 (0.0%)	13.16 (0.0%)
			30	3.33 (0.0%)	3.41 (0.0%)	4.21 (0.0%)	3.40 (0.0%)
		0.010	10	11.38 (0.0%)	12.01 (0.0%)	14.51 (0.0%)	11.97 (0.0%)
			30	3.67 (0.0%)	3.74 (0.1%)	4.52 (0.0%)	3.74 (0.1%)
		0.016	10	11.46 (0.0%)	12.47 (0.0%)	13.62 (0.0%)	12.39 (0.0%)
			30	3.98 (0.0%)	4.10 (0.1%)	4.72 (0.1%)	4.09 (0.1%)
		0.090	10	14.72 (0.0%)	15.87 (0.0%)	17.64 (0.0%)	15.82 (0.0%)
			30	4.22 (0.0%)	4.34 (0.1%)	5.45 (0.1%)	4.33 (0.1%)

Tabela E.2: EQM e contribuição relativa do vício ao EQM (entre parênteses) das estimativas, considerando variâncias iguais para os erros de medida endógenos. Dados com distribuição assimétrica.

ρ_{21}^C	ρ_{21}^T	$\bar{\sigma}_1^2$	n_0	\hat{T}^*	\hat{T}_e^*	\bar{T}	$2\hat{\tau}_c$
0.0-0.25	0.0-0.25	0.000	10	5.75 (0.0%)	6.08 (0.0%)	5.92 (0.0%)	6.17 (0.0%)
			30	1.63 (0.0%)	1.66 (0.0%)	1.64 (0.0%)	1.66 (0.0%)
		0.010	10	5.97 (0.0%)	6.39 (0.0%)	6.01 (0.0%)	6.46 (0.0%)
			30	1.49 (0.0%)	1.50 (0.1%)	1.50 (0.0%)	1.51 (0.1%)
		0.016	10	7.13 (0.0%)	7.19 (0.1%)	7.15 (0.0%)	7.27 (0.1%)
			30	0.92 (0.0%)	0.94 (0.0%)	0.93 (0.0%)	0.94 (0.0%)
		0.090	10	6.24 (0.0%)	6.29 (0.2%)	6.24 (0.0%)	6.39 (0.2%)
			30	1.01 (0.0%)	1.02 (0.0%)	1.01 (0.0%)	1.02 (0.0%)
	0.25-0.50	0.000	10	6.42 (0.0%)	6.66 (0.2%)	6.42 (0.0%)	6.75 (0.2%)
			30	1.54 (0.0%)	1.56 (0.0%)	1.55 (0.0%)	1.56 (0.0%)
		0.010	10	4.93 (0.0%)	3.97 (0.6%)	5.00 (0.0%)	4.06 (0.7%)
			30	1.23 (0.0%)	1.26 (0.0%)	1.24 (0.0%)	1.26 (0.0%)
		0.016	10	5.84 (0.0%)	5.75 (0.1%)	5.86 (0.0%)	5.81 (0.1%)
			30	0.93 (0.0%)	0.94 (0.0%)	0.93 (0.0%)	0.94 (0.0%)
		0.090	10	2.56 (0.0%)	2.63 (0.1%)	2.59 (0.0%)	2.66 (0.1%)
			30	2.18 (0.0%)	2.19 (0.0%)	2.19 (0.0%)	2.20 (0.0%)
0.25-0.40	0.0-0.25	0.000	10	5.91 (0.0%)	5.91 (0.0%)	5.99 (0.0%)	5.96 (0.0%)
			30	1.48 (0.0%)	1.45 (0.0%)	1.48 (0.0%)	1.45 (0.0%)
		0.010	10	7.45 (0.0%)	7.74 (0.1%)	7.50 (0.0%)	7.84 (0.1%)
			30	1.01 (0.0%)	1.03 (0.0%)	1.01 (0.0%)	1.03 (0.1%)
		0.016	10	3.84 (0.0%)	3.98 (0.0%)	3.88 (0.0%)	4.03 (0.0%)
			30	0.77 (0.0%)	0.79 (0.0%)	0.78 (0.0%)	0.79 (0.0%)
		0.090	10	2.19 (0.0%)	2.27 (0.0%)	2.19 (0.0%)	2.31 (0.0%)
			30	1.50 (0.1%)	1.53 (0.2%)	1.52 (0.1%)	1.54 (0.2%)
	0.25-0.50	0.000	10	6.50 (0.0%)	6.51 (0.0%)	6.66 (0.0%)	6.54 (0.0%)
			30	0.60 (0.0%)	0.59 (0.0%)	0.60 (0.0%)	0.59 (0.0%)
		0.010	10	4.04 (0.0%)	4.15 (0.0%)	4.06 (0.0%)	4.17 (0.0%)
			30	1.16 (0.0%)	1.18 (0.0%)	1.17 (0.0%)	1.18 (0.0%)
		0.016	10	2.93 (0.0%)	3.10 (0.0%)	2.93 (0.0%)	3.12 (0.0%)
			30	2.37 (0.0%)	2.38 (0.0%)	2.49 (0.0%)	2.38 (0.0%)
		0.090	10	4.50 (0.0%)	4.56 (0.0%)	4.53 (0.0%)	4.63 (0.1%)
			30	0.86 (0.0%)	0.87 (0.0%)	0.86 (0.0%)	0.87 (0.0%)
0.40-0.60	0.0-0.25	0.000	10	3.30 (0.0%)	3.38 (0.2%)	3.33 (0.0%)	3.43 (0.3%)
			30	0.90 (0.0%)	0.92 (0.0%)	0.90 (0.0%)	0.92 (0.0%)
		0.010	10	8.35 (0.0%)	8.19 (0.5%)	8.53 (0.0%)	8.31 (0.5%)
			30	1.13 (0.0%)	1.12 (0.3%)	1.15 (0.0%)	1.12 (0.3%)
		0.016	10	5.98 (0.1%)	6.24 (0.7%)	6.00 (0.1%)	6.41 (0.8%)
			30	1.47 (0.0%)	1.49 (0.1%)	1.48 (0.0%)	1.50 (0.1%)
		0.090	10	7.78 (0.0%)	7.98 (0.1%)	7.89 (0.0%)	8.11 (0.1%)
			30	1.30 (0.0%)	1.31 (0.0%)	1.30 (0.0%)	1.32 (0.1%)
	0.25-0.50	0.000	10	3.91 (0.1%)	3.87 (0.0%)	4.03 (0.1%)	3.90 (0.1%)
			30	0.85 (0.0%)	0.85 (0.1%)	0.90 (0.0%)	0.85 (0.1%)
		0.010	10	8.35 (0.0%)	7.55 (0.0%)	8.90 (0.0%)	7.59 (0.0%)
			30	0.73 (0.0%)	0.71 (0.0%)	0.77 (0.0%)	0.71 (0.0%)
		0.016	10	2.27 (0.0%)	2.24 (0.3%)	2.34 (0.0%)	2.26 (0.3%)
			30	1.12 (0.0%)	1.11 (0.0%)	1.19 (0.0%)	1.11 (0.0%)
		0.090	10	2.01 (0.0%)	2.01 (0.4%)	2.07 (0.0%)	2.02 (0.6%)
			30	1.12 (0.0%)	1.09 (0.2%)	1.17 (0.0%)	1.09 (0.3%)

Tabela E.3: EQM ($\times 10^{-2}$) considerando variâncias diferentes para os erros de medida endógenos. Dados com distribuição simétrica.

ρ_{21}^C	ρ_{21}^T	n_0	\hat{T}^*	\hat{T}_e^*	\bar{T}	$2\hat{\tau}_C$
0.10-0.30	0.10-0.30	10	19.36	20.44	32.24	20.69
		30	7.01	7.13	12.56	7.14
	0.40-0.70	10	18.27	19.55	27.69	19.78
		30	5.51	5.63	8.21	5.62
0.40-0.60	0.10-0.30	10	18.10	19.10	25.91	19.27
		30	6.39	6.48	9.63	6.49
	0.40-0.70	10	14.96	15.92	20.91	16.01
		30	5.12	5.20	7.19	5.20
0.70-0.90	0.10-0.30	10	15.41	16.54	21.59	16.63
		30	5.47	5.58	7.62	5.58
	0.40-0.70	10	12.78	13.46	15.71	13.47
		30	4.56	4.65	5.78	4.66

A contribuição relativa do vício ao EQM em todas as celas é 0.0%.

Tabela E.4: EQM e contribuição relativa do vício ao EQM (entre parênteses) considerando variâncias diferentes para os erros de medida endógenos. Dados com distribuição assimétrica.

ρ_{21}^C	ρ_{21}^T	n_0	\hat{T}^*	\hat{T}_e^*	\bar{T}	$2\hat{\tau}_C$
0.10-0.25	0.0-0.25	10	4.04 (0.0%)	4.10 (0.1%)	4.05 (0.0%)	4.15 (0.1%)
		30	1.73 (0.0%)	1.74 (0.0%)	1.75 (0.0%)	1.74 (0.0%)
	0.25-0.40	10	5.29 (0.0%)	5.45 (0.1%)	5.30 (0.0%)	5.50 (0.2%)
		30	2.27 (0.0%)	2.29 (0.0%)	2.29 (0.0%)	2.29 (0.0%)
0.25-0.40	0.0-0.25	10	3.17 (0.0%)	3.17 (0.1%)	3.17 (0.0%)	3.20 (0.1%)
		30	3.26 (0.0%)	3.31 (0.0%)	3.29 (0.0%)	3.31 (0.0%)
	0.25-0.40	10	5.67 (0.0%)	5.85 (0.1%)	5.70 (0.0%)	5.93 (0.1%)
		30	0.73 (0.0%)	0.73 (0.0%)	0.73 (0.0%)	0.73 (0.0%)
0.40-0.60	0.0-0.25	10	4.95 (0.0%)	4.84 (0.7%)	5.06 (0.0%)	4.93 (0.8%)
		30	2.46 (0.0%)	2.49 (0.0%)	2.48 (0.0%)	2.49 (0.0%)
	0.25-0.40	10	3.48 (0.0%)	3.44 (0.1%)	3.71 (0.0%)	3.47 (0.1%)
		30	1.67 (0.0%)	1.65 (0.2%)	1.81 (0.0%)	1.65 (0.3%)

Tabela E.5: EQM ($\times 10^{-2}$) das estimativas considerando variâncias diferentes para os erros de medida exógenos. Dados com distribuição simétrica.

ρ_{21}^C	ρ_{21}^T	n_0	\tilde{T}	\tilde{T}_e	\bar{T}	$2\hat{\tau}_C$
0.10-0.30	0.10-0.30	10	18.81	19.80	30.29	19.99
		30	5.79	5.90	10.23	5.91
	0.40-0.60	10	17.90	18.85	28.05	19.08
		30	5.83	5.95	8.39	5.96
0.40-0.60	0.10-0.30	10	18.78	19.89	27.58	20.14
		30	5.86	5.96	8.28	5.96
	0.40-0.60	10	15.39	16.27	20.93	16.38
		30	4.31	4.39	6.54	4.39
0.70-0.90	0.10-0.30	10	15.54	16.46	20.25	16.59
		30	4.72	4.80	6.15	4.80
	0.40-0.60	10	12.28	13.11	15.44	13.07
		30	3.33	3.41	4.15	3.41

A contribuição relativa do vício ao EQM em todas as celas é 0.0%.

Tabela E.6: EQM e contribuição relativa do vício ao EQM (entre parênteses) das estimativas considerando variâncias diferentes para os erros de medida exógenos. Dados com distribuição assimétrica.

ρ_{21}^C	ρ_{21}^T	n_0	\tilde{T}	\tilde{T}_e	\bar{T}	$2\hat{\tau}_C$
0.10-0.25	0.0-0.25	10	4.67 (0.0%)	5.09 (0.0%)	4.71 (0.0%)	5.18 (0.0%)
		30	2.22 (0.0%)	2.21 (0.2%)	2.22 (0.0%)	2.21 (0.2%)
	0.25-0.40	10	10.84 (0.0%)	11.49 (0.1%)	10.87 (0.0%)	11.70 (0.1%)
		30	1.98 (0.0%)	2.00 (0.0%)	1.98 (0.0%)	2.00 (0.0%)
0.25-0.40	0.0-0.25	10	3.97 (0.0%)	3.98 (0.0%)	3.98 (0.0%)	4.04 (0.1%)
		30	2.90 (0.0%)	2.94 (0.2%)	2.93 (0.0%)	2.94 (0.2%)
	0.25-0.40	10	3.54 (0.0%)	3.56 (0.0%)	3.59 (0.0%)	3.60 (0.0%)
		30	0.83 (0.0%)	0.84 (0.0%)	0.84 (0.0%)	0.84 (0.0%)
0.40-0.60	0.0-0.25	10	2.89 (0.0%)	3.02 (0.3%)	2.91 (0.0%)	3.05 (0.3%)
		30	1.95 (0.0%)	1.96 (0.0%)	2.02 (0.0%)	1.96 (0.0%)
	0.25-0.40	10	2.98 (0.0%)	2.99 (0.0%)	3.07 (0.0%)	3.00 (0.1%)
		30	0.71 (0.0%)	0.71 (0.2%)	0.73 (0.0%)	0.72 (0.2%)

Apêndice F

Dados dos exemplos

Tabela F.1: Pressão arterial média (mmHg) de cães antes e depois de uma aplicação de $MgSO_4$.

Cão	Grupo	Antes	Depois	Cão	Grupo	Antes	Depois
1	Indometacina	148	132	13	Nifedipina	82	68
2	Indometacina	100	76	14	Nifedipina	90	72
3	Indometacina	120	108	15	Nifedipina	102	84
4	Indometacina	116	96	16	Nifedipina	110	100
5	Indometacina	140	128	17	Nifedipina	86	58
6	Indometacina	92	88	18	Nifedipina	90	76
7	Indometacina	112	100	19	Nifedipina	90	78
8	Indometacina	120	108	20	Nifedipina	84	72
9	Indometacina	128	100	21	Nifedipina	98	80
10	Indometacina	124	100	22	Nifedipina	116	80
11	Indometacina	108	92	23	Nifedipina	90	80
12	Indometacina	100	80	24	Nifedipina	94	76

Figura F.1: Boxplot da pressão arterial média (mmHg) de cães, antes (pré-teste) e depois (pós-teste) de uma aplicação de $MgSO_4$, e efeito da aplicação.

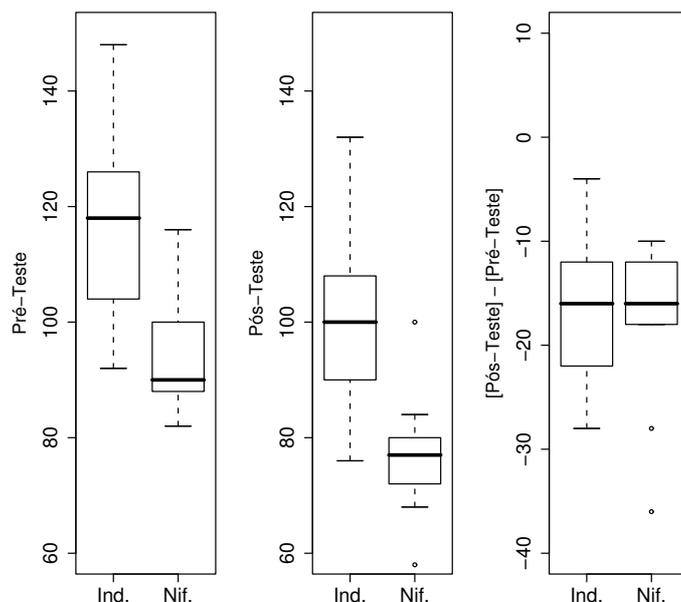


Tabela F.2: Resistência homogênea do sistema respiratório (RHSR) de ratos.

Rato	Gás	Pré-Teste	Pós-Teste	Rato	Gás	Pré-Teste	Pós-Teste
1	<i>T</i>	0.0828	0.0347	15	<i>C</i>	0.0477	0.0461
2	<i>T</i>	0.0274	0.0172	16	<i>C</i>	0.0313	0.0114
3	<i>T</i>	0.0115	0.0022	17	<i>C</i>	0.0437	0.0417
4	<i>T</i>	0.0112	0.0157	18	<i>C</i>	0.0358	0.0021
5	<i>T</i>	0.0065	0.0149	19	<i>C</i>	0.0477	0.0488
6	<i>T</i>	0.0025	0.0000	20	<i>C</i>	0.0579	0.0591
7	<i>T</i>	0.0357	0.0147	21	<i>C</i>	0.0664	0.0986
8	<i>T</i>	0.0232	0.0291	22	<i>C</i>	0.0871	0.0529
9	<i>T</i>	0.0356	0.0092	23	<i>C</i>	0.0517	0.0577
10	<i>T</i>	0.0458	0.0490	24	<i>C</i>	0.1102	0.1018
11	<i>T</i>	0.0836	0.0208	25	<i>C</i>	0.0776	0.0329
12	<i>T</i>	0.0638	0.0221	26	<i>C</i>	0.0970	0.0110
13	<i>T</i>	0.0909	0.0972	27	<i>C</i>	0.0157	0.0258
14	<i>T</i>	0.0281	0.0204	28	<i>C</i>	0.1203	0.1171
				29	<i>C</i>	0.0279	0.0176

Ventilação com *T*: Ar Sintético/Hélio-Oxigênio

Ventilação com *C*: Ar Sintético/Ar Sintético

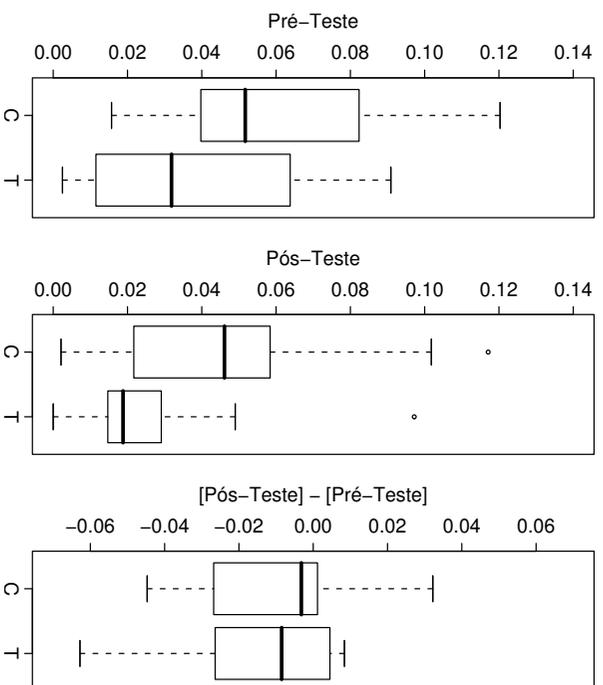


Figura F.2: Boxplot da resistência homogênea do sistema respiratório (RHSR) de ratos antes (pré-teste), depois (pós-teste) e efeito da ventilação com Ar Sintético, ou com Hélio e Oxigênio.

Tabela F.3: Notas em Matemática na FUVEST e habilidades em Cálculo (que assumimos conhecidas e iguais às notas de Cálculo I), de 104 estudantes do IME no ano 2004.

Est.	FUVEST	Cálculo	Est.	FUVEST	Cálculo	Est.	FUVEST	Cálculo
1	10	5.5	36	18	5.1	71	26	3.2
2	16	8.1	37	21	5.5	72	17	6.0
3	21	8.8	38	26	6.9	73	18	6.9
4	17	6.5	39	20	4.3	74	21	6.5
5	22	6.8	40	25	7.8	75	21	5.2
6	17	7.7	41	20	5.9	76	22	5.4
7	25	9.0	42	33	9.1	77	25	5.6
8	19	2.4	43	21	6.9	78	17	6.8
9	18	7.5	44	17	5.0	79	24	4.6
10	21	5.0	45	17	6.5	80	20	6.5
11	31	7.9	46	26	7.9	81	14	6.8
12	19	7.6	47	22	5.1	82	19	5.0
13	20	8.5	48	24	6.6	83	22	7.7
14	21	7.2	49	29	5.1	84	18	7.9
15	13	9.7	50	24	5.4	85	13	5.3
16	27	7.6	51	24	5.1	86	13	8.8
17	24	8.0	52	19	6.3	87	25	8.8
18	34	9.8	53	22	5.8	88	16	5.0
19	27	8.8	54	19	5.2	89	21	7.6
20	21	5.7	55	22	6.3	90	17	5.9
21	21	5.7	56	21	7.8	91	14	6.4
22	14	6.0	57	16	6.4	92	14	6.7
23	22	5.1	58	25	7.7	93	13	5.7
24	27	5.4	59	22	5.6	94	13	6.4
25	14	3.6	60	29	7.7	95	22	7.7
26	15	7.7	61	30	5.1	96	14	8.1
27	21	5.6	62	24	5.5	97	18	6.1
28	20	5.1	63	24	7.2	98	16	6.1
29	17	5.5	64	22	6.1	99	22	7.2
30	35	6.9	65	19	5.6	100	21	8.4
31	23	6.2	66	13	5.8	101	18	5.4
32	15	6.7	67	34	7.2	102	21	6.6
33	19	4.2	68	24	9.3	103	23	9.2
34	27	6.7	69	28	8.3	104	20	6.4
35	25	5.6	70	25	5.9			

Apêndice G

Interpretação da medida de associação entre a variável auxiliar e os valores latentes

A interpretação de B definida em (1.1) depende do enfoque sob o qual é assumido. Para interpretar B , consideremos o modelo linear não estocástico

$$y_s = A + Bx_s + e_s, \quad s = 1, \dots, N, \quad (\text{G.1})$$

em que $A = \mu_y - B\mu_x$, $e_s = y_s - \mu_y - B(x_s - \mu_x)$ e μ_y , μ_x e B estão definidos em (1.1). De um lado, sob o enfoque descritivo (freqüentista ou baseado no planejamento), Kish & Frankel (1974) definem os parâmetros A e B como os valores do vetor $(A \ B)'$ que minimizam a função $\sum_{s=1}^N (y_s - a - bx_s)^2$. Outros autores como Särndal, Swensson & Wretman (1992, p. 191) mencionam que “By way of interpretation, $(A \ B)'$ is the coefficient vector of the best fitting plane of the type $y = A + Bx$ ”, e também que “ B measures the number of units of y per unit of x in the finite population”. Além disso, assinalam que “An analyst interested in a regression slope, B , implicitly considers that a linear regression model is consistent with the finite population scatter”.

Por outro lado, sob o enfoque de super-população (baseado em modelos), no qual assume-se que a população finita é uma amostra aleatória ou realização de uma ‘super-população’ infinita, e que as observações das variáveis dependente e independente seguem o modelo linear clássico (ver, Pfeffermann & Holmes (1985) e Little (2004), por exemplo), considera-se que y_s^* é aleatório e que pode ser representado pela soma de duas quantidades, uma das quais é um componente determinístico $\alpha + \beta x_s$ e a outra é um componente aleatório ϵ_s , i.e.,

$$y_s^* = \alpha + \beta x_s + \epsilon_s, \quad s = 1, \dots, N, \quad (\text{G.2})$$

em que α e β são coeficientes conhecidos do modelo da regressão linear. Adicionamos o super-índice * a y_s para distinguir o seu caráter aleatório, induzido por ϵ_s , daquele de y_s definido em (G.1) que é uma quantidade fixa. No modelo (G.2), assume-se que o vetor $\epsilon = (\epsilon_1, \dots, \epsilon_s, \dots, \epsilon_N)'$ é tal que $\mathbb{E}(\epsilon) = \mathbf{0}$ e que sua matriz de covariâncias, \mathbf{V} , é definida positiva; portanto $\mathbb{E}_{\Psi}(y_s^*/x_s) = \alpha + \beta x_s$, $s = 1, \dots, N$, em que a esperança, $\mathbb{E}_{\Psi}(\cdot)$, é avaliada com respeito ao modelo de super-população parametrizado por $\Psi = (\alpha, \beta, \mathbf{V})$. Em particular, no caso de erros homocedásticos, tem-se $\mathbf{V} = \sigma^2 \mathbf{I}_N$, $\sigma^2 > 0$, em que \mathbf{I}_N denota a matriz identidade da ordem N . Conseqüentemente, Neter, Kutner, Nachtsheim & Wasserman (1996, p. 12) interpretam o parâmetro β como “the change in the mean of the probability distribution of y^* per unit increase in x ” e Sen & Srivastava (1990, p. 6) assinalam que “for every unit increase in x , y^* increases by β ”.

Sob o enfoque de populações finitas, as interpretações apresentadas acima podem ser não adequadas porquanto a média de y é uma constante que não depende de x . Alternativamente, podemos escrever

$$B = \sum_{\substack{s=1 \\ x_s \neq \mu_x}}^N \frac{(x_s - \mu_x)^2}{\sum_{s=1}^N (x_s - \mu_x)^2} \times \frac{(y_s - \mu_y)}{(x_s - \mu_x)} = \sum_{s=1}^N w_s b_s,$$

em que $w_s = (x_s - \mu_x)^2 / \sum_{s=1}^N (x_s - \mu_x)^2$ e $b_s = (y_s - \mu_y) / (x_s - \mu_x)$ quando $x_s \neq \mu_x$. Logo pode-se considerar B definida em (1.1) como uma média ponderada das N inclinações, b_s , dos segmentos de reta determinados pelos pontos (x_s, y_s) e (μ_x, μ_y) . Essas inclinações podem ser interpretadas como as medidas de variação de y por unidade de variação de x . Deste modo, B pode ser interpretado como uma medida de associação baseada numa média ponderada das variações, em que os pesos w_s dependem do afastamento entre x_s e μ_x .

Referências Bibliográficas

- Aoki, R., Achcar, J. A., Bolfarine, H. & Singer, J. M. (2003). Bayesian Analysis of Null Intercept Errors-in-Variables Regression for Pretest-Posttest data. *Journal of Applied Statistics*, **30**, 3-12. [5](#)
- Bolfarine, H. (1991). Finite-population prediction under error-in-variables superpopulation models. *The Canadian Journal of Statistics*, **19**, 191-207. [2](#), [23](#)
- Bolfarine, H. & Zacks, S. (1992). *Prediction Theory for Finite Populations*. New York: Springer-Verlag. [2](#), [3](#), [23](#), [36](#), [37](#)
- Bolfarine, H., Zacks, S., Elian, S. & Rodrigues, J. (1994). Optimal Prediction of the Finite Population Regression Coefficient. *Sankhyā B*, **56**, 1-10. [3](#)
- Bonate, P. L. (2000). *Analysis of Pretest-Posttest Designs*. Boca Raton: Chapman & Hall/CRC. [5](#), [8](#), [54](#), [70](#)
- Brogan, D. R. & Kutner, M. H. (1980). Comparative Analysis of Pretest-Posttest Research Designs. *The American Statistician*, **34**, 229-232. [5](#)
- Cassel, C. M., Särndal, C. E. & Wretman, J. H. (1976). Some Results on Generalized Difference Estimation and Generalized Regression Estimation for Finite Populations. *Biometrika*, **63**, 615-620. [2](#), [23](#), [36](#)
- Chen, J. & Qin, J. (1993). Empirical Likelihood Estimation for Finite Populations and the Effective Usage of Auxiliary Information. *Biometrika*, **80**, 107-116. [2](#)
- Cumberland, W. G. & Royall, R. M. (1988). Does Simple Random Sampling Provide Adequate Balance?. *Journal of the Royal Statistical Society, Series B*, **50**, 118-124. [37](#)
- Deville, J. C. & Särndal, C. E. (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, **87**, 376-382. [2](#)
- Fuller, W. A. (1975). Regression Analysis for Sample Survey. *Sankhyā C*, **37**, 117-132. [2](#)
- Harville, D. A. (1997). *Matrix Algebra From A Statistician's Perspective*. New York: Springer. [14](#)
- Henderson, C. R., Kempthorne, O., Searle, S. R. & von Krosigk, C. M. (1959). The Estimation of Environmental and Genetic Trends From Records Subject to Culling. *Biometrics*, **15**, 192-218. [23](#), [36](#)

- Hinkelmann, K. & Kempthorne, O. (1994). *Design and Analysis of Experiments. Volume I. Introduction to Experimental Design*. New York: John Wiley & Sons, Inc. 55
- Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, **81**, 945-970. 55
- Holt, D., Smith, T. M. F. & Winter, P. D. (1980). Regression Analysis of Data from Complex Surveys. *Journal of the Royal Statistical Society, Series A*, **143**, 474-487. 2
- Kish, L. & Frankel, M. R. (1974). Inference from Complex Samples (with discussion). *Journal of the Royal Statistical Society, Series B*, **36**, 1-37. 2, 115
- Knoke, J. D. (1991). Nonparametric Analysis of Covariance for Comparing Change in Randomized Studies with Baseline Values Subject to Error. *Biometrics*, **47**, 523-533. 5, 10
- Laird, N. (1983). Further Comparative Analyses of Pretest-Posttest Research Designs. *The American Statistician*, **37**, 329-330. 5
- Lencina, V. B., Singer, J. M. & Stanek III, E. J. (2005). Much Ado About Nothing: the Mixed Models Controversy Revisited. *International Statistical Review*, **73**, 9-20. 7
- Leon, S., Tsiatis, A. A. & Davidian, M. (2003). Semiparametric Estimation of Treatment Effect in a Pretest-Posttest Study. *Biometrics*, **59**, 1046-1055. 5, 55
- Li, W., Stanek III, E. J. & Singer, J. M. (2008). Design-Based Random Permutation Models with Auxiliary Information. (Submitted). 2, 11, 15, 17, 23, 35
- Little, R. J. (2004). To Model or Not to Model? Competing Modes of Inference for Finite Population Sampling. *Journal of the American Statistical Association*, **99**, 546- 556. 115
- Morrison, D. F. (1983). *Applied Linear Statistical Methods*. New Jersey: Prentice-Hall. 35
- Neter, J., Kutner, M. H., Nachtsheim, C. J. & Wasserman, W. (1996). *Applied Linear Statistical Models*. Boston: WCB/McGraw-Hill. 116
- Pfeffermann, D. & Holmes, D. J. (1985). Robustness Considerations in the Choice of a Method of Inference for Regression Analysis of Survey Data. *Journal of the Royal Statistical Society, Series A*, **148**, 268-278. 115
- Pfeffermann, D. & Smith, T. M. F. (1985). Regression Models for Grouped Populations in Cross-Section Surveys. *International Statistical Review*, **53**, 37-59. 2
- Robinson, G. K. (1991). That BLUP Is a Good Thing: The Estimation of Random Effects. *Statistical Science*, **6**, 15-51. 23
- Rubin, D. B. (2005). Causal Inference Using Potential Outcomes: Design, Modeling, Decisions. *Journal of the American Statistical Association*, **100**, 322-331. 55
- Sen, A. & Srivastava, M. (1990). *Regression Analysis. Theory, Methods, and Applications*. New York: Springer-Verlag. 116

- Singer, J. M. & Andrade, D. F. (1997). Regression Models for the Analysis of Pretest/Posttest Data. *Biometrics*, **53**, 729-735. [5](#)
- Singer, J. M., Seoanes, M. & Ogando, M. A. (1988). Estudo do efeito da infusão aguda de $MgSO_4$ nas funções hemodinâmicas e renais em cães. São Paulo - IME - USP, SEA-Relatório de Análise Estatística. [4](#)
- Särndal, C. E., Swensson, B. & Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag. [2](#), [3](#), [23](#), [36](#), [115](#)
- Stanek III, E. J. (1988). Choosing a Pretest-Posttest Analysis. *The American Statistician*, **42**, 178-183. [5](#)
- Stanek III, E. J., Lencina, V. B., Singer, J. M., González, L. M., Li, W. & San Martino, S. (2008). Predicting the Latent Value of a Sample Subject with Mixed Models: What does σ_i^2 mean?. (Submitted). [2](#), [11](#), [26](#), [35](#), [74](#)
- Stanek III, E. J. & Singer, J. M. (2004). Predicting Random Effects From Finite Population Clustered Samples With Response Error. *Journal of the American Statistical Association*, **99**, 1119-1130. [1](#), [2](#), [11](#), [26](#), [35](#)
- Stanek III, E. J. & Singer, J. M. (2008). Predicting Random Effects with an Expanded Finite Population Mixed Model. In Press in *Journal of Statistical Planning and Inference*. [48](#), [63](#)
- Stanek III, E. J., Singer, J. M. & Lencina, V. B. (2004). A Unified Approach to Estimation and Prediction under Simple Random Sampling. *Journal of Statistical Planning and Inference*, **121**, 325-338. [11](#), [13](#)
- Voss, D. T. (1999). Resolving the Mixed Models Controversy. *The American Statistician*, **53**, 352-356. [7](#)
- Wu, C. & Sitter, R. R. (2001). A Model-Calibration Approach to Using Complete Auxiliary Information From Survey Data. *Journal of the American Statistical Association*, **96**, 185-193. [2](#), [23](#), [36](#)
- Yang, L. & Tsiatis, A. A. (2001). Efficiency Study of Estimators for a Treatment Effect in a Pretest-Posttest Trial. *The American Statistician*, **55**, 314-321. [5](#)