

Mirian Martins Ribeiro

**Mortalidade adulta por níveis de escolaridade
no estado e no município de São Paulo: uma
proposta de estimação a partir do Censo
Demográfico de 2010**

Belo Horizonte, MG
UFMG/Cedeplar
Agosto de 2016

Mirian Martins Ribeiro

**Mortalidade adulta por níveis de escolaridade no estado
e no município de São Paulo: uma proposta de estimação
a partir do Censo Demográfico de 2010**

Tese apresentada ao curso de Doutorado em Demografia do Centro de Desenvolvimento e Planejamento Regional da Faculdade de Ciências Econômicas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do Título de Doutor em Demografia.

Orientador: Prof. Cássio Maldonado Turra

Co-orientadora: Prof^a. Cristine Campos de Xavier Pinto

Belo Horizonte, MG
Centro de Desenvolvimento e Planejamento Regional
Faculdade de Ciências Econômicas - UFMG
Agosto de 2016

Ficha Catalográfica

Ribeiro, Mirian Martins.

R484m Mortalidade adulta por níveis de escolaridade no estado e no
2016 município de São Paulo [manuscrito] : uma proposta de
 estimação a partir do censo demográfico de 2010 / Mirian
 Martins Ribeiro. – 2016.

xiv, 122 f. : il., tabs.

Orientador: Cássio Maldonado Turra.
Coorientadora: Cristine Campos de Xavier Pinto.
Tese (doutorado) - Universidade Federal de Minas Gerais,
Centro de Desenvolvimento e Planejamento Regional.
Inclui bibliografia (f. 108-117).

1. Mortalidade – São Paulo (Estado) – Teses. 2. São Paulo
(Estado) – Censo demográfico – Teses. I. Turra, Cássio
Maldonado. II. Pinto, Cristine Campos de Xavier.
III. Universidade Federal de Minas Gerais. Centro de
Desenvolvimento e Planejamento Regional. IV. Título.

CDD: 304.6009816



Curso de Pós-Graduação em Demografia da Faculdade de Ciências Econômicas

ATA DE DEFESA DE TESE DE MIRIAN MARTINS RIBEIRO Nº. REGISTRO 2011661220. Às quatorze horas do dia oito de agosto de dois mil e dezesseis, reuniu-se na *Faculdade de Ciências Econômicas* da Universidade Federal de Minas Gerais a Comissão Examinadora de TESE, indicada “*ad referendum*” pelo Colegiado do Curso em 02/08/2016, para julgar, em exame final, o trabalho final intitulado “**Mortalidade adulta por níveis de escolaridade no estado e no município de São Paulo: uma proposta de estimação a partir do Censo Demográfico de 2010**”, requisito final para a obtenção do Grau de *Doutor em Demografia*, área de concentração em Demografia. Abrindo a sessão, o Presidente da Comissão, Prof. Cássio Maldonado Turra, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata, para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores, com a respectiva defesa da candidata. Logo após, a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição do resultado final. A Comissão aprovou a candidata por unanimidade. O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora. Belo Horizonte, 08 de agosto de 2016.

Prof. Cássio Maldonado Turra
(Orientador) (CEDEPLAR/FACE/UFGM)

Profa. Cristine Campos de Xavier Pinto
(Coorientadora) (FGV/SP)

Profa. Simone Wajnman
(CEDEPLAR/FACE/UFGM)

Prof. Bernardo Lanza Queiroz
(CEDEPLAR/FACE/UFGM)

Profa. Marília Miranda Forte Gomes
(UnB)

Prof. Flávio Henrique Miranda de Araújo Freire
(UFRN)

Prof. José Irineu Rangel Rigotti
 Coordenador do Curso de Pós-Graduação
 em Demografia

AGRADECIMENTOS

Deixo meus agradecimentos aos que contribuíram direta ou indiretamente com minha formação.

Ao Cássio, pela forma como conduziu a orientação deste trabalho. Pela competência, cobrança e estímulo combinados com respeito às minhas limitações, que foram muitas durante o processo de desenvolvimento da tese. À Cristine, por aceitar essa coorientação e atender às minhas dificuldades com paciência e cordialidade inacreditáveis. Meu carinho por vocês é imenso.

À professora Ignez Perpetuo pela orientação e apoio na graduação e no mestrado; por ter proporcionado minhas primeiras experiências em pesquisa.

À professora Mônica Viegas, pelas oportunidades de trabalho em pesquisas tão interessantes nos caminhos da economia da saúde, que contribuíram para a minha continuidade na vida acadêmica e para minha formação no Doutorado.

À professora Carla Jorge, por permitir minha participação em trabalho por ela idealizado e por ser responsável pelo meu interesse nos temas relacionados à mortalidade.

Ao professor Roberto Nascimento pelo carinho e respeito com seus alunos; por nos incentivar a priorizar qualidade de vida e estabilidade emocional.

À Professora Simone, minha orientadora de curso, a quem recorri em momentos de angústia e dúvidas, sendo sempre bem recebida.

Ao Professor Irineu, coordenador da Demografia, pelo seu profissionalismo e sensibilidade no atendimento às minhas dificuldades.

Aos funcionários do Cedeplar, em especial Cecília, Sebastião e Carol.

À Universidade Federal dos Vales do Jequitinhonha e Mucuri, em especial ao Departamentos de Ciências Exatas e de Administração e à Direção da FACSAB, pela licença concedida para minha qualificação.

À Banca de qualificação: Simone, Bernardo e Flávio pelos questionamentos e contribuições que resultaram em reflexões importantes para o fechamento da tese.

À Ana Carolina Maia, pela amizade e oportunidades de trabalho ao longo dos meus anos no Cedeplar.

Aos amigos Daniel, Claudinha Aguiar, Heder, Laurinha e Virna pelo companheirismo, carinho, sinceridade e pelos doces momentos de alegria que compartilhamos ao longo de tantos anos de convivência.

Aos amigos de Téofilo Otoni, Geórgia, Gustavo, Ivana, Marconi e Mirelle, pelo acolhimento na cidade e na UFVJM, pela amizade e pelo alento em momentos difíceis pela distância da família e de outros amigos.

Ao papai e a mamãe, pelo amor, cuidado e apoio incondicional às nossas escolhas. Aos meus queridos irmãos Marcela e Átila pelo companheirismo e união em momentos de alegria, euforia e aflições.

Ao tio Argentino, à tia Rachel e à Vovó Leopoldina (em lembrança) por me receberem quando me mudei para Belo Horizonte para estudar.

Ao Eltinho, pelo companheirismo, compreensão, carinho e amor.

Ao CNPq pelas bolsas concedidas.

LISTA DE ABREVIATURAS E SIGLAS

DO - Declaração de óbito

EM - Expectativa-Maximização

IBGE - Instituto Brasileiro de Geografia e Estatística

IM - Imputação Múltipla

MCMC - Método de Monte Carlo via cadeias de Markov

MS - Ministério da Saúde

MV - Máxima Verosimilhança

MVN - Modelo Multivariado Normal

SIM - Sistema de Informações sobre Mortalidade

SUMÁRIO

ÍNDICE DE FIGURAS E QUADROS	X
ÍNDICE DE TABS	XII
RESUMO	XIV
ABSTRACT	XV
1. INTRODUÇÃO	1
2. ASPECTOS TEÓRICOS E EMPÍRICOS DO GRADIENTE EDUCACIONAL DA MORTALIDADE	5
2.1. CONTEXTO.....	5
2.2. EVIDÊNCIAS DO GRADIENTE EDUCACIONAL DA MORTALIDADE EM PAÍSES DESENVOLVIDOS.....	7
2.3. EVIDÊNCIAS DO GRADIENTE EDUCACIONAL DA MORTALIDADE EM PAÍSES EM DESENVOLVIMENTO	11
2.4. EVIDÊNCIAS DO GRADIENTE EDUCACIONAL DA MORTALIDADE NO BRASIL.....	15
2.5. MECANISMOS DE ASSOCIAÇÃO ENTRE ESCOLARIDADE, SAÚDE E MORTALIDADE ADULTA	19
2.5.1. <i>Os modelos econômicos</i>	22
2.5.2. <i>A abordagem psicossocial</i>	23
3. DADOS FALTANTES: CONCEITOS, PREMISSAS E MÉTODOS DE IMPUTAÇÃO	26
3.1. CONCEITOS E PREMISSAS.....	26
3.2. PRINCIPAIS TÉCNICAS DE IMPUTAÇÃO	29
3.3. IMPUTAÇÃO MÚLTIPLA (IM).....	31
3.2.1. <i>Procedimento geral para imputação múltipla</i>	32
3.2.2. <i>Escolha do modelo</i>	33
3.2.3. <i>Etapa de imputação com dado faltante ignorável</i>	35
3.2.4. <i>Considerações sobre o número de imputações e de modelos</i>	36
3.2.6. <i>Limitações da IM</i>	38
3.4. O ALGORITMO EM.....	38
4. BASE DE DADOS E METODOLOGIA.....	41
4.1. FONTE DOS DADOS BÁSICOS.....	41
4.1.1. CENSO DEMOGRÁFICO 2010.....	41
4.1.2. SISTEMA DE INFORMAÇÕES SOBRE MORTALIDADE - SIM	42
4.2. POPULAÇÃO DE INTERESSE E DIMENSÃO DA AMOSTRA	43
4.3. APLICAÇÃO DA IMPUTAÇÃO MÚLTIPLA PARA OS DADOS FALTANTES DE ESCOLARIDADE	46
4.4. APLICAÇÃO O ALGORITMO EM PARA O DADO FALTANTE DE ESCOLARIDADE.....	49
4.5. SELEÇÃO DAS VARIÁVEIS DOS MODELOS	49
4.6. VALIDADE DOS DADOS IMPUTADOS	54
4.7. MODELO ANALÍTICO PARA MORTALIDADE	55
4.8. CORREÇÃO DO NÍVEL DOS ÓBITOS NO CENSO	56
5. RESULTADOS DAS IMPUTAÇÕES DA ESCOLARIDADE.....	57
5.1. IMPUTAÇÃO POR <i>HOTDECK</i> : CONSISTÊNCIA DOS RESULTADOS	57

5.2.	IMPUTAÇÃO MÚLTIPLA COM REGRESSÃO LOGÍSTICA ORDENADA: CONSISTÊNCIA DOS RESULTADOS	60
5.3.	IMPUTAÇÃO POR REGRESSÃO MULTIVARIADA NORMAL (MVN) COM ALGORITMO EM E MCMC: CONSISTÊNCIA DOS RESULTADOS	74
5.4.	DISTRIBUIÇÕES DAS ESCOLARIDADES DO RESPONSÁVEL PELO DOMICÍLIO E DAS IMPUTADAS POR IM E MNV	80
5.5.	DISTRIBUIÇÕES DA ESCOLARIDADE IMPUTADA, DO RESPONSÁVEL PELO DOMICÍLIO E A OBSERVADA NO SIM	81
5.6.	CONSIDERAÇÕES FINAIS	86
6.	MORTALIDADE POR NÍVEIS DE ESCOLARIDADE	88
6.1.	MORTALIDADE FEMININA NO MUNICÍPIO DE SÃO PAULO: ESTIMATIVAS DE DIFERENCIAIS POR ESCOLARIDADE IMPUTADA, POR ESCOLARIDADE DO CHEFE E ESCOLARIDADE DO SIM.....	88
6.2.	MORTALIDADE MASCULINA NO MUNICÍPIO DE SÃO PAULO: ESTIMATIVAS DE DIFERENCIAIS POR ESCOLARIDADE IMPUTADA, POR ESCOLARIDADE DO CHEFE E ESCOLARIDADE DO SIM	94
6.3.	DIFERENCIAIS DE MORTALIDADE POR NÍVEL DE ESCOLARIDADE POR IDADE E SEXO: RESULTADOS COM CORREÇÃO PARA SUB E SOBRE ENUMERAÇÃO DOS ÓBITOS	98
6.4.	DISCUSSÃO	102
7.	CONSIDERAÇÕES FINAIS	106
	APÊNDICE A	118

ÍNDICE DE FIGURAS E QUADROS

FIGURA 1: Proporção de observações segundo a diferença entre escolaridade observada e imputada, <i>hotdeck</i> com 5 repetições, subamostra de 10000 pessoas, estado de São Paulo, 2010 ...	58
FIGURA 2: Proporção de observações segundo a diferença entre escolaridade observada e imputada, imputação múltipla com regressão logística ordenada (IM1 a IM9), subamostra de 10000 pessoas, São Paulo, 2010.....	65
FIGURA 3: Distribuição das mulheres falecidas e vivas por escolaridade segundo a idade, IM modelo IM6, 3 categorias de escolaridade, estado de São Paulo, 2010	70
FIGURA 4: Distribuição dos homens falecidos e vivos por escolaridade segundo a idade, IM modelo IM6, 3 categorias de escolaridade, estado de São Paulo, 2010	71
FIGURA 5: Distribuição das mulheres falecidas e vivas por escolaridade segundo idade, IM modelo IM9, 4 categorias de escolaridade, estado de São Paulo, 2010	72
FIGURA 6: Distribuição dos homens falecidos e vivos por escolaridade segundo idade, IM modelo IM9, 4 categorias de escolaridade, estado de São Paulo, 2010	73
FIGURA 7: Distribuição das mulheres falecidas e vivas por escolaridade segundo idade, imputação MVN, 3 categorias de escolaridade, estado de São Paulo, 2010.....	76
FIGURA 8: Distribuição dos homens falecidos e vivos por escolaridade segundo idade, imputação MVN, 3 categorias de escolaridade, estado de São Paulo, 2010.....	77
FIGURA 9: Distribuição das mulheres falecidas e vivas por escolaridade segundo idade, imputação MVN, 4 categorias de escolaridade, estado de São Paulo, 2010.....	78
FIGURA 10: Proporção de homens falecidos e vivos por escolaridade segundo idade, imputação MVN, 4 categorias de escolaridade, estado de São Paulo, 2010.....	79
FIGURA 11: Distribuição percentual dos óbitos de pessoas entre 25 e 59 anos segundo grupos etários, populações femininas do estado e município de São Paulo, 2010, 2012-2013	89
FIGURA 12: Taxas de mortalidade (por mil) segundo escolaridade do responsável pelo domicílio, escolaridade imputada por MVN e observada no SIM, população feminina do município de São Paulo, 2010, 2012-2013	91
FIGURA 13: Taxas de mortalidade (por mil) e intervalos de confiança estimados (IC) segundo escolaridade do responsável pelo domicílio, escolaridade imputada por MVN e observada no SIM, população feminina do município de São Paulo, 2010, 2012-2013	92
FIGURA 14: Razão entre as taxas de mortalidade estimadas pelo Censo, por escolaridade do responsável pelo domicílio e escolaridade imputada por MVN, e as taxas do SIM, população feminina município de São Paulo, 2010, 2012-2013	93
FIGURA 15: Distribuição percentual dos óbitos de pessoas entre 25 e 59 anos segundo grupos etários, populações masculinas do estado e município de São Paulo, 2010, 2012-2013	94
FIGURA 16: Taxas de mortalidade (por mil) segundo escolaridade do responsável pelo domicílio, escolaridade imputada por MVN e observada no SIM, população masculina do município de São Paulo, 2010, 2012-2013	96
FIGURA 17: Taxas de mortalidade (por mil) e intervalos de confiança estimados (IC) segundo escolaridade do responsável pelo domicílio, escolaridade imputada por MVN e observada no SIM, população masculina do município de São Paulo, 2010, 2012-2013	97

FIGURA 18: Razão entre as taxas de mortalidade estimadas pelo Censo, por escolaridade do responsável pelo domicílio e escolaridade imputada por MVN, e as taxas do SIM, população masculina do município de São Paulo, 2010, 2012-2013	98
FIGURA 19: Taxas de mortalidade por idade e sexo segundo níveis de escolaridade, estado e município de São Paulo, 2010.....	101
FIGURA 20: Razão entre as taxas de mortalidade masculinas e femininas por idade e sexo, segundo níveis de escolaridade, estado e município de São Paulo, 2010	101
FIGURA 21: Razão entre as taxas de mortalidade dos grupos de menor e maior escolaridade segundo idade e sexo, estado e município de São Paulo, 2010.....	102

ÍNDICE DE TABS

TABELA 1: Proporção de pessoas em domicílios unipessoais segundo idade e sexo, estado de São Paulo, 2010	44
TABELA 2: Número de pessoas e domicílios por área de ponderação, Brasil e Unidades da Federação, 2010	45
TABELA 3: Número de pessoas e óbitos entre 25 e 59 anos por área de ponderação, Brasil e Unidades da Federação, 2010	46
TABELA 4: Percentual de perda de informação de escolaridade por idade e sexo, estado de São Paulo, 2010	48
TABELA 5: Escolaridade média observada e imputada, <i>hotdeck</i> com escolaridade discreta, subamostra de 10000 pessoas, estado de São Paulo, 2010	59
TABELA 6: Distribuição dos indivíduos segundo escolaridade imputada e observada, <i>hotdeck</i> com 3 categorias de escolaridade, subamostra de 10000 pessoas, estado de São Paulo, 2010	60
TABELA 7: Distribuição dos indivíduos segundo escolaridade imputada e observada, <i>hotdeck</i> com 4 categorias de escolaridade, subamostra de 10000 pessoas, estado de São Paulo, 2010	60
TABELA 8: Modelos de regressão logística ordenada para imputação, 3 categorias de escolaridade, IM1 a IM3	62
TABELA 9: Modelos de regressão logística ordenada para imputação, 3 categorias de escolaridade, IM4 a IM6	63
TABELA 10: Modelos de regressão logística ordenada para imputação, 4 categorias de escolaridade, IM7 a IM9	64
TABELA 11: Distribuição dos indivíduos segundo escolaridade imputada e observada por sexo, IM com regressão logística com 3 categorias de escolaridade (IM6), para uma subamostra de 10000 pessoas, estado de São Paulo, 2010	66
TABELA 12: Distribuição dos indivíduos segundo escolaridade imputada e observada por sexo, IM com regressão logística com 4 categorias de escolaridade (IM9), subamostra de 10000 pessoas, estado de São Paulo, 2010	66
TABELA 13: Distribuição dos indivíduos segundo escolaridade imputada e observada por idade, IM com regressão logística com 3 categorias de escolaridade (IM6), subamostra de 10000 pessoas, estado de São Paulo, 2010	67
TABELA 14: Distribuição dos indivíduos segundo escolaridade imputada e observada por idade, IM com regressão logística com 4 categorias de escolaridade (IM9), subamostra de 10000 pessoas, estado de São Paulo, 2010	68
TABELA 15: Escolaridade média com dados imputados e observados, por sexo, imputação por MVN, subamostra de 10000 pessoas, estado de São Paulo, 2010	74
TABELA 16: Escolaridade média com dados imputados e observados, por grupos etários, imputação por MNV, subamostra de 10000 pessoas, estado de São Paulo, 2010	75
TABELA 17: Distribuição percentual dos óbitos segundo escolaridade observada e imputada por IM6 e MVN para os responsáveis pelo domicílio, segundo sexo, 3 categorias de escolaridade, estado de São Paulo, 2010	80

TABELA 18: Distribuição percentual dos óbitos segundo escolaridade observada e imputada por IM6 e MVN para os responsáveis pelo domicílio, segundo o sexo, estado de São Paulo, 2010.....	81
TABELA 19: Distribuição percentual dos óbitos por escolaridade do responsável pelo domicílio e observada no SIM, 3 categorias de escolaridade, estado e município de São Paulo, 2010.....	82
TABELA 20: Distribuição percentual dos óbitos por escolaridade do responsável pelo domicílio e observada no SIM, 4 categorias de escolaridade, estado e município de São Paulo, 2010.....	83
TABELA 21: Distribuição percentual dos óbitos por escolaridade imputada por modelo IM6 e observada no SIM, 3 categorias de escolaridade, estado e município de São Paulo, 2010.....	84
TABELA 22 Distribuição percentual dos óbitos por escolaridade imputada por modelo IM6 e observada no SIM, estado e município de São Paulo, 2010	84
TABELA 23: Distribuição percentual dos óbitos por escolaridade imputada por MVN e observada no SIM, 3 categorias de escolaridade, estado e município de São Paulo, 2010.....	85
TABELA 24: Distribuição percentual dos óbitos por escolaridade imputada por MVN e observada no SIM, 4 categorias de escolaridade, estado e município de São Paulo, 2010.....	86
TABELA 25: Distribuição dos óbitos de pessoas entre 25 e 59 anos, estado e município de São Paulo 2010, 2012-2013	87
TABELA 26: Resultado dos coeficientes e razões das taxas de incidência (IRR) dos modelos de Poisson, população feminina do município de São Paulo, 2010, 2012-2013.....	90
TABELA 27: Resultado dos coeficientes e razões das taxas de incidência (IRR) dos modelos de Poisson, população masculina do município de São Paulo, 2010, 2012-2013.....	95
TABELA 28: Resultado dos coeficientes e razões das taxas de incidência (IRR) dos modelos de Poisson, com níveis de mortalidade ajustados por idade sexo e educação, população feminina do estado e do município de São Paulo, 2010, 2012-2013	99
TABELA 29: Resultado dos coeficientes e razões das taxas de incidência (IRR) dos modelos de Poisson, com níveis de mortalidade ajustados por idade sexo e educação, população masculina do estado e do município de São Paulo, 2010, 2012-2013.....	100
QUADRO 1: Regras de conversão do campo escolaridade em séries para escolaridade em anos, para transição do uso dos modelos antigos e novos da declaração de óbito	43
QUADRO 2: Variáveis selecionadas para as regressões	50
QUADRO 3: Matriz de correlações tetracóricas das variáveis componentes do indicador socioeconômico para domicílio.....	53

RESUMO

O objetivo da tese foi investigar diferenciais educacionais da mortalidade adulta a partir dos dados da amostra do Censo Demográfico de 2010. Trata-se de uma proposta de estimativa a partir de uma base de dados alternativa à principal fonte de estudos de mortalidade no Brasil, o Sistema de Informações sobre Mortalidade (SIM) do Ministério da Saúde. Dentre as vantagens em usar o Censo, destacam-se: disponibilidade dados (microdados) tanto de sobreviventes quanto de óbitos, dentro de um mesmo período de referência; e variedade de informações sobre domicílios e indivíduos sobreviventes conviventes com os que sofreram óbitos. Por outro lado, as únicas informações individuais para os mortos são idade e sexo. Diante desta limitação, uma proposta secundária da tese foi apresentar uma metodologia de imputação da escolaridade dos indivíduos que sofreram óbito, utilizando imputação múltipla por regressão logística ordenada e imputação por modelo multivariado normal (MVN) usando algoritmo *Expectation-Maximization* (EM) e simulações. Para isso, foi selecionada a amostra do estado de São Paulo para desenvolver e validar a metodologia, para que futuramente ela possa ser aplicada ao Brasil. O modelo MVN se mostrou mais adequado na análise de consistência e validação dos métodos, etapa que contou com os dados de mortalidade do SIM para ser realizada. Após imputação da escolaridade dos mortos, as taxas de mortalidade por idade, sexo e escolaridade foram estimadas usando coeficientes obtidos por regressão de Poisson. Os resultados apontaram gradiente educacional da mortalidade significativo, sobretudo para os homens e para o município de São Paulo, comparativamente ao estado. Um ponto importante foi o comportamento do diferencial educacional da mortalidade por idade. A partir da revisão de literatura, uma hipótese era de um diferencial decrescente com idade. No entanto, o diferencial foi praticamente constante por idade. Para explorar esses resultados torna-se necessário entender a forma com a transição da mortalidade tem ocorrido no estado e como as medidas de período podem influenciar os resultados. É importante entender as relações entre período, idade e coorte para analisar o comportamento dos diferenciais de mortalidade por escolaridade. Além disso, deve-se levar em conta como o aumento do acesso à escolaridade tem elevado o nível educacional das coortes mais jovens no Brasil. Deve-se avançar nessas discussões, que não foram privilegiadas na tese. Quanto ao uso do Censo para estudos de diferenciais de mortalidade, os resultados dão uma perspectiva otimista. Embora exista uma expectativa de melhoria da cobertura e qualidade das informações de escolaridade do SIM, o Censo pode ser uma importante fonte para estudos de diferenciais socioeconômicos da mortalidade por ofertar outras características que podem complementar a análise do efeito da escolaridade.

Palavras-chave: diferenciais de mortalidade; mortalidade adulta; censo demográfico

ABSTRACT

The aim of the thesis was to investigate differentials in adult mortality based on educational levels from the demographic sample data of 2010 Census. This is a draft estimate from an alternative database to the main source of mortality studies in Brazil, a Mortality Information System (SIM) of the Ministry of Health. Among the highlighted advantages of using the Census, are: the availability of data (micro data) on both the living and the dead within a reference period and variety of information on cohabiting households and surviving individuals with those who suffered death. On the other hand, the only personal information on the dead are age and sex. Given this limitation, a secondary proposal of the thesis was to present an allocation methodology on educational levels of individuals who have suffered death by using multiple imputation ordered logistics regression and imputation by normal multivariate model (MVN) using Expectation-Maximization (EM) algorithm and simulations. To achieve this, a sample was selected from the State of Sao Paulo to develop and validate the methodology so as to make it applicable in Brazil in the future. The MVN model was more appropriate in the consistency and validation analysis of the methods; this phase depended on mortality data from SIM to be conducted. After obtaining information on the educational levels of the dead, rates of mortality by age, sex and education were estimated using coefficients obtained by Poisson regression. The results show significant educational gradient in mortality, especially for men and the city of São Paulo as compared to the state. An important point was the differential behavior in mortality by educational level with age. From the literature review, one of the hypotheses was a decreasing differential with age. However, the difference was almost constant by age. To further analysis, it is necessary to understand how transition of mortality has occurred in the state and how periodic measures may influence the results. It is important to understand the relationships between period, age and cohort to analyze the differential behaviors of educational mortality. Furthermore, the increased access to education which has elevated educational level of younger cohorts in Brazil should be taken into account. It should however be pointed out in these discussions, that they were not privileged in the thesis. Regarding the use of census to study differentials in mortality, results give an optimistic perspective. Meanwhile, there is an expectation to improve the coverage and quality of information on the educational levels of SIM. Census can be an important source for the studies of socioeconomic and mortality differentials offering other features that can complement the analysis of the effect of education.

Keywords: mortality differentials; adult mortality; census

1. INTRODUÇÃO

O objetivo desta tese é investigar os diferenciais da mortalidade adulta por escolaridade no estado e no município de São Paulo, a partir dos dados do Censo Demográfico de 2010. Grande parte dos estudos sobre diferenciais socioeconômicos da mortalidade no país é voltada para outras dimensões investiga diferenciais regionais (e.g Wood & Carvalho, 1988; Duarte *et al* 2002; Simões, 1999; Simões, 2002). Estudos que considerem a educação têm focaco no papel da escolaridade materna sobre a mortalidade infantil ou na infância (e.g Barros & Sawyer, 1993). A partir dos anos 2000, surgiram investigações no Brasil sobre o papel da educação do indivíduo que usam técnicas indiretas de estimação da mortalidade ou imputação de dados de escolaridade para informações faltantes (Perez & Turra, 2008; Perez, 2010; Guedes *et al*, 2011).

O retorno da questão de mortalidade para o Censo de 2010 criou novas possibilidades de investigação do tema. Trata-se de uma alternativa ao Sistema de Informações em Mortalidade (SIM), que é a principal fonte para estudos de mortalidade, mas que possui um percentual expressivo de perda de informação socioeconômica. De forma pioneira, estudos recentes investigaram o papel das características socioeconômicas dos domicílios na mortalidade jovem (Pereira, 2014; Pereira, 2016) e o diferencial de mortalidade por escolaridade do responsável pelo domicílio (Silva, 2014; Silva *et al* 2016), a partir de informações do Censo.

Mas é possível avançar em relação às estimativas existentes. Investigações sobre o efeito das características individuais na mortalidade são dificultadas pela falta deste tipo de informação no Censo. Os quesitos relacionados ao óbito, que é aplicado a todos os domicílios particulares permanentes do universo, só abrangem o sexo e a idade do falecido. Nesse contexto, outra contribuição pretendida neste trabalho é usar metodologias adequadas para imputar os dados de escolaridade dos indivíduos que faleceram durante o período de referência do Censo.

Logo, os principais objetivos da tese são:

- desenvolver um método de imputação da escolaridade dos indivíduos que sofreram óbito, usando técnicas adequadas destacadas pela literatura sobre dados faltantes;
- estimar e analisar o gradiente educacional da mortalidade para São Paulo - estado e capital;
- investigar como o gradiente em mortalidade adulta se comporta em função da idade e do sexo.

A amostra do estado de São Paulo foi selecionada para desenvolver e validar a metodologia de imputação. Espera-se que futuramente os métodos possam ser aplicados para estudos de diferenciais educacionais de mortalidade no Brasil.

O crescimento do interesse pelos diferenciais socioeconômicos da mortalidade é datado do início do século XIX. Até meados do século XX houve uma predominância do uso de classes ocupacionais na investigação destes diferenciais (Antonovsky, 1967; Elo, 2009). A partir dos anos 1960, a escolaridade passou a ser uma dimensão socioeconômica privilegiada neste tipo de estudo. Ela é considerada a variável com menor grau de causalidade reversa com mortalidade; a menos afetada pela proximidade com a morte; e é determinada no início da fase adulta, se mantendo relativamente estável ao longo da vida (Kitagawa & House, 1963; Preston & Taubman, 1994; Elo & Preston, 1996; Montez *et al*, 2012). Em relação à qualidade da informação, a escolaridade tende a ser reportada com certa precisão em pesquisas de inquérito (Liberatos *et al*, 1988; Montez *et al*, 2012).

Tanto em países desenvolvidos e quanto naqueles em desenvolvimento, observa-se um gradiente educacional da mortalidade expressivo: quanto maior a escolaridade menor a mortalidade. O efeito da escolaridade tem sido mais expressivo que o da renda (Kitagawa & Houser, 1973; Rogers *et al*, 1999b). Ele interage com idade e sexo: os diferenciais são maiores para os homens e decrescem com a idade em vários países desenvolvidos (Elo & Preston, 1996; Ross *et al*, 2012; Valkonen, 1989; Kunst e Mackenbach, 1994). O gradiente também está presente em países em desenvolvimento da Europa (Mackenbach *et al*, 2008; Shkolnikov *et al* 1998; Denisova, 2009; Leinsalu, 2003), da Ásia (Liang *et al*, 2000; Subramanian *et al*, 2006; Hurt *et al*, 2004), da África (Berhane *et al*, 2002; Walque & Filmer, 2013) e da América Latina e Caribe (Koch *et al*, 2007; Manzelli, 2014; Sandoval e Turra, 2015).

Nos Estados Unidos, os diferenciais têm se expandido desde a década de 1960 (Rogot *et al*, 1992; Pappas *et al*, 1993). Na Europa, os países de economia em transição apresentam os maiores diferenciais de mortalidade por educação (Mackenbach *et al*, 2008) e experimentaram um aumento da desigualdade a partir da década de 1990, com o fim do regime socialista (Shkolnikov *et al* 1998; Leinsalu, 2003).

Dentre as principais abordagens explicativas da associação entre mortalidade e educação, destacam-se os modelos econômicos, inspirados no modelo de produção de saúde de Grossman (1972), e a abordagem psicossocial, privilegiada pela demografia social (House *et al*, 1988; Williams, 1990). Nos modelos de inspiração econômica, a educação é *proxy* de capital humano e, por isso, guarda uma relação com a renda: pessoas com melhores níveis de educação exibem salários médios maiores. Portanto, suas funções orçamentárias são diferenciadas daquelas que representam os menos escolarizados. A dotação de recursos permite aos mais escolarizados consumir maiores quantidades de bens e serviços para produção de saúde como, por exemplo, alimentos saudáveis, maiores espaços para viver e áreas menos poluídas para residirem. Além da sua relação com a renda, a educação melhora o discernimento dos indivíduos e muda a forma como os mesmos processam a informação de saúde (Preston & Taubman, 1994; Sickles & Taubman, 1997; Grossman, 2008).

Na abordagem da sociodemografia, os fatores psicossociais e os serviços de saúde mediam a relação entre fatores socioeconômicos e a saúde. Existem fatores psicossociais e comportamentais como estresse, prática religiosa, estado civil, hábitos de alimentação, prática de exercício físico, tabagismo e suporte social, que influenciam o estado de saúde e a probabilidade de morte. Tais fatores, por sua vez, variam em função das características socioeconômicas dos indivíduos, sobretudo em função da escolaridade (Williams, 1990).

Com base nas evidências empíricas encontradas para outros países e para o Brasil e nas abordagens teóricas sobre o tema, espera-se que os resultados da tese apontem uma mortalidade adulta inversamente relacionada à escolaridade; e que exista um comportamento diferenciado por idade e sexo, sendo maior entre os homens. O gradiente da escolaridade na mortalidade seria sensível à região de residência, tendo em vista a heterogeneidade econômica, cultural e comportamental e de oferta de serviços de saúde observadas no Brasil (Wood e Carvalho, 1988).

A tese é composta por 7 capítulos, sendo o primeiro a introdução. O capítulo 2 apresenta as principais evidências sobre os diferenciais de mortalidade em países desenvolvidos e em desenvolvimento e os arcabouços teóricos mais relevantes na demografia. O capítulo 3 tem por objetivo apresentar o arcabouço teórico e empírico sobre dados faltantes (*missing*), privilegiando as principais técnicas de imputação e de estimação para lidar com o problema. No capítulo 4 são discutidos os aspectos metodológicos dos dados do Censo Demográfico de 2010, os métodos de imputação e de estimação da mortalidade. O capítulo 5 apresenta os resultados da imputação da escolaridade para os indivíduos que experimentaram óbito, assim como a discussão e análise de consistência dessas estimativas, segundo os diferentes métodos adotados. No capítulo 6, são apresentados os resultados das estimativas de mortalidade por escolaridade, segundo idade e sexo são. O capítulo 7 trás as conclusões e a discussão das principais contribuições e limitações do trabalho.

2. ASPECTOS TEÓRICOS E EMPÍRICOS DO GRADIENTE EDUCACIONAL DA MORTALIDADE

2.1. Contexto

Embora o termo socioeconômico seja abrangente, para estudos de mortalidade alguns autores reconhecem a renda, a educação e a ocupação como indicadores principais (Preston & Taubman, 1994; Adler, 1994). Outros acrescentam a posição dos indivíduos na comunidade, o poder e a riqueza como importantes dimensões do *status* socioeconômico (Krieger *et al*, 1997).

O interesse pelo gradiente socioeconômico da mortalidade cresceu durante a primeira metade do século XIX nos países ocidentais, quando uma ampliação dos diferenciais de longevidade segundo classes sociais foi percebida. Até meados do século XX, o uso de classes ocupacionais para associar mortalidade e *status* socioeconômicos era predominante (Antonovsky, 1967; Liberatos *et al*, 1988; Elo, 2009).

Durante o século XVI, a população mundial experimentou um baixo crescimento, prevalecendo altas taxas de fecundidade e mortalidade. Entre o século XVII e meados do século XIX, houve uma acentuada elevação da esperança de vida dos estratos ocupacionais médios e altos das populações ocidentais. A população mundial experimentou um rápido crescimento, grande parte concentrado no Ocidente. Entretanto, os ganhos de vida para a população dos estratos ocupacionais mais baixos foram bem mais suaves. Em alguns casos foram observadas perdas na expectativa de vida entre classes proletárias, que haviam surgido com a revolução industrial (Antonovsky, 1967).

A partir dos anos 1850, houve queda do diferencial de mortalidade por ocupação nos países industrializados, período em que os ganhos médios de sobrevivência foram maiores entre as classes mais pobres. A população mundial cresceu rapidamente até 1930, sobretudo a ocidental. A partir desse panorama, muitos cientistas acreditavam que o declínio da mortalidade nos países industrializados seria acompanhado pelo estreitamento dos diferenciais entre classes (Antonovsky, 1967).

Entretanto, o gradiente permaneceu considerável ao longo do século XX, o que manteve o interesse crescente nos diferenciais socioeconômicos da mortalidade na agenda de pesquisa de países desenvolvidos. Em estudo clássico para as populações da Inglaterra e Gales, usando dados do *Occupational Mortality Reports* de 1920 a 1963, Pamuk (1985) trabalhou com cinco categorias de trabalhadores, que vão de "não qualificados" a "ocupações gerenciais", e encontrou associação expressiva entre mortalidade e classes ocupacionais. Os diferenciais de mortalidade adulta foram decrescentes no início dos anos 1920, mas voltaram a crescer nas décadas de 1950 e 1960. As evidências não corroboravam a ideia de convergência da mortalidade, ao contrário apontavam um crescimento dos diferenciais de mortalidade por classes.

Nos Estados Unidos, o *National for Health Statistics Center* - NHSC discutia as dificuldades em usar classes ocupacionais para estudo de mortalidade, devido à incompatibilidade entre os registros de morte e os censos. Além disso, a análise ficava limitada aos adultos, devido à dificuldade de incluir crianças e idosos em estudos que considerassem ocupação com indicador de posição social (NCHS, 1961; 1968; 1969). Por este motivo, vários estudos passaram a adotar índices socioeconômicos que combinavam indicadores de renda, educação e ocupação.

Kitagawa e Houser (1973), em estudo clássico e pioneiro para a década de 1960 nos Estados Unidos, pontuavam que a renda, a educação e a ocupação, que estão associados entre si, possuem relações muito distintas com a mortalidade. A renda e a ocupação possuem relação de causação reversa mais forte com mortalidade. A proximidade com a morte faria com essas variáveis se modificassem de forma considerável, distorcendo seus efeitos sobre a mortalidade. A educação seria a dimensão socioeconômica menos afetada pela proximidade com a morte, pelo menos para a população adulta. Sendo determinada no início da fase adulta, sua tendência seria se manter estável ao longo do ciclo de vida. Por este motivo, eles privilegiaram a educação como uma dimensão socioeconômica para estudos da mortalidade.

Nas décadas subsequentes, vários pesquisadores de países desenvolvidos passaram a adotar a educação como um bom indicador para estudos de gradiente socioeconômico da mortalidade. Além das vantagens destacadas, ela é importante na determinação da ocupação, da renda e da riqueza (Preston & Taubman, 1994; Elo & Preston, 1996; Montez

et al, 2012) e costuma ser mais relevante na determinação da saúde para indivíduos aposentados, desempregados ou fora da força de trabalho. Com relação à qualidade da informação, a educação costuma ser reportada com maior precisão em pesquisas de inquérito, se comparadas à renda e à riqueza (Liberatos *et al*, 1988; Montez *et al*, 2012). Em países desenvolvidos, sobretudo Estados Unidos e Reino Unido, o monitoramento das desigualdades socioeconômicas da mortalidade tem sido sistemático e a educação tem sido um indicador socioeconômico central (Cutler & Lleras-Muney, 2007).

2.2. Evidências do gradiente educacional da mortalidade em países desenvolvidos

Grande parte dos estudos sobre diferenciais de mortalidade por educação trabalha com a população adulta, iniciando com indivíduos de 25 anos ou mais, considerando esta uma idade suficientemente elevada para que o grau de escolaridade dos indivíduos já esteja definido. Nos parágrafos seguintes, serão destacados alguns dos principais trabalhos sobre o tema e seus principais resultados para este grupo de países.

Nos Estados Unidos, o estudo dos diferenciais socioeconômicos da mortalidade na década de 1960 de Kitagawa e Houser (1973) tem sido um importante marco na agenda de pesquisa do tema. Eles estabeleceram a educação como principal indicador para pesquisa de gradiente socioeconômico da mortalidade. Por meio de dados dos registros de morte e do censo de 1960, eles encontraram diferenciais de escolaridade maiores para população em idade ativa, comparada aos idosos, e para as mulheres. Os diferenciais variaram inversamente à idade, sendo praticamente extintos a partir dos 65 anos para homens e mulheres não brancas. Para as mulheres brancas deste grupo etário, os diferenciais ainda foram expressivos, embora menores que para idades mais jovens. O efeito encontrado para educação foi maior que o observado para renda, no caso das mulheres. No caso dos homens, o efeito da renda é maior, mas provavelmente as limitações da causalção reversa confundem o seu papel na mortalidade (Kitagawa & Houser, 1973).

Para a primeira metade da década de 1980, Elo e Preston (1996) apresentam estudo sobre diferenciais de escolaridade na mortalidade para a população economicamente ativa (25-64 anos) e população em idade de aposentadoria (65 anos ou mais). Para ambos os sexos, foram encontrados diferenciais de mortalidade por grupos de escolaridade. Os contrastes entre os grupos de menor e maior escolaridade (0-8 anos e 16 ou mais anos de estudo)

foram maiores para o grupo etário mais jovem 25-64 anos e para os homens. Em uma análise multivariada em que foram incorporados controles por renda, residência e características individuais, os efeitos da educação caem substancialmente, sendo o impacto maior também para a população masculina em idade ativa. Dentre as variáveis consideradas, a renda apresentou maior associação com a escolaridade, tendo sido a que mais afetou seu efeito sobre a mortalidade (Elo & Preston, 1996). Estudos com dados da década de 1990 também apontam que a renda diminui substancialmente o efeito da educação. Entretanto, o nível de educação permanece relevante, mesmo quando controlado por renda, raça, *status* marital e ocupação (Rogers *et al*, 1999b).

Estudo recente usou dados do *National Health Survey-Linked Mortality Files (NHIS-LMF)* de 1986 a 2000, para a população de 25 anos ou mais, e corroborou os diferenciais por escolaridade. O efeito é mais expressivo para os homens, mas os diferenciais por sexo decrescem com educação. Os ganhos maiores ocorrem para homens com educação superior (*college degree*), que apresentam uma mortalidade vertiginosamente menor que aqueles com educação inferior (Ross *et al*, 2012).

Com relação ao comportamento do gradiente no tempo, para os EUA houve um aumento dos diferenciais, assim como ressaltou Pamuk (1985). Rogot e outros autores (1992) identificaram diferenciais de mortalidade por escolaridade entre os homens maiores àqueles encontrados na década de 1960, por Kitagawa e Houser. Segundo Pappas e outros pesquisadores, embora a mortalidade tenha declinado para a população total entre 1960 e 1986, os ganhos não foram igualmente distribuídos entre grupos populacionais. Tomando a população de 24 a 64 anos, os mais escolarizados apresentaram maior queda da mortalidade comparado aos menos escolarizados. Uma análise por raça indica que tanto a população negra quanto a branca apresentavam maiores diferenças de mortalidade por grupos de escolaridade em 1986, comparativamente a 1960 (Pappas *et al*, 1993).

A despeito das diferenças entre os dados e as metodologias utilizadas, a confiabilidade dos resultados destes estudos tem sido relativamente elevada. Quando as metodologias são compatibilizadas ou substituídas para fins comparativos, resultados de diversos estudos se validam. De fato parece ter havido um aumento destes diferenciais ao longo das décadas de 1970 e 1980, pelo menos para a população masculina branca. Para as mulheres, o

diferencial diminuiu entre as idades 25-64 anos e se manteve estável para aquelas entre 65 e 74 anos (Preston e Elo, 1995).

Por outro lado, Duleep (1989) estimou mortalidade entre 1973 e 1978, usando dados do *Social Security Administration* (SSA), e encontrou resultados contrastantes: uma aparente queda dos diferenciais por escolaridade dos homens de 25 a 64 anos, comparativamente ao estudo de Kitagawa e House. Porém, o próprio autor chamou atenção para o fato de que os dados da SSA constituem uma subamostra da população do estudo antecessor, e que possivelmente há viés em suas estimativas. A SSA pode excluir homens de baixo nível socioeconômico, com mortalidade mais baixa, e homens de elevado nível socioeconômico, com mortalidade acima da média do grupo, aproximando a mortalidade entre os grupos.

No século XXI, a discussão continua presente e relevante na agenda de pesquisa dos EUA. Até final dos anos 2000, o crescimento do diferencial de mortalidade era atribuído a uma queda mais acelerada da mortalidade entre os mais escolarizados em relação à aos menos escolarizados (Pappas *et al*, 1993; Preston & Elo, 1995). Mas estudos recentes tem apontado um crescimento da mortalidade para grupos populacionais menos escolarizados.

Há indícios de crescimento do diferencial de mortalidade por educação entre mulheres brancas de 45-84 anos, associado a um crescimento da mortalidade entre mulheres com educação secundária ou inferior, entre 1986 e 2006 (Montez et al, 2011). Resultados de Hayward et al (2015) corroboram esses achados. Além deles, Sasson (2016) indica que, entre 1990 e 2010, a esperança de vida aos 25 anos para as mulheres menos escolarizadas diminuiu cerca de 3 anos, em grande parte devido a um aumento da mortalidade entre as idades de 45 e 64 anos. Em contrapartida, houve melhorias consideráveis na esperança de vida dos mais escolarizados.

No mesmo sentido, Case e Deaton (2015) encontraram um crescimento da mortalidade de 34 por 100 mil, entre adultos brancos não hispânicos de 45 a 54 anos, no período de 1999 a 2013. A maior parte desta mudança está relacionada ao aumento das taxas de morte por causas externas, sobretudo intoxicação por drogas, álcool e suicídio, que foram muito maiores, comparativamente aos não hispânicos negros. Tais mortes foram concentradas entre indivíduos com educação secundária ou inferior. Para aqueles com bacharelado, a queda das taxas de morte pelas referidas causas se confirmam (Case & Deaton, 2015). Como não foram realizadas padronizações por idade, cabe discussão sobre estes resultados.

De qualquer forma, os achados recentes não somente contradizem a tendência de queda da mortalidade observada até então, como também apontam a educação como um fator importante para explicar o seu comportamento segundo algumas causas específicas. Olshansky e outros pesquisadores haviam chamado atenção para uma possível queda da esperança de vida nos EUA, em grande parte em função do efeito negativo do crescimento da obesidade. A tendência de crescimento do problema poderia levar à reversão dos ganhos de longevidade experimentados pela população no século XX (Olshansk *et al.*, 2005).

Na Europa, Kitagawa e Houser (1973) também influenciaram estudos na área, aumentando o interesse pelos diferenciais de escolaridade, sobretudo para fins comparativos. Até a década de 1960, as investigações sobre os determinantes socioeconômicos da mortalidade adulta priorizavam as classes ocupacionais (Antonovsky, 1967; Pamuk, 1985; Townsend & Davidson, 1988; Elo 2009). Mas estes indicadores dificultavam as comparações entre países devido à adoção de diferentes estruturas de classificações das ocupações adotadas por cada um (Valkonen, 1989).

Em estudo que considerou a população de 35 a 54 anos, Valkonen (1989) comparou as desigualdades da mortalidade por escolaridade entre seis países europeus: Dinamarca, Noruega, Suécia, Finlândia, Inglaterra e Gales, Hungria. As taxas de mortalidade mostraram-se negativamente associadas à escolaridade, semelhante ao observado nos EUA. Os dados entre países variaram, mas no geral o período de abrangência foi de 1970 a 1981. A Hungria apresentou as maiores taxas de mortalidade para todos os níveis de escolaridade. Para a população masculina, com exceção da Noruega, as taxas de mortalidade decaem monotonicamente com o nível de escolaridade. Para o grupo de menor escolaridade, as taxas de mortalidade variavam entre 40 e 60 e foram maiores em relação às taxas observadas para o grupo de maior escolaridade. Os diferenciais relativos entre os países são bem semelhantes para todos os níveis de escolaridade, tornando as curvas de mortalidade em função da escolaridade quase paralelas. Para as mulheres, o padrão dos diferenciais varia entre os países.

Kunst e Mackenback (1994) estendem a comparação para nove países: Holanda, Dinamarca, Noruega, Suécia, Finlândia, Inglaterra e Gales, França, Itália e Estados Unidos. Com base em dados longitudinais da década de 1970 e do início dos anos 1980, foram encontradas relações inversas entre mortalidade e educação. Os Estados Unidos, a França e

a Itália possuíam as maiores iniquidades. Outro estudo para as décadas de 1990 e 2000 apontou a França como um dos países da Europa Ocidental com maior iniquidade de mortalidade por educação (Mackenback, 2008).

Em estudo comparativo de onze países, foram encontrados diferenciais de mortalidade masculina por nível de educação na Áustria, sendo a taxa de mortalidade do grupo de menor educação 86% maior que o grupo mais escolarizado. Para a Dinamarca e Inglaterra esses diferenciais são, respectivamente, de 37% e 36%. Há diferenciais também entre as mulheres, porém menores expressivos que os observados para os homens (Huisman *et al* 2004). Para a Finlândia, observam-se efeitos significativos da educação e da renda familiar na mortalidade adulta prematura no período de 1985–1994, embora os resultados sejam diferenciados por sexo (Elo, Martikainen & Smith, 2006).

2.3. Evidências do gradiente educacional da mortalidade em países em desenvolvimento

Em países em desenvolvimento, a investigação dos diferenciais socioeconômicos da mortalidade é mais recente, mais escassa e mais restrita, devido à limitação em relação à disponibilidade de dados de qualidade e/ou adequados (Subramanian *et al*, 2006; Elo, 2009; Saikia & Bhlat, 2008; Walque & Filmer, 2013). Dentre os estudos que usam a escolaridade como um indicador socioeconômico, boa parte é direcionada para o diferencial da escolaridade na mortalidade infantil ou na infância (Caldwell, 1989). No caso dos países de economia em transição, o interesse pelo tema cresceu com o final da era socialista, sendo os estudos bastante recentes (Shkolnikov *et al*, 1998; Mackenbach *et al*, 2008; Leinsalu *et al*, 2003).

Na Ásia existem estudos para China, Índia e Bangladesh que apontam mortalidade diferenciada por nível educacional (Liang *et al*, 2000; Subramanian *et al*, 2006; Hurt *et al*, 2004). A China é um país diferenciado, com um sistema de organização social muito específico e grandes disparidades entre as sociedades urbana e rural. Após a mudança para a economia de mercado, no final da década de 1970, o padrão de vida da população apresentou melhorias. Mas ao final da década de 1990, a maior parte dos idosos do país era analfabeta. Há apontamentos de um gradiente educacional da mortalidade para a população acima de 60 anos. Um ano adicional de escolaridade se associa a 5% de redução no risco

de morte. O efeito da educação permanece mesmo quando são consideradas variáveis de urbanização e presença de bens de luxo. Porém, o efeito da educação é mais forte para área urbana e é diferenciado por sexo. Dentre as mulheres, o efeito da educação se mostra bem maior: um ano adicional de educação formal reduz o efeito do risco de morte em 22%, enquanto para os homens essa redução é de apenas 6% (Liang *et al*, 2000).

Na Índia, há uma escassez de estudos que utilizam a escolaridade devido à falta de dados. Um estudo considerou o número de analfabetos presentes no domicílio e apontou que o risco de mortalidade entre homens diminuía cerca 60% e o de mulheres 40%, quando o número de analfabetos passava de nenhum para dois (Saikia & Bhlat, 2008).

Hurt e parceiros (2004) encontraram diferenciais de escolaridade para mulheres e homens de uma região rural de Bangladesh, ao sudeste de Dhaka. Foi acompanhada uma coorte de mulheres casadas com 45 anos ou mais e seus respectivos maridos, no período de 1982 a 1998. Aproximadamente 22% delas haviam recebido educação formal e 59% educação corânica. O fator explicativo mais importante para a mortalidade feminina foi a escolaridade. As mulheres com alguma educação formal apresentaram uma mortalidade substancialmente menor, comparadas aquelas sem nenhuma escolaridade. A educação corânica também se mostrou benéfica, embora em grau menor que a formal. Dentre os homens, somente a educação formal mostrou-se significativamente associada à mortalidade. Por outro lado, os maridos cujas esposas receberam algum tipo de educação (formal ou corânica) apresentaram menores riscos de morte (Hurt *et al*, 2004).

Para o continente africano, Walque e Filmer trazem importantes achados, com base em dados da *Demographic Health Survey* (DHS). Foram considerados 46 países: 33 da África Subsaariana e 13 de outras regiões¹. A mortalidade adulta foi estimada com base nas informações de sobrevivência de irmãos das mulheres de 15 a 49 anos entrevistadas, nos períodos de 1975-79, 1980-84, 1985-89, 1990-94, 1995-99 e 2000-04. Para estimar mortalidade por situação de residência e escolaridade, foram consideradas as informações prestadas pelos respondentes, que são os irmãos vivos (Walque & Filmer, 2013).

¹ Os países de outras regiões são: Bolívia, Brasil, Peru, República Dominicana, Guatemala, Camboja, Haiti, Indonésia, Yemen, Jordânia, Marrocos, Nepal e Filipinas.

Os países do sudeste da África Subsaariana apresentaram expressivo aumento da mortalidade a partir dos anos 1980, o que é relacionado à elevada prevalência de HIV. Para os demais países Subsaarianos, a mortalidade cresceu a partir de 1980 ou início de 1990, inclusive em países como Benin, Guiné, Mali, Mauritânia e Nigéria, onde a prevalência de HIV é relativamente baixa. Os países das demais regiões apresentaram mortalidade constante ou declinante nos mesmos períodos. Quanto à escolaridade, na África Subsaariana, os homens com educação primária ou superior têm mortalidade 13% maior, comparados aqueles com maior escolaridade. Para as mulheres, esse diferencial é de 30%. Ademais, elas apresentam mortalidade menor que a dos homens, sendo os diferenciais por sexo maiores para o grupo mais escolarizado. Ao longo do tempo, o gradiente educacional aumentou entre países Subsaarianos, principalmente, entre os homens. Porém, esse aumento foi consideravelmente maior entre os países com elevada prevalência de HIV. O gradiente também está presente nos demais países, porém as taxas de mortalidade caíram ao longo do tempo, sobretudo entre os mais escolarizados, mas cresceram para mulheres menos escolarizadas maiores de 25 anos (Walque & Filmer, 2013).

Na Etiópia, dados para população acima de 15 anos, de 1987 a 1996, colocam a alfabetização do indivíduo como um fator relacionado à mortalidade, sobretudo nas áreas rurais. Pessoas alfabetizadas apresentavam menor mortalidade que aquelas não alfabetizadas. O efeito se mostra cumulativo com a idade² (Berhane *et al*, 2002).

Com relação à Europa, os países de economia de transição apresentam iniquidades de mortalidade superiores às observadas nos países da Europa Ocidental. A Hungria, a República Tcheca e a Polônia apresentam as maiores iniquidades, enquanto populações como da Suécia e os bascos (Espanha) possuem as menores. Na maior parte dos países do leste europeu e dos bálticos, as iniquidades educacionais de mortalidade são maiores que a média da Europa. Outro achado importante é que na Europa Ocidental, há uma menor iniquidade educacional da mortalidade por todas as causas, grande parte explicada pelo diferencial menor da mortalidade por doenças cardiovasculares. Já para o leste do continente, o diferencial por estas doenças é maior, contribuindo para os grandes diferenciais de mortalidade geral (Mackenbach *et al*, 2008).

² Os autores chamam atenção para o fato da amostra ser pequena para idades acima de 65 anos, o que deve ser levado em consideração para análise dos resultados (Berhane *et al*, 2002).

A Rússia também apresenta um marcado gradiente educacional da mortalidade (Shkolnikov *et al* 1998; Shkolnikov *et al* 2006). No período de 1979 a 1989, a mortalidade entre pessoas de 20 a 69 anos, possuía uma forte associação negativa com a escolaridade, porém declinante com a idade. Essa relação era mais forte entre os homens e variava segundo causas de morte, sendo expressivamente maior entre as causas violentas e acidentes. Os diferenciais se expandiram na década de 1990 (Shkolnikov *et al* 1998). Outros dados corroboram os achados: os indivíduos com melhores níveis de escolaridade se beneficiam da longevidade, sendo que aqueles com grau universitário apresentam risco de mortalidade 40 pontos percentuais menor. Esses resultados têm por base os dados do *Longitudinal Monitoring Survey* para o período 1994 a 2007 (Denisova, 2009).

A Estônia, que alcançou autonomia em 1991, foi um dos países com abertura comercial mais intensa na Europa Central. Entre as décadas de 1960 e 2000, o comportamento da mortalidade no país foi similar ao observado para os demais países em transição, com expectativa de vida masculina aumentando em média 1 ano e a feminina em torno de 4 anos. A mortalidade aumentou no início dos anos 1990, com a abertura econômica, voltando a declinar após 1994. Dados do censo apontam que houve um acentuado crescimento dos diferenciais de mortalidade por nível de educação entre 1989 e 2000. A expectativa de vida condicionada à sobrevivência aos 25 anos aumentou para os graduados em universidade e declinou para a população com menor escolaridade. Embora essa tendência tenha sido observada para todas as idades, ela foi mais expressiva entre os grupos mais jovens (Leinsalu, 2003).

Para a América Latina, embora esse tipo de investigação seja ainda rara³, alguns registros corroboram os resultados encontrados para os demais países: a mortalidade varia inversamente ao nível de educação (Koch *et al*, 2007; Manzelli, 2014; Sandoval e Turra, 2015). Na Argentina, a razão entre a mortalidade de adultos com 13 ou mais anos de estudo e menos de 8 anos era substancialmente grande em 2010, seguindo uma relação decrescente com a idade. As iniquidades eram maiores entre os homens, sobretudo a partir

³ Existem estudos sobre o gradiente socioeconômico da mortalidade, porém eles usam indicadores disponíveis (Rosero-Bixby, 1996; Rosero-Bixby & Dow, 2009; Rofman, 2004; dentre outros), sendo gradiente educacional pouco explorado (Manzelli, 2014).

dos 30 anos. Estas estimativas referem-se a pessoas com 25 anos ou mais, construídas com dados do Censo Nacional da Argentina e estatísticas dos *Mortality Files-Vital Statistic*, ambos de 2010 (Manzelli, 2014).

No Chile, Koch e outros autores utilizaram uma amostra de 920 pessoas maiores de 20 anos, retirada de um estudo longitudinal conduzido para a população urbana de San Francisco de Mostazal, entre 1997 e 1999. As coortes com nível menor escolaridade apresentaram risco de morte maior em relação às de maior escolaridade (Koch *et al*, 2007). O trabalho de Sandoval e Turra (2015) aponta taxas de mortalidade consideravelmente menores para grupos mais escolarizados no período de 2001-2003. Entre os homens com treze ou mais anos de estudo, as taxas de mortalidade eram de 2,6 por mil, enquanto os menos escolarizados (0 a 8 anos de estudo) exibiam taxas de 19,7 por mil. Entre as mulheres, estas taxas são respectivamente 1,6 e 15,8 por mil. Um resultado interessante do trabalho é que, semelhantemente ao observado para os Estados Unidos e países da Europa Central e Meridional, o Chile experimentou um aumento da mortalidade entre os indivíduos menos escolarizados entre os períodos de 1991-1993 e 2001-2003. Para os homens menos escolarizados, a taxa de mortalidade passou de 17,2 para 19,7 por mil. A população mais escolarizada, em contraposição, apresentou queda da mortalidade no período.

Na Costa Rica foram encontrados resultados conflitantes para a população idosa, tomando-se por base uma pesquisa amostral representativa da população. A mortalidade de indivíduos com 60 anos ou mais se mostrou diretamente relacionada com os níveis de educação e riqueza. A hipótese é de que o resultado esteja relacionada ao comportamento dos indivíduos de maior *status* socioeconômico que tornam sua mortalidade por doenças cardíacas maior que os de menor *status* (Rosero-Bixby & Dow, 2009).

2.4. Evidências do gradiente educacional da mortalidade no Brasil

Os primeiros estudos sistematizados de mortalidade no Brasil datam da década de 1940, quando a esperança de vida ao nascer era estimada em 50 anos (Yunes & Ronchezal, 1974; IBGE, 2010a). A partir desta década, observou-se uma queda vertiginosa da mortalidade.

Entre 1980 e 2010, a esperança de vida ao nascer e_0 passou de 62,7 para 73,5 anos (IBGE, 2010a; IBGE, 2010b).

Grande parte da variação da mortalidade é atribuída à queda da mortalidade infantil. Estima-se que, no mesmo período, a população com menos de um ano foi responsável por aproximadamente 56% dos ganhos totais em e_0 (Simões, 2002). Mas essa queda ocorreu de forma desigual entre as grandes regiões brasileiras e em diferentes grupos socioeconômicos. Neste cenário, um número expressivo de estudos foi produzido sobre as relações da mortalidade infantil (e na infância) com o nível de desenvolvimento regional e com fatores socioeconômicos (Simões, 2002; Sawyer & Soares, 1982; Sastry, 1997; Simões, 1999; Duarte *et al*, 2002). Dentre estes, uma parte identifica associação entre mortalidade infantil e escolaridade materna (Barros & Sawyer, 1993; Sastry, 1997; Garcia & Santana, 2011; Lima, 2013).

Para outros estratos da população, os aspectos socioeconômicos da mortalidade têm sido investigados usando indicadores de renda. Para a população total, o estudo clássico de Wood e Carvalho (1988) estima esperança de vida ao nascer por nível socioeconômico e aponta grandes distâncias da mortalidade entre os grupos de renda. Em 1970, os indivíduos das famílias com maior nível de renda alcançavam vantagens de até 12 anos de esperança de vida ao nascer, em relação aos de famílias de menor renda. As grandes desigualdades regionais de mortalidade estariam, dessa forma, relacionadas às desigualdades regionais de renda. Se a distribuição de renda fosse menos desigual, assim também seria a mortalidade. Semelhantemente, diferenciais de mortalidade encontrados segundo situação do domicílio eram explicados, em grande parte, pelos diferenciais de renda entre os domicílios urbanos e rurais.

Mas a escassez de dados tem sido um empecilho ao avanço dos estudos nesta área, sobretudo para diferenciais de escolaridade para a população adulta. Embora tenha havido melhoria no grau de cobertura dos óbitos e das causas de morte do "Sistema de Informações sobre Mortalidade" - SIM (Brasil, 2004; Lima & Queiroz, 2014a), o preenchimento de variáveis socioeconômicas apresenta um elevado grau de dados faltantes (Guedes *et al*, 2011). O problema pode ser pior para a população adulta, dada a possibilidade da informação da certidão ser prestada por alguém que não conhece bem o indivíduo falecido (Perez, 2010). Essa tem sido uma limitação para uso de dados das

certidões de óbitos para investigação de gradiente educacional da mortalidade (Guedes *et al*, 2011; Perez 2010).

A despeito dessa limitação de informação, há esforços recentes para produção de estimativas de mortalidade infantil e adulta por escolaridade, com base em técnicas indiretas (Perez e Turra, 2008; Perez, 2010; Guedes *et al*, 2011). Perez (2010) desenvolveu uma metodologia que permite estimar taxas de mortalidade por nível de escolaridade, utilizando dados de retrospectivos de pesquisas domiciliares e um método estocástico. Trata-se de método baseado na técnica de estimação indireta de mortalidade adulta por meio de dados de orfandade de Brass⁴. Utilizando dados da Pesquisa de Padrões de Vida (PPV) e da Pesquisa por Amostra de Domicílios (PNAD) de 1996, foram estimadas medidas de mortalidade para a população feminina de 20 a 69 anos, com base informações de morte e escolaridade das mães dos entrevistados. As taxas de mortalidade foram decrescentes em relação à escolaridade. As coortes mais jovens apresentaram mortalidade menor, em parte devido ao efeito da maior escolarização em relação às coortes mais velhas. Foram encontrados diferenciais expressivos entre os grupos de escolaridade, embora os resultados tenham variado em função da amostra utilizada. As mulheres sem escolaridade apresentaram mortalidade até 70% maior do que aquelas com escolaridade entre 1 e 8 anos.

Guedes *et al* (2011) também sugerem o uso de dados da PNAD para aplicação do método de Brass para estimar mortalidade na infância⁵ e o método de orfandade para a mortalidade adulta. No caso da mortalidade na infância, eles obtêm as estimativas de mortalidade segundo a escolaridade da mãe, informação disponível na PNAD. Para a mortalidade adulta, é utilizada a escolaridade dos respondentes. Após estimada a TAB de vida da população feminina, os autores sugerem estimar mortalidade masculina com base em relações conhecidas sobre os diferenciais de sexo da mortalidade no Brasil, adotando TABs de vida modelo.

⁴ Descrição da técnica pode ser vista em Brass W., Hill K. Estimating adult mortality from orphanhood. Proceedings of International Population Conference. International Union for the Scientific Study of Population, Liege, Vol. 3: 111-23. 1973.

⁵ Ver Brass W. Methods for Estimating Fertility and Mortality From Limited and Defective Data. Chapel Hill, N.C., University of North Carolina, 1975.

Os trabalhos de Gomes et al (2013; 2015) usam dados da SABE: Saúde, bem-estar e envelhecimento para população acima de 60 anos e encontram associação negativa entre escolaridade e mortalidade no município de São Paulo. Há diferenciação por sexo: o efeito da escolaridade é maior para os homens.

Com o retorno do quesito sobre mortalidade⁶ no domicílio do Censo Demográfico de 2010, surgiram novas perspectivas de estimação de mortalidade por características socioeconômicas dos domicílios e indivíduos. Mas o grande desafio tem sido o fato que, para as pessoas falecidas, as únicas informações individuais prestadas pelos respondentes é a idade e o sexo. Por outro lado, há evidências de boa cobertura dos óbitos. Problemas maiores na qualidade da cobertura são encontrados para as idades mais avançadas (Queiroz & Sawyer, 2012).

Os trabalhos recentes de Pereira (2014), Pereira & Queiroz (2016) e Silva (2014), Silva et al (2016) são exemplos dos que fizeram uso do Censo 2010 para estimar efeitos de fatores socioeconômicos na mortalidade de jovens e adultos. Pereira e Queiroz (2016) analisaram o risco de morte de jovens entre 15 e 29 anos, segundo o nível socioeconômico dos domicílios e condições sociais dos municípios e unidades da federação onde eles estão inseridos, a partir de modelo hierárquico. Além das informações da amostra do Censo, foram utilizados dados sobre desenvolvimento humano originados da PNUD (2013) e outros indicadores específicos. Foi constatado que as características dos domicílios são responsáveis por 89,2% da variabilidade no risco de morte dos jovens. Os domicílios de menor status socioeconômico apresentaram uma probabilidade de apresentar óbito 7,7 vezes maior em relação aos de maior status. Também foram encontrados diferenciais por raça/cor e sexo dos responsáveis pelos domicílios, sendo os responsáveis negros e do sexo feminino mais prováveis de reportar mortalidade jovem no domicílio. Entretanto, o foco do estudo não foi investigar o gradiente educacional da mortalidade.

O trabalho de Silva et al (2016) investigou diferenciais de mortalidade por escolaridade. Dada a inexistência da informação de escolaridade do falecido, foi utilizada a do responsável pelo domicílio, sob o pressuposto de que o regime de mortalidade não difere

⁶ O quesito sobre óbito no domicílio já esteve presente no questionário da amostra do Censo Demográfico de 1980.

significativamente entre indivíduos de um mesmo domicílio. Foram analisadas as populações de 15 a 59 anos e de 60 a 80 anos de idade. As probabilidades de morte para a população entre 15 e 60 anos tiveram comportamento inverso à escolaridade. A população inserida em domicílios com responsáveis com ensino superior apresentou mortalidade menos elevada em relação aos "sem instrução ou ensino fundamental incompleto". Os diferenciais guardam relação também com o sexo. Os homens sem instrução ou com ensino fundamental incompleto apresentam uma probabilidade de morte 2,48 maior que aqueles com ensino superior, sendo a maior razão observada no Nordeste (2,64) e a menor do Norte (1,5). A expectativa de vida aos 15 anos de homens com o ensino superior mostrou-se 6,27 anos maior que aqueles com ensino Fundamental completo. Para as mulheres, a razão entre as probabilidades de morte do nível menor e maior de escolaridade alcança 2,58, embora o nível de mortalidade seja menor entre elas. Para as idades maiores de 70 anos, o comportamento é invertido: associação positiva entre mortalidade e escolaridade, exceto para região Sudeste.

2.5. Mecanismos de associação entre escolaridade, saúde e mortalidade adulta

A educação tem mostrado significativa associação negativa com mortalidade, mesmo após controlar os efeitos de renda, ocupação e outros fatores, conforme estudos destacados nas seções 2.3 e 2.4. Entretanto, sua aplicação para estudos de diferenças de escolaridade tornou-se frequente a partir da década de 1960, conforme já discutido. Até meados do século XX, os estudos sobre diferenças socioeconômicas da mortalidade privilegiavam a posição dos indivíduos na sociedade representada, em grande medida, por sua classe ocupacional. As abordagens teóricas para explicações destes diferenciais se desenvolveram neste contexto.

Towsend e Davidson (1988) classificaram os principais arcabouços teóricos sobre o diferencial de saúde por classe social em cinco grupos: explicação do artefato (*artefact explanation*), teorias da seleção natural ou social (*drift hypothesis*), explicações estruturalistas ou materialistas e abordagem comportamental e cultural. Esses arcabouços eram aplicados a estudos que adotavam, em sua maior parte, classes ocupacionais como medida socioeconômica.

A abordagem do artefato, exposta por Illsley (1986) e Bloor *et al* (1987), sugere que saúde e classes sociais são variáveis artificiais usadas para medir o fenômeno social e que superestimam a importância dos diferenciais de mortalidade. Assim, a incapacidade de redução das iniquidades na mortalidade entre as classes ocupacionais na Grã-Bretanha⁷, por exemplo, estaria no fato de que os ganhos em saúde eram, em grande medida, compensados pelo encolhimento do peso das classes mais pobres. Os indivíduos que se movem para classes mais ricas tem saúde melhor ou melhorada após mobilidade, comparativamente aos que permanecem nas classes mais pobres, o que explicaria o diferencial artificial. Mas essa hipótese é limitada para explicar o panorama observado no período pós II Guerra, pois mesmo com uma lenta mobilidade social na Inglaterra, os diferenciais persistiam (DHSS, 1980; Townsend & Davidson, 1988).

A abordagem da seleção natural e social - *drift hypothesis* - assume que a variável socioeconômica é dependente. Os indivíduos mais saudáveis são selecionados para as melhores posições socioeconômicas. O resultado é que os mais saudáveis possuem maiores chances de mobilidade ascendente e desfrutam de melhores posições na sociedade, o que explicaria a menor mortalidade entre eles (Blane *et al*, 1993; West, 1991). A forma e a magnitude da seleção dependem das políticas de educação e de construção das estruturas ocupacionais. Isto indica que o efeito da seleção dos indivíduos segundo suas características biológicas e de saúde deve variar entre sociedades, ou mesmo dentro de uma mesma população ao longo do tempo (West, 1991).

Recentemente, essa abordagem tem sido usada nos estudos que privilegiam a escolaridade como *proxy* de condição socioeconômica. A saúde aparece como a variável independente que tem efeito sobre os níveis de escolaridade dos indivíduos. Um efeito da saúde sobre a educação seria plausível na infância, de forma que crianças com saúde frágil teriam probabilidades maiores de experimentar baixa escolaridade e, conseqüentemente, seriam adultos menos saudáveis. A maior limitação desta abordagem é que ela não é capaz de lidar com casos como dos Estados Unidos, por exemplo, onde um número reduzido de crianças é incapaz de frequentar escola por causa de condições de saúde, no entanto o gradiente educacional permanece em expansão (Cutler & Lleras-Muney, 2007).

⁷ O debate destes autores gira em torno dos resultados apontados DHSS (1980) e revisados posteriormente por Townsend e Davidson (1988).

As teorias estruturalistas/materialistas enfatizam o papel dos fatores econômicos e socioestruturais na distribuição da saúde e do bem estar. As condições sob as quais os indivíduos vivem e trabalham os pressionam para o consumo de bens não saudáveis. Neste contexto, indivíduos de classes sociais mais baixas estariam mais expostos a ambientes insalubres. Eles possuem trabalhos mais perigosos, vivem em domicílios em piores condições e possuem recursos restritos para investimento e consumo de cuidados com saúde (Townsend & Davidson 1988). Recentemente, neomaterialistas pontuam que as iniquidades em saúde resultam do processo de acumulação de experiências e exposição a riscos que variam em função dos recursos materiais (Lynch, 2000; Lynch *et al*, 2000).

Na abordagem comportamental e cultural, as desigualdades de saúde são produto de fatores comportamentais e culturais assumidos pelos indivíduos em uma sociedade moderna e industrial. O comportamento em relação à saúde, estilo de vida é de responsabilidade dos indivíduos, que tomam atitudes prejudiciais ou benéficas a eles e a seus filhos, dependendo das decisões em relação ao consumo de produtos prejudiciais ou saudáveis, a prática exercícios físicos, ou ao cuidado preventivo com a saúde. Neste cenário, a educação é um aspecto individual que torna o indivíduo inato ou apto para adotar comportamento que o torne saudável (Fuchs, 1974).

Singh-Manoux e Marmot argumentam que a seleção natural, o materialismo e as abordagens comportamentais se complementam. Mas, os processos sociais são importantes para explicar os padrões comportamentais adotados por cada sociedade e, conseqüentemente, explicam os resultados observados na saúde. Desta forma, são os comportamentos dos grupos, e não dos indivíduos, que moldam as trajetórias em relação à saúde (Singh-Manoux & Marmot, 2005).

Nos estudos demográficos, embora Crimmins (1993) aponte uma ausência de uma teoria unificada sobre a mortalidade, quando se trata de diferenciais socioeconômicos, as abordagens explicativas mais importantes podem ser divididas em dois grandes grupos (Preston e Taubman, 1994): os modelos econômicos e a abordagem psicossocial.

Nas seções 2.5.1 e 2.5.2, esses modelos explicativos são abordados e elucidam a relação entre mortalidade e fatores socioeconômicos, com destaque para o papel da escolaridade na mortalidade entre adultos.

2.5.1. Os modelos econômicos

Segundo Preston & Taubman (1994), os modelos econômicos são os mais sistematizados para relacionar saúde e condições socioeconômicas. Eles são inspirados no modelo de produção de saúde de Grossman (1972), que por sua vez, se baseou na teoria do capital humano de Becker (1965).

Dentro desta abordagem, os indivíduos tomam decisões que ajudam a formar seu estoque de saúde H em qualquer período. A morte ocorre quando H atinge um valor abaixo do nível crítico. Os indivíduos decidem sobre investimento em saúde e consumo de determinados bens e serviços de saúde que afetam sua idade à morte. Há duas funções envolvidas neste processo: a função orçamentária e a função de produção de saúde.

Na função orçamentária, a renda deve se igualar à soma dos produtos entre preço e quantidade de bens consumidos. A renda é dada pelo produto entre salário-hora e horas trabalhadas, somada aos retornos da poupança e das transferências recebidas pelo indivíduo. A função de produção de saúde descreve como H varia em função de alguns insumos, dentre os quais se destacam: quantidade consumida de bens de saúde, quantidade de conhecimento ou tecnologia médica disponível para a população, disponibilidade de serviços médicos, dotação genética e características do ambiente em que os indivíduos se encontram inseridos. O consumo de bens de saúde é determinado pelos indivíduos, enquanto os demais insumos são determinados exogenamente (Rosenzweig & Schultz, 1983; Preston & Taubman, 1994).

O papel da educação é servir de *proxy* para capital humano. Ela guarda uma relação com a renda, visto que pessoas com melhores níveis de educação exibem, em média, melhores salários. Portanto, suas funções orçamentárias são diferenciadas daquelas que representam os menos escolarizados. A dotação de recursos permite aos mais escolarizados consumir maiores quantidades de bens e serviços para produção de saúde como, por exemplo, alimentos saudáveis, maiores espaços para viver, áreas menos poluídas para residirem. Além disso, a educação melhora o discernimento dos indivíduos e muda a forma como processam informação de saúde (Preston & Taubman, 1994; Sickles & Taubman, 1997; Grossman, 2008).

Os preços dos produtos e serviços de saúde são centrais, pois são fonte de diferenciais de mortalidade por meio de dois mecanismos. De um lado, embora os indivíduos se deparem com preços iguais no mercado, há fatores que funcionam como subsídio como, por exemplo, o acesso aos bens e serviços de saúde por meio de programas públicos ou seguros de saúde. O segundo mecanismo é o custo de oportunidade em se obter o cuidado médico, que depende tanto dos salários dos indivíduos quanto do tempo de espera pelo cuidado. Mesmo com salários baixos, o custo de oportunidade em obter o serviço médico pode se elevar, se o tempo de espera para isso for grande. Sendo assim, o papel das políticas públicas de acesso aos serviços de saúde estaria nas mudanças que elas causam no preço relativo dos bens e serviços de saúde (Preston & Taubman, 1994).

Dentre as principais limitações dos modelos econômicos, destacam-se a dificuldade na incorporação dos fatores ambientais e da formação dos gostos e hábitos individuais. Independente das características dos indivíduos, há fatores ambientais (incidência de doenças transmissíveis, por exemplo) que são impostos e alteram o estoque de saúde da população. Os modelos econômicos tratam esses efeitos como endógenos, sendo determinados pelas escolhas individuais (Preston & Taubman, 1994). Com relação à formação dos gostos, ela é vista como dada nos modelos econômicos: trata-se de um aspecto autônomo ao indivíduo e, portanto, distribuída aleatoriamente na sociedade. Não há nenhuma consideração sobre a influência do contexto social em que os indivíduos se inserem sobre suas escolhas (Preston & Taubman, 1994; Williams, 1990).

Em contraposição, as abordagens psicossociais privilegiam os fatores negligenciados pelos modelos econômicos, conforme será apresentado na seção 2.5.2.

2.5.2. A abordagem psicossocial

Embora a literatura que investiga o papel das relações sociais na saúde date do século XIX, o tema foi revitalizado na década de 1970 com o surgimento da teoria do suporte social. O interesse pelo tema foi impulsionado pelos resultados do relatório de um comitê consultivo do Serviço de Saúde Pública (*Public Health Services*) dos Estados Unidos (US, 1964), que relacionou morbidade e mortalidade por várias causas ao consumo de cigarro (House *et al*, 1988). O relatório apresenta o comportamento, o estilo de vida e as atitudes individuais como independentes da estrutura social. As teorias psicossociais, ao contrário, partem da

premissa de que as estruturas sociais moldam os valores e influenciam o comportamento dos indivíduos. Há um padrão de resposta dos grupos sociais às realidades e as restrições que lhes são impostas (Williams, 1990).

Mas foi a partir do início dos anos 1990 que autores adeptos da abordagem psicossocial passaram a incorporar o papel das características socioeconômicas em sua estrutura analítica (Preston & Taubman, 1994; House *et al*, 1990; Williams, 1990). Os fatores psicossociais e os serviços de saúde funcionam como mediadores da relação entre fatores socioeconômicos e saúde. Para entender essa interação, é preciso identificar a ligação entre os fatores psicossociais e saúde; e a ligação entre as características socioeconômicas e os fatores psicossociais (Williams, 1990).

A morbidade e a mortalidade estão associadas a fatores psicossociais e comportamentais, como estresse, prática religiosa, estado civil hábitos de alimentação e exercício físico, tabagismo e suporte social. Em geral, pessoas casadas, sobretudo os homens, exibem menor mortalidade que aquelas não casadas (Roger, 1995; Lillard & Waite, 1995). Também há evidências de que pessoas envolvidas intensamente em preceitos religiosos apresentem menor mortalidade, se comparadas às menos envolvidas, o que é visto como um resultado do papel dos laços e suporte sociais na saúde dos indivíduos (Rogers, 1996). Com relação ao estresse, ele tem sido medido por variáveis como desemprego, divórcio, morte de entes queridos, dentre outras. As evidências apontam uma relação positiva entre ele e a presença de uma série de doenças. Os laços sociais e o controle pessoal também são vistos como fatores que agem nos processos biológicos e afetam a vulnerabilidade a doenças e mudam as chances de morte (Williams, 1990).

Os fatores psicossociais, por sua vez, variam em função de características socioeconômicas dos indivíduos, tais como renda, ocupação e escolaridade. A obesidade, por exemplo, varia negativamente com a escolaridade, assim como a quantidade de cigarros fumados diariamente. Entre 1974 e 1985, a prevalência de fumantes nos Estados Unidos declinou cinco vezes mais rápido entre pessoas com algum curso universitário comparadas às pessoas menos escolarizadas (NCHS, 1981). Entre as décadas de 1980 e 1990, o tabagismo é apontado como uma das explicações para a expansão dos diferenciais educacionais da mortalidade. Estima-se que cerca de um quarto do aumento dos diferenciais entre as mulheres seja explicado por diferenciais educacionais na prática do tabagismo (Meara et al

2008). Pampel *et al* (2010) realça a influência de dos fatores socioeconômicos, principalmente da escolaridade, no estilo de vida dos indivíduos, com destaque para uso de tabaco, prática de exercícios físicos e de alimentação. No caso das bebidas alcoólicas, embora haja uma relação positiva entre seu consumo e escolaridade, o alcoolismo está associado negativamente a esta dimensão (NCHS, 1981).

A educação provê aos indivíduos um estoque de conhecimento que influencia a forma como eles interagem com outros, na intervenção em situações de crise e emergências, na busca por bens e serviços médicos, na prevenção de doenças e na promoção da saúde. Pessoas mais educadas são mais propensas a adotar comportamento saudável, a procurar cuidado médico, a perceberem problemas de saúde, a uma melhor comunicação com médicos e a um maior entendimento para seguir suas recomendações. Em contrapartida, indivíduos menos educados são mais suscetíveis a características psicossociais negativas, do ponto de vista da saúde. Eles são mais prováveis em adotar comportamento de risco, em apresentar estresse, isolamento social e falta de controle social, que são importantes determinantes da morbidade e mortalidade (Rogers *et al*, 1999b, p.118).

A principal crítica à abordagem está na ausência de teoria unificada que explique as interações entre os fatores socioeconômicos, psicossociais e os determinantes próximos da mortalidade. Em geral, são apontados padrões e tendências, mas com dados insuficientes para ir além da especulação sobre as relações causais por traz destes achados.

3. DADOS FALTANTES: CONCEITOS, PREMISSAS E MÉTODOS DE IMPUTAÇÃO

3.1. Conceitos e premissas

Consideram-se dados faltantes ou *missing* aqueles para os quais não se tem informação para todas as unidades de análise ou observações. Para Little e Rubin (2002), ao considerar uma matriz de dados em que as linhas representam observações ou unidades de análise e as colunas representam variáveis, as entradas para as quais não se têm dados são consideradas *missing*.

Heitjan e Rubin (1990, 1991) estendem o conceito para dados que tenham sofrido transformações como agrupamento, arredondamento, censura, truncamento e outras, que tenham resultado em perda parcial de informação. O conceito mais abrangente permite que métodos para tratar dados faltantes sejam adotados para lidar com problemas de qualidade de dados observados. Muitos problemas estatísticos, que não envolvem necessariamente perda de dados, podem ser reformulados e tratados com métodos criados para lidar com dados faltantes (Schafer, 1997).

Em alguns casos, a ausência dos dados pode ser intencional. Em pesquisas de inquérito, por exemplo, o desenho amostral pode contemplar ausência de informações para algumas unidades, com o intuito de reduzir custos e tempo de participação dos respondentes, incluir mais quesitos, evitar dados faltantes não planejados e aumentar qualidade das respostas prestadas (Enders, 2010; Mistler & Enders, 2012; Little & Rhemtulla, 2013).

Seja qual for o caso, os mecanismos de ligação entre o dado faltante e os seus determinantes são importantes para a escolha do método estatístico que irá lidar com o possível viés gerado pelos dados faltantes nas estimativas. Rubin (1976)⁸ criou um sistema de classificação de mecanismos com base em hipóteses sobre a probabilidade de cada valor

⁸ Rubin (1976) é responsável por desenvolver uma teoria sobre dados faltantes que se tornou referência na literatura sobre o tema.

de uma variável ser ou não faltante e os fatores associados a ela. Foram estabelecidos três mecanismos: i) dados faltantes completamente aleatórios (*missing completely at random* – MCAR); ii) dados faltantes aleatórios (*missing at random* – MAR); e iii) dados faltantes não aleatórios (*missing not at random* – NMAR). Essa classificação passou a ser amplamente adotada e utilizada em trabalhos que envolvem tratamento de dados faltantes (Graham *et al*, 1997; Schafer & Graham, 2002; Enders, 2010; Camargos *et al*, 2011). Os três mecanismos de Rubin são bem descritos e exemplificados em Little e Rubin (2002) e Enders (2010).

Quando os dados faltantes são classificados como MCAR, as razões para a falta do dado não se encontram relacionadas a quaisquer respostas observadas. Considerando

- $Y = y_{ij}$ representa a matriz ($n \times p$) de variáveis de interesse com dados faltantes, com i -ésima linha $y_i = (y_{i1}, \dots, y_{ip})$, sendo $j = 1, \dots, p$;
- y_i representa os valores de Y para a observação i , sendo $i = 1, \dots, n$;
- $M = (M_{ij})$ é a matriz de indicação de dados faltantes, sendo $M_{ij} = 1$ se y_{ij} é faltante e $M_{ij} = 0$ se não;
- $f(M|Y, \phi)$ representa a distribuição de M condicionada a Y ;
- ϕ é um ou mais parâmetro(s) não conhecido(s), que descreve a relação entre M e Y ;
- Y_{obs} representa os valores observados de Y ; e
- Y_{mis} os valores faltantes de Y (*missing*);

sob o mecanismo MCAR,

$$f(M|Y, \phi) = f(M|\phi), \text{ para todo } Y, \phi. \quad \text{Eq 3.1}$$

O mecanismo MAR, também denominado mecanismo ignorável, é uma hipótese menos restritiva. Neste caso, as razões da ausência de Y estão relacionadas a outras variáveis com respostas completas, mas não à variável com dados faltantes. Assim,

$$f(M|Y, \phi) = f(M|Y_{obs}, \phi), \text{ para todo } Y_{mis}, \phi. \quad \text{Eq 3.2}$$

Por fim, o mecanismo NMAR ou não ignorável ocorre quando a distribuição de M é condicionada aos valores da própria variável com dado faltante na matriz Y (Y_{mis}):

$$f(M|Y, \Phi) = f(M|Y_{obs}, Y_{mis}, \Phi) \text{ para todo } Y_{mis}, \Phi. \quad Eq. 3.3$$

Na realidade, não há como estimar o(s) valor(es) do(s) parâmetro(s) ϕ . As técnicas desenvolvidas permitem trabalhar com imputação partindo de alguma hipótese sobre estes mecanismos. A maior parte delas assume hipótese MCAR, mas são pouco utilizadas por produzirem estimativas viesadas dos parâmetros de interesse, se a hipótese não é realista. Mas em sua apresentação da teoria de dados faltantes, Rubin (1976) apresenta a imputação múltipla (IM) e a estimação por máxima verossimilhança (algoritmo EM) como capazes de produzir estimativas válidas sob as hipóteses MAR e MCAR. Esses métodos serão discutidos nas seções 3.3 e 3.4.

Embora haja métodos sugeridos para imputação sob a hipótese de NMAR, eles são bem restritivos uma vez que os pressupostos assumidos são bastante fortes (Enders, 2010). Logo, entender as consequências de assumir as hipóteses de MAR ou MCAR, quando elas não são completamente satisfeitas, tem sido um ponto de interesse na literatura. A gravidade do problema depende do objetivo de cada análise. Se o objetivo for investigar a distribuição marginal de Y , o mecanismo é menos importante. Se a distribuição condicional de Y for de interesse, ele se torna relevante (Little & Rubin, 2002).

Quanto às consequências de violação da hipótese MAR, Graham *et al* (1997, p.354) compararam resultados de uma série de simulações usando diferentes métodos de imputação e apontam que a violação de MAR ou MCAR se desponta como um problema de menor gravidade. Outros estudos corroboram estes achados e sugerem que técnicas que envolvam imputação múltipla e estimação por máxima verossimilhança podem prover estimativas razoáveis, mesmo com dados NMAR. Em geral, as consequências da violação dependem do grau de relação entre os dados faltantes e suas causas. A recomendação tem sido adotar modelos diferentes de imputação para identificar a sensibilidade dos resultados aos modelos (Collins *et al*, 2001; Schafer & Graham, 2002; Allison 2009). As principais estratégias de imputação são discutidas nas seções 3.2 a 3.4.

3.2. Principais técnicas de imputação

A imputação tem sido uma técnica bastante usada para lidar com dados faltantes. Trata-se de uma forma de estimar valores para estes dados a partir das unidades com dados observados. A sua aplicabilidade tem se expandido para outras situações. A imputação múltipla já foi usada para substituir dados de idade com problema de preferência por dígitos. A substituição foi realizada usando a idade reportada e outras variáveis disponíveis no estudo (Heitjan & Rubin, 1990). Em outro projeto, a técnica foi usada para compatibilizar classificações ocupacionais entre os censos de 1970 e 1980 nos Estados Unidos. Os códigos de ocupações do censo de 1980, que não estavam presentes em 1970, foram tratados como *missing* e a imputação foi a estratégia assumida para minimizar o problema (Clogg *et al*, 1991; Rubin & Schenker, 1991).

Os principais métodos de imputação são discutidos e exemplificados em Little & Rubin (2002), Allison (2009) e Enders (2010). A imputação pode se apoiar em modelos explícitos, em que os pressupostos para construção da distribuição são explícitos, ou modelagem implícitas, em que os pressupostos não são explícitos, mas cuja escolha implica em uma distribuição subjacente. Neste último caso, é necessário avaliar se a distribuição subjacente é razoável. Os métodos podem envolver imputação de um único valor (imputação simples) ou de mais valores (imputação múltipla).

Dentre os métodos de modelos explícitos, destacam-se as imputações pela média, pela média condicional e por regressão estocástica. A imputação pela média consiste em substituir os valores faltantes pela média obtida para a variável usando os dados observados. Embora seja um método simples, ele é desvantajoso por reduzir artificialmente a variância da variável tratada ao adotar valores do centro da distribuição.

O método da imputação pela média condicional trata da substituição dos valores faltantes pela média da variável tratada condicionada aos valores assumidos pelas variáveis completas. Supondo um que o indivíduo i apresente dado faltante para variável Y_K e que possua as demais variáveis (Y_{p-1}) completas, o valor de Y_{ik} será:

$$\hat{Y}_{ik} = \tilde{\beta}_{K0,1,2,\dots,K-1} + \sum_{j=1}^{K-1} \tilde{\beta}_{Kj,1,2,\dots,K-1} Y_{ij}. \quad Eq\ 3.4$$

Assim como a média incondicional, a média condicional subestima a variância da variável tratada, embora em menor nível. Além do mais é necessário assumir pressuposto de MCAR (ou pelo menos MAR) para adotá-las.

A imputação por regressão estocástica adiciona um termo de erro aleatório z_{ik} com distribuição normal:

$$\hat{Y}_{ik} = \tilde{\beta}_{K0,1,2,\dots,K-1} + \sum_{j=1}^{K-1} \tilde{\beta}_{Kj,1,2,\dots,K-1} y_{ij} + \hat{z}_{ik} \quad \text{Eq 3.5}$$

Este método possui a vantagem de produzir estimativas válidas, pois amplia a extensão de valores possíveis. Porém, provoca perda de eficiência em estimativas de parâmetros no modelo de análise, por não adicionar incerteza em relação ao dado imputado. Essa incerteza, conforme será discutido no item 3.3, está relacionada à variabilidade amostral e às causas do dado faltante.

No grupo de modelos implícitos, destacam-se as imputações por substituição, por *cold deck* e por *hot deck*. No método da substituição, os dados faltantes são substituídos por informações de unidades que inicialmente não estavam na amostra. A imputação é feita ainda na etapa de coleta de dados, em caso de pesquisas de inquérito. O problema é que as unidades não respondentes podem diferir sistematicamente da respondente. O *cold deck* é o método em que os dados faltantes são supridos por valores constantes doados por unidades de uma fonte externa de dados. Quando estes dois métodos são usados, em geral, os dados passam a ser considerados completos, ignorando-se as consequências de sua imputação (Little & Rubin, 2002).

O método do *hot deck* consiste na imputação de valores de outras unidades de análises similares. O processo pode envolver esquemas elaborados para a escolha das unidades doadoras de informações. As unidades de análise devem ser classificadas com base em características observadas dos doadores, dentro de uma determinada classe, para serem selecionados aleatoriamente. A vantagem do método é que ele tende a subestimar a variância dos dados imputados em um grau menor, se comparado à imputação pela média e por regressão. De qualquer forma, por não adicionar incerteza, ele também pode causar viés em medidas de associação e estimadores, assim como os demais métodos de imputação simples.

Para resolver o problema da incerteza, Rubin (1978) introduziu o método de Imputação Múltipla (IM), que propõe a imputação de mais de um valor para cada dado faltante. Assim, são obtidos vários bancos de dados completos que são analisados individualmente e combinados.

Outro método bastante usado para lidar com estimativas com *missing* é a estimação por Máxima Verossimilhança (MV). Neste caso, os dados faltantes são tratados como variáveis aleatórias que podem ser removidas da função de verossimilhança, como se não tivessem sido amostrados. O problema do dado faltante é contornado durante o ajuste do modelo. O algoritmo expectativa-maximização (*expectation-maximization* - EM), que tem grande parte do seu desenvolvimento atribuído a Dempster *et al* (1977), é um processo iterativo por MV bastante usando. Assim como a IM, EM é uma estratégia moderna e superior frente aos outros métodos para tratamento de dados faltantes.

Com base nas suas vantagens, IM e o EM são recomendados mesmo em casos em que NMAR não pode ser descartado. A aplicação destes métodos esteve limitada dada a dificuldade prática de lidar com os cálculos envolvidos. Mas a partir dos anos 1990, o desenvolvimento de programas estatísticos e a incorporação de IM e EM em seus pacotes tem incentivado suas aplicações (Rubin, 1996; Schafer & Graham, 2002; Alisson, 2009). Dada a importância e aplicabilidade destes modelos, eles serão discutidos separadamente nas seções 3.3 e 3.4.

3.3. Imputação Múltipla (IM)

O método de IM foi proposto por Rubin (1978) como uma forma de imputação que refletisse a variabilidade e incorporasse a incerteza na distribuição dos dados faltantes. O método foi desenvolvido sob uma perspectiva teórica Bayesiana e envolve repetições que simulam uma distribuição *a posteriori* para os valores faltantes. O interesse é estimar a distribuição preditiva dos valores faltantes com base na distribuição dos dados observados. Essa distribuição depende dos modelos adotados, de forma que o ponto central da discussão está na sensibilidade das inferências frente às escolhas. Revisões abrangentes do método foram realizadas por Rubin (1987) e Rubin (2004). A descrição apresentada nesta

seção é baseada em grande parte nos trabalhos de Rubin (1978; 2004) e Little e Rubin (2002).

A IM tem algumas vantagens sobre a imputação simples que a torna atraente para o propósito deste trabalho. Inicialmente, a IM permite o uso de informações disponíveis ao pesquisador para gerar incerteza em torno dos valores faltantes. Com isso, a eficiência das estimativas obtidas com os dados imputados torna-se maior. Essa incerteza refere-se à variabilidade amostral e às razões pelas quais o dado é faltante na pesquisa. A produção de valores múltiplos para o dado faltante deve refletir a variabilidade amostral. Quando apenas um modelo é usado, a IM pode ser chamada de método das repetições. Mas é possível utilizar mais de um modelo e, assim capturar a incerteza sobre as causas da ausência do dado. Assim, as diferenças de resultados observadas servem como base para análise de sensibilidade aos modelos. Elas indicam a (in)capacidade dos dados observados em prever os dados faltantes.

Outra vantagem da IM é que quando o método de repetições é utilizado, as inferências de interesse podem ser obtidas combinando os resultados múltiplos por meio de técnicas para dados completos. Outra possibilidade seria repetir métodos para dados completos em todos os bancos obtidos. Isso torna o método acessível mesmo para usuários que não dominam métodos estatísticos mais sofisticados. Quando modelos explícitos são utilizados para a imputação, torna-se computacionalmente mais viável o uso da IM para realizar inferências, comparativamente a outras análises matemáticas sofisticadas.

3.2.1. Procedimento geral para imputação múltipla

A IM trata de imputar m valores de uma distribuição *a posteriori* com base em um ou mais modelos escolhidos. Sendo Y a matriz de dados de uma amostra $n \times p$, em que n representa as unidades e p as variáveis, assume-se: que Y possui linhas independentes; uma distribuição *a posteriori* $pos(\theta)$ do parâmetro θ , de forma que uma extração aleatória de θ de $pos(\theta)$ seja possível. A imputação é, portanto, uma extração aleatória da distribuição preditiva dos valores faltantes a partir dos valores observados, com base no modelo escolhido. A IM pode ser resumida em cinco etapas.

- 1) A etapa 1 envolve a escolha do(s) modelo(s) que relacionam Y_{obs} às covariáveis de X observadas. Nesta etapa, o parâmetro ou conjunto de parâmetros de interesse será representado por θ .
- 2) A etapa 2 consiste na estimação da distribuição a *posteriori* com base no parâmetro θ estimados a partir dos modelos escolhidos na etapa 1.
- 3) Considerando $\ell = 1, \dots, r_K$, sendo r_K o número de réplicas do K -ésimo modelo, estimam-se os valores $y_{K\ell}$ da distribuição a *posteriori* e cria-se um vetor M com m dados imputados.
- 4) Com os valores de $y_{K\ell}$ com $K = 1, \dots, K$; $\ell = 1, \dots, r_K$, formam-se I vetores M com dados completos, sendo $I = \sum_{k=1}^K r_k$.
- 5) A etapa 5 trata da análise dos conjuntos de dados imputados e validade dos valores produzidos.

3.2.2. Escolha do modelo

A escolha do modelo é a parte crítica da IM. Ela envolve tanto a escolha funcional como das variáveis independentes. O modelo não necessariamente deve representar uma relação de causa e efeito. Trata-se de um dispositivo que deve preservar importantes aspectos da distribuição conjunta das variáveis envolvidas no modelo. A relação entre essas variáveis, expressa por uma regressão, por exemplo, deve ser automaticamente preservada após a imputação.

Supondo que os dados faltantes são ignoráveis (MAR), a distribuição de Y_{mis} está condicionada à distribuição dos valores observados para uma ou mais covariáveis X e Y_{obs}

$Pr(Y_{mis}|X, Y_{obs})$, sendo

$$Pr(Y_{mis}|X, Y_{obs}) = Pr(X, Y) / \int Pr(X, Y) dY_{mis}. \quad Eq 3.6$$

Se os dados faltantes forem ignoráveis, $Pr(X, Y)$ é modelada como independente e identicamente distribuída (*i.i.d.*):

$$Pr(X, Y) = \int Pr(X, Y|\theta) Pr(\theta) d\theta = \int [\prod_{i=1}^N f_{XY}(X_i, Y_i|\theta)] Pr(\theta) d\theta \quad eq. 3.7$$

A função $f_{XY}(X_i, Y_i|\theta)$ representa a distribuição conjunta de X e Y em função de um ou mais parâmetros θ , podendo ser decomposta em

$$f_{XY}(X_i, Y_i|\theta) = f_{Y|X}(Y_i|X_i, \theta_{Y|X}) f_X(X_i|\theta_X) \quad eq. 3.8$$

em que $\theta_{Y|X}$ e θ_X são funções de θ . Supondo que X_i é completamente observada, quando $\theta_{Y|X}$ e θ_X são modelados *a priori* como independentes, eles assumem uma distribuição *a posteriori* independente e a distribuição *a posteriori* de $\theta_{Y|X}$ não requer uma especificação para $f_X(\cdot|\cdot)$.

Qualquer variável envolvida pode ser usada como dependente ou independente na análise *a posteriori* do processo de imputação. Algumas vezes torna-se necessário criar indicadores para diferenciar grupos que possam apresentar comportamento diferenciado dentro da amostra e proceder com as imputações separadamente (Schafer & Graham, 2002).

O método mais comum de predição de Y a partir de covariáveis X é o modelo de regressão linear normal. Algumas transformações nas variáveis podem ser usadas quando a hipótese de normalidade é quebrada. Variáveis binárias ou ordinais, por exemplo, podem ser imputadas sob normalidade e então arredondadas. Simulações apontam que variáveis que inicialmente não seguem uma distribuição normal de probabilidade são tratadas como normais e os resultados das regressões são satisfatórios (Schafer, 1997; Schafer, 1999).

Quando a variável dependente é dicotômica ou é limitada a poucos valores, um modelo de regressão logística é apropriado. Supondo que Y assumira valores 0 ou 1. Assim tem-se que

$$f(Y_i|X_i, \theta) = \text{logit}^{-1}(X_i\theta)^{Y_i} [1 - \text{logit}^{-1}(X_i\theta)]^{1-Y_i} \text{ e a função inversa}$$

$$\text{logit}^{-1}(a) = \frac{\exp(a)}{1 + \exp(a)} \text{ que corresponde a função logística } \text{logit}(a) = \log \frac{a}{1-a}.$$

Em grandes amostras, a distribuição do(s) parâmetro(s) θ pode se aproximar de uma normal. Assim, pode-se obter a média *a posteriori* de θ por aproximação usando máxima verossimilhança. O primeiro passo é estimar θ (que será denominado θ^*) de $N(\hat{\theta}, V(\hat{\theta}))$. O segundo é calcular $\text{logit}^{-1}(X_i\theta^*)$ para $i \in \text{mis}$. O terceiro passo seria retirar n_0 (0, 1) números aleatórios independentes, u_i , $i \in \text{mis}$. Se $u_i > \text{logit}^{-1}(X_i\theta^*)$, $Y_i = 0$; do contrário, $Y_i = 1$.

Quanto à escolha das variáveis, observa-se que a inclusão de preditores tende a aumentar a plausibilidade da hipótese MAR (Schafer, 1977; Van Buren, 1999). No entanto, a escolha do número de variáveis deve levar em conta a viabilidade em função de perda de graus de liberdade e problemas de multicolinearidade.

Van Buren e outros (1999) propõem uma estratégia que envolve quatro passos para esse processo. Inicialmente, devem-se incluir todas as variáveis de interesse para a análise com dados completos, com recomendação de um limite entre 15 e 25 variáveis. No segundo passo, incluem-se as variáveis com possíveis relações com os dados faltantes da variável de interesse na imputação. O terceiro passo trata da adição de variáveis que expliquem consideravelmente a variância da variável de interesse. Na etapa final, removem-se as variáveis que apresentem uma proporção elevada de *missing* dentre as observações que vão sofrer a imputação.

Em algumas situações, é possível realizar a imputação sem adoção de um modelo, usando o método de aproximação Bayesiana por amostragem *bootstrap* (*approximate Bayesian bootstrap* - *ABB*) desenvolvido por Rubin e Schenker (1986). O método *ABB* parte da imputação de uma amostra de n_1 valores de forma aleatória dos Y_{obs} para criar Y_{obs}^* . Em seguida são extraídos aleatoriamente e com reposição $n_0 = n - n_1$ componentes de Y_{mis} a partir de Y_{obs}^* . O método é apropriado para amostras grandes (Rubin 2004).

3.2.3. Etapa de imputação com dado faltante ignorável

A tarefa de imputação consiste em obter os valores para Y_{mis} com base na sua distribuição *a posteriori*,

$$Pr(Y|X, Y) = \int Pr(Y_{mis}|X, Y, \theta) Pr(\theta|X, Y) d\theta. \quad Eq 3.9$$

A ideia é estimar um valor da distribuição a *posteriori* $Pr(\theta|X, Y_{obs})$, que para fins práticos assume a notação θ^* , e estimar um valor para Y_{mis} a partir de sua distribuição condicional e do valor de θ^* . O processo é repetido m vezes criando m extrações da distribuição conjunta a *posteriori* de (Y_{mis}, θ) . Considerando o caso mais simples de Y_i univariado e $\theta_{Y|X}^*$ com o valor de obtido de $\theta_{Y|X}$ por meio de sua distribuição a *posteriori*, cada valor faltante de Y_i é imputado de forma independente, de acordo com $f_{Y|X}(Y_i|X_i, \theta_{Y|X}^*)$.

Se o dado faltante é ignorável, com base no Teorema de *Bayes*, o valor de $Y_{i,mis}$ tem distribuição a *posteriori* independente e definida por θ por meio de $\theta_{Y|X}$:

$$Pr(Y_{mis}|X, Y_{obs}, \theta) = \prod Pr(Y_{i,mis}|X_i, Y_{i,obs}, \theta_{Y|X}). \quad Eq 3.10$$

3.2.4. Considerações sobre o número de imputações e de modelos

Embora não exista uma resposta para a quantidade de imputações e de modelos a serem adotados, existem situações nas quais se é possível perceber se os dados faltantes representam um problema maior, sendo do tipo NMAR. Em algumas situações, poucas réplicas de um modelo apontam diferenças pequenas entre respostas e ausências que resultam em grandes diferenças entre os dados imputados (Rubin, 1978).

A eficiência de uma estimativa baseada em m imputações em relação a um número infinito de imputações é aproximadamente $(1 + \lambda/m)^{-1/2}$, sendo λ a proporção de dados faltantes. Em aplicações típicas resultados consistentes são obtidos usando $3 \leq m \leq 5$ (Schafer, 1997; Rubin, 2004).

3.2.5. Análise dos bancos completos

Após o exercício de imputação, tem-se I vetores M de valores imputados. Considerando a utilização de apenas um modelo de imputação, sendo este razoável, se a variância entre os

valores imputados for relativamente grande, tornam-se necessárias restrições adicionais para procedimento da análise pretendida.

O erro padrão das estimativas obtidas com a IM tem duas fontes: variância intra e inter imputações. A variância intra imputação V_W é a média aritmética das variâncias dos m bancos de dados formados:

$$V_W = \frac{1}{m} \sum_{t=1}^m SE_t^2$$

em que SE_t^2 é o quadrado do erro padrão do conjunto de dados t . A variância V_W reflete a variabilidade amostral.

A presença de *missing* deve acrescentar algum ruído às estimativas dos parâmetros e, conseqüentemente, aumentar o erro padrão. A imputação simples não lida com isso, pois assume um valor como real para preenchimento dos dados faltantes. A IM, por outro lado, provê fonte adicional de erro amostral. As variações observadas nas estimativas de θ entre os m conjuntos de dados representam flutuações amostrais adicionais presentes quando há dados *missing*. Trata-se da variância inter imputações - V_B - que é dada por

$$V_B = \frac{1}{m} \sum_{t=1}^m (\hat{\theta}_t - \bar{\theta})^2$$

em que $\hat{\theta}_t$ é o parâmetro estimado do conjunto de dados t , e $\bar{\theta}$ é a estimativa da média pontual dada por

$$\bar{\theta} = \frac{1}{m} \sum_{t=1}^m \hat{\theta}_t.$$

A variância total V_T é dada por

$$V_T = V_W + V_B + \frac{V_B}{m}$$

em que $\frac{V_B}{m}$ representa o erro adicional cuja a fonte é a estimativa da média. Assim ele reflete a variância da média e serve como um fator de correção para uso de imputações com m finito.

3.2.6. Limitações da IM

Uma das grandes desvantagens associadas à IM tem sido a quantidade de trabalho envolvida no processo. Como se adotam múltiplos bancos de dados, o método requer mais espaço de armazenamento e mais tempo para análise de múltiplos resultados. No entanto, conforme discutido no item 3.2.4, quando a fração de dados faltante é muito pequena, o valor de m pode ser bem modesto, minimizando estes problemas. Ademais, com o desenvolvimento de novos pacotes computacionais desde os anos 1990, o esforço para aplicação da IM tem se tornado substancialmente menor, não requerendo mais trabalho que os métodos de imputação simples (Schafer, 1999).

Com relação à robustez dos resultados, uma das principais críticas tem sido o uso de simulação, que pode acrescentar ruído para os dados imputados. Entretanto, Rubin (1996; 2004) destaca que a proposta da IM não é criar informação, mas representar os dados observados e sua incerteza. As evidências apontam que poucas simulações são necessárias para se alcançar resultados válidos. Alguns estudos indicam que a imputação múltipla apresenta resultados melhores que os de imputação simples, mesmo quando modelos mais simples são utilizados (Heitjan & Rubin, 1990).

3.4. O algoritmo EM

O algoritmo EM é um método iterativo para estimação por MV para dados incompletos cuja formalização e estabelecimento de suas propriedades são atribuídas a Dempster *et al* (1977). A descrição do método realizada nesta seção tem como referências básicas as apresentações de Schafer (1997), Little e Rubin (2002) e Enders (2010).

Cada interação procedida pelo método envolve uma etapa de estimação da expectativa (passo *E*) e outra de maximização (passo *M*). Considerando a função log-verossimilhança $\ell(\theta|Y)$ como sendo o logaritmo natural da função de verossimilhança $L(\theta|Y)$, os passos de EM consistem em: i) substituir os valores faltantes usando algum método mais simples (imputação pela média ou assumir uma distribuição uniforme, por exemplo); ii) estimar parâmetros desejados por uma função de verossimilhança $\ell(\theta^{(0)}|Y_{obs})$; iii) reestimar os valores faltantes usando os parâmetros estimados em ii; iv) reestimar os parâmetros $\ell(\theta^{(1)}|Y_{obs})$ e assim por diante, $\ell(\theta^{(2)}|Y_{obs}) \dots \ell(\theta^{(t)}|Y_{obs})$, até que o valor de θ alcance a convergência na *t-ésima* interação.

Considerando a função com dados completos $\ell(\theta|Y)$ fatorada em

$$\ell(\theta|Y) = l(\theta|Y_{obs}) + \ln P(Y_{mis}|Y_{obs}, \theta) + c, \quad \text{eq 3.12}$$

a distribuição preditiva do dado faltante, denotada por P , captura a relação entre Y_{mis} e θ . Como essa distribuição não é conhecida previamente, ela pode ser substituída pela média da distribuição de $P(Y_{mis}|Y_{obs}, \theta^{(t)})$, sendo $\theta^{(t)}$ uma estimativa preliminar de θ . Tal média produz

$$Q(\theta|\theta^{(t)}) = l(\theta|Y_{obs}) + H(\theta|\theta^{(t)}) + c \quad \text{eq 3.13}$$

sendo

$$H(\theta|\theta^{(t)}) = \int \ln P(Y_{mis}|Y_{obs}, \theta) P(Y_{mis}|Y_{obs}, \theta^{(t)}) dY_{mis}, \quad \text{eq 3.14}$$

de modo que

$$Q(\theta|\theta^{(t)}) = \int \ell(\theta|Y) P(Y_{mis}|Y_{obs}, \theta^{(t)}) dY_{mis}. \quad \text{eq 3.15}$$

Se $\theta^{(t+1)}$ for o valor de θ que maximiza $Q(\theta|\theta^{(t)})$, ele se torna um estimador melhor que $\theta^{(t)}$ uma vez que a verossimilhança obtida com os dados observados é no mínimo tão elevada quando $\theta^{(t)}$

$$\ell(\theta^{(t+1)}|Y_{obs}) \geq \ell(\theta^{(t)}|Y_{obs}).$$

Logo,

$$\ell(\theta^{(t+1)}|Y_{obs}) - \ell(\theta^{(t)}|Y_{obs}) = Q(\theta^{(t+1)}|\theta^{(t)}) - Q(\theta^{(t)}|\theta^{(t)}) + H(\theta^{(t)}|\theta^{(t)}) - H(\theta^{(t+1)}|\theta^{(t)}).$$

eq 3.16

Como $\theta^{(t+1)}$ é escolhida de forma que $Q(\theta^{(t+1)}|\theta^{(t)}) \geq Q(\theta^{(t)}|\theta^{(t)})$ para todo θ , então $Q(\theta^{(t+1)}|\theta^{(t)}) - Q(\theta^{(t)}|\theta^{(t)}) \geq 0$. Considerando que

$$H(\theta^{(t)}|\theta^{(t)}) - H(\theta^{(t+1)}|\theta^{(t)}) = \int \ln \frac{P(Y_{mis}|Y_{obs}, \theta^{(t)})}{P(Y_{mis}|Y_{obs}, \theta^{(t+1)})} P(Y_{mis}|Y_{obs}, \theta^{(t)}) dY_{mis} \quad \text{eq 3.17}$$

e como $x \log x$ é convexa, então esse termo da equação também é não negativo.

A etapa E , que consiste em encontrar um valor esperado para os dados faltantes com base nos observados. Trata do cálculo de $Q(\theta|\theta^{(t)})$ pela média da verossimilhança $\ell(\theta|Y)$ sobre $P(Y_{mis}|Y_{obs}, \theta^{(t)})$. A etapa M é a mesma que envolve a maximização de uma função de verossimilhança qualquer que, neste caso, é $Q(\theta|\theta^{(t)})$, com $t = 0, 1, 2, \dots$.

Uma das vantagens do método é que, sob certas condições, a convergência é garantida para um ponto estacionário de $\ell(\theta|Y_{obs})$. Isso ocorre, uma vez que cada interação aumenta a probabilidade $\ell(\theta|Y_{obs})$ e, se ℓ é bem delimitada, a sequência $\ell(\theta^{(t)}|Y_{obs})$ converge para o valor estacionário de $\ell(\theta|Y_{obs})$. A taxa de convergência é linear e proporcional aos dados observados. Desse modo, uma das desvantagens é o tempo gasto para a convergência, quando a proporção de dados faltantes é muito elevada.

4. BASE DE DADOS E METODOLOGIA

4.1. Fonte dos dados básicos

4.1.1. Censo Demográfico 2010

As estimativas pretendidas foram realizadas com base nos microdados da amostra do Censo Demográfico de 2010 do IBGE⁹. Ela tem sido apontada como uma fonte importante para estudos de mortalidade e seus diferenciais socioeconômicos (Queiroz & Sawyer, 2012). O questionário básico e o da amostra incluem um quesito sobre morte de indivíduos residentes em domicílios particulares, para o ano de referência da pesquisa: agosto de 2009 a julho de 2010. O respondente também é inquirido sobre o sexo e idade da pessoa falecida.

Dentre as vantagens em se utilizar o censo demográfico para o estudo de mortalidade, destaca-se o fato de prover também informações socioeconômicas e geográficas do domicílio e da comunidade, onde os indivíduos se encontravam inseridos. Dentre as desvantagens, destaca-se a forma de mensuração de mortalidade que pode ser fonte de sub ou superenumeração de óbitos. Como a pergunta é realizada no nível do domicílio, os óbitos ocorridos em domicílios dissolvidos no período de referência do Censo ou de domicílios unipessoais, onde houve óbitos, não são captados. Há ainda a possibilidade de óbitos serem reportados mais de uma vez, caso a pessoa falecida tenha transitado em mais de um domicílio no período de referência. Outra possível fonte de erros relaciona-se ao período de referência, que pode gerar subestimação da óbitos por erros de memória. Erros de declaração de idade também tem sido um problema recorrente para todos os censos.

Por outro lado, uma avaliação dos dados do Censo aponta boa qualidade das informações de óbitos em relação à estrutura etária da mortalidade. Problemas maiores na cobertura foram encontrados para as idades mais avançadas, sobretudo acima de 60 anos. De

⁹ A extração do banco de pessoas e domicílios foi realizada usando o programa *Stata* 14.0 (Stata.Corp, 2015a) por meio da plataforma *Datazoom*, desenvolvida pelo Departamento de Economia da PUC-Rio.

qualquer forma, estas devem ser questões avaliadas, assim como suas consequências para os resultados pretendidos em cada estudo (Queiroz e Sawyer, 2012).

4.1.2. Sistema de Informações sobre Mortalidade - SIM

Para a análise de consistência das imputações e das estimativas de mortalidade, foram utilizados os dados do Sistema de Informações sobre Mortalidade (SIM) do Ministério da Saúde. O SIM é a principal fonte de estudos de mortalidade no Brasil, mas pouco utilizada para investigações de diferenciais de escolaridade por apresentar elevado grau de perda desta variável.

A partir de 2009 foram concluídas novas declarações de óbito (DO) com recomendação do Ministério da Saúde de que fossem privilegiadas em relação às antigas a partir de 2011 (MS 2009; MS 2011). Um dos quesitos com mudanças foi a escolaridade do falecido, que passou a ser coletada em ciclos de ensino seguida da série concluída. São seis categorias: sem escolaridade, ensino fundamental I, ensino fundamental II, nível médio, superior e série do ciclo. Com isso, há uma expectativa de melhoria na cobertura e qualidade da informação. Entretanto, como circulam ainda os dois modelos de declaração, os dados continuam sendo disponibilizados no formato antigo, por meio de compatibilização: 0, 1 a 3, 4 a 7, 8 a 11 e 12 ou mais anos de escolaridade.

A despeito destas mudanças, o percentual de perda ainda é elevado nas Unidades da Federação. Mas, para o município de São Paulo, a escolaridade é reportada para 92,3% dos óbitos entre adultos de 25 a 59 anos, no período de agosto de 2012 a julho de 2013¹⁰. Esse dado torna a base atrativa para investigar diferencial de escolaridade e utilizá-la para avaliar consistência das imputações. Para os 7,7% de perda, foi aplicada a imputação simples por *hotdeck* supondo que eles sejam MCAR.

Contudo, há de se manter cautela em relação à qualidade da informação prestada. Uma primeira fonte de erro é a forma de coleta do dado, uma vez que outra pessoa presta a informação sobre a escolaridade do indivíduo falecido. Outra fonte de erros é a compatibilização utilizada pelo MS antes de publicar os dados, que é apresentada no QUADRO 1. Para as informações de séries não declaradas, a observação é incluída no

¹⁰ Essa informação foi obtida a partir da observação dos microdados do MS/SIM.

grupo de menor escolaridade dentro do ciclo de ensino, o que causa distorção caso o dado faltante corresponda a uma série mais elevada.

Em termos de cobertura de óbitos, os registros de São Paulo possuem alta confiabilidade. Estudo de Queiroz (2012) aponta grau de cobertura de praticamente 100% para o estado, dado que torna a base do SIM adequada para avaliar as estimativas obtidas com o Censo.

QUADRO 1: Regras de conversão do campo escolaridade em séries para escolaridade em anos, para transição do uso dos modelos antigos e novos da declaração de óbito

COLETA EM DOIS CAMPOS DO NOVO FORMULÁRIO			CONVERSÃO PARA O CAMPO ANTIGO - anos de escolaridade FUTURA FORMA DE DIVULGAÇÃO, COM BASE NA COLETA FEITA PELO NOVO FORMULÁRIO			
Nível	descreve nível	Série	Escolar agregado 2	Descreve escolar agregado 2	Escolar agregado 1	Descreve escolar agregado 1
0	Sem escolaridade	desabilita	1	Nenhuma	00	Sem escolaridade
1	Fundamental I (1a a 4a série)	1	2	De 1 a 3	01	Fundamental I incompleto
1	Fundamental I (1a a 4a série)	2	2	De 1 a 3	01	Fundamental I incompleto
1	Fundamental I (1a a 4a série)	3	2	De 1 a 3	01	Fundamental I incompleto
1	Fundamental I (1a a 4a série)	4	3	De 4 a 7	02	Fundamental I completo
1	Fundamental I (1a a 4a série)	branco	2	De 1 a 3	10	Fundamental I incompleto ou inespecificado
2	Fundamental II (5a a 8a série)	5	3	De 4 a 7	03	Fundamental II incompleto
2	Fundamental II (5a a 8a série)	6	3	De 4 a 7	03	Fundamental II incompleto
2	Fundamental II (5a a 8a série)	7	3	De 4 a 7	03	Fundamental II incompleto
2	Fundamental II (5a a 8a série)	8	4	De 8 a 11	04	Fundamental II completo
2	Fundamental II (5a a 8a série)	branco	3	De 4 a 7	11	Fundamental II incompleto ou inespecificado
3	Médio (antigo 2o grau)	1	4	De 8 a 11	05	Ensino médio incompleto
3	Médio (antigo 2o grau)	2	4	De 8 a 11	05	Ensino médio incompleto
3	Médio (antigo 2o grau)	3	4	De 8 a 11	06	Ensino médio completo
3	Médio (antigo 2o grau)	branco	4	De 8 a 11	12	Ensino médio incompleto ou inespecificado
4	Superior incompleto	desabilita	5	De 12 ou mais	07	Superior incompleto
5	Superior completo	desabilita	5	De 12 ou mais	08	Superior completo
9	Ignorado	desabilita	9	Ignorado	09	Ignorado
	branco	desabilita	branco	não informado	branco	não informado

Fonte: Construído por MS (2011)

4.2. População de interesse e dimensão da amostra

O interesse foi estimar a mortalidade adulta entre 25 e 59 anos para o estado e o município de São Paulo. A adoção do limite inferior se justifica por se uma idade suficientemente alta para a fixação do nível de escolaridade dos indivíduos. Quanto ao limite superior, a exclusão da população idosa se deve ao menor grau de qualidade das informações de óbitos para este grupo. Conforme destacado na seção 4.1, as evidências apontam para uma redução na cobertura de óbitos em idades acima de 60 anos, o que, pelo menos em parte, pode ser explicado pela maior prevalência de domicílios unipessoais e pessoas institucionalizadas.

De fato, a proporção de pessoas vivendo em domicílios unipessoais¹¹ (Tabela 1) e em instituições são expressivamente maiores nas idades mais avançadas. Menos de 1 a cada mil pessoas abaixo de 60 anos vivia institucionalizada em São Paulo, enquanto esse valor

¹¹ Domicílios constituídos por uma pessoa responsável pelo domicílio.

atingia 4,6 e 5, respectivamente, para homens e mulheres de 60 anos ou mais. Como a mortalidade é crescente com a idade, a exclusão dos grupos mais velhos deve diminuir substancialmente o problema de cobertura de óbitos em função do tipo de unidade doméstica.

TABELA 1: Proporção de pessoas em domicílios unipessoais segundo idade e sexo, estado de São Paulo, 2010

Idade	domicílio pessoal		todos os domicílios	
	<i>n</i>	%	<i>n</i>	%
<25	5577	0.4	1357746	100.0
≥25 e <59	72333	4.0	1791293	100.0
≥60	58132	13.6	428572	100.0
Total	136042	3.8	3577611	100.0

Fonte: IBGE - Amostra do Censo Demográfico 2010 para São Paulo

Com relação à condição do domicílio, foram considerados apenas os particulares, para os quais o quesito sobre óbito foi aplicado. Eles abrangem 99,3% do total de domicílios e de pessoas da amostra. A amostra de São Paulo contém 3.671.485 de pessoas distribuídas em 1.126.686 de domicílios particulares. São 1.806.099 adultos entre 25 e 59 anos, 49% da amostra.

A menor unidade geográfica disponível para a amostra é a área de ponderação. São 10184 áreas no Brasil, sendo São Paulo a UF com maior número de áreas: 1879 (18,4%). Cada área de ponderação do estado possui, em média, 1929 pessoas e 599 domicílios particulares ocupados. O número médio de óbitos e de pessoas entre 25 e 59 anos por área é, respectivamente, 3,1 e 953 (TABELA 2 e TABELA 3)

TABELA 2: Número de pessoas e domicílios por área de ponderação, Brasil e Unidades da Federação, 2010

UF	Nº de áreas de ponderação	Pessoas				Domicílios			
		Média	Desvio padrão	Mínimo	Máximo	Média	Desvio padrão	Mínimo	Máximo
Brasil	10184	2019.30	851.99	361	17064	601.72	249.88	112	5475
Rondônia	93	2094.45	655.04	1078	3726	619.56	203.15	326	1182
Acre	31	3013.48	1002.09	1249	5578	777.52	311.40	268	1553
Amazonas	105	2803.35	901.01	1397	5920	610.75	198.08	256	1512
Roraima	28	2269.21	533.87	1506	3549	575.50	123.45	417	874
Pará	292	2491.85	929.62	1173	8837	601.25	220.74	250	2347
Amapá	36	2168.11	711.93	1270	4207	503.42	158.23	243	934
Tocantins	160	1669.43	661.48	539	4033	481.92	191.58	163	1281
Maranhão	321	2469.32	727.77	1329	4267	606.06	175.49	318	1127
Piauí	257	1930.48	779.06	617	6971	526.83	212.30	166	1890
Ceará	362	2334.20	671.34	1254	4505	645.73	183.40	349	1157
Rio Grande do Norte	226	1876.27	594.52	777	4140	528.37	165.23	213	1177
Paraíba	281	2031.71	753.33	629	4999	576.15	214.81	201	1451
Pernambuco	388	2292.49	854.96	729	10732	647.59	240.36	188	3354
Alagoas	151	2314.11	760.13	930	6800	611.43	201.71	253	1963
Sergipe	113	2168.50	762.18	949	5232	612.65	207.89	272	1345
Bahia	690	2244.35	658.34	848	6044	641.58	185.45	247	1942
Minas Gerais	1306	1912.13	625.30	403	4431	586.71	191.06	143	1361
Espírito Santo	192	2076.86	712.20	1246	6368	649.09	223.29	401	1813
Rio de Janeiro	538	2118.90	1610.65	843	17064	681.80	504.41	392	5475
São Paulo	1879	1929.75	839.60	361	14574	599.62	257.60	112	4319
Paraná	655	1967.76	831.21	709	10527	618.64	256.13	217	3269
Santa Catarina	495	1756.44	649.39	734	4349	558.27	207.37	216	1408
Rio Grande do Sul	826	1674.96	692.70	622	8428	562.38	237.34	212	3182
Mato Grosso do Sul	111	2481.75	1114.10	986	9088	778.17	353.67	337	2848
Mato Grosso	209	1840.04	673.65	554	3987	562.32	203.58	194	1241
Goiás	388	1815.95	725.05	455	4474	580.62	222.53	169	1384
Distrito Federal	51	2273.12	742.23	1236	4598	686.47	264.17	401	1808

Fonte: IBGE - Amostra do Censo Demográfico 2010

TABELA 3: Número de pessoas e óbitos entre 25 e 59 anos por área de ponderação, Brasil e Unidades da Federação, 2010

UF	Nº de áreas de ponderação	Óbitos entre 25 e 59 anos				Adultos vivos entre 25 e 59 anos			
		Média	Desvio padrão	Mínimo	Máximo	Média	Desvio padrão	Mínimo	Máximo
Brasil	10184	3.14	2.35	0	43	927.45	391.89	192	8743
Rondônia	93	3.05	1.87	0	8	955.08	301.90	493	1704
Acre	31	3.97	2.75	0	14	1170.29	503.36	385	2433
Amazonas	105	2.98	2.04	0	9	1058.67	362.05	439	2865
Roraima	28	2.89	2.25	0	8	903.93	194.59	600	1359
Pará	292	3.41	2.35	0	18	1004.64	358.03	473	3764
Amapá	36	3.42	2.18	0	8	853.97	266.75	398	1693
Tocantins	160	2.44	2.00	0	11	706.16	292.28	228	1981
Maranhão	321	3.21	2.31	0	13	956.42	267.23	496	1758
Piauí	257	2.60	2.07	0	14	808.72	349.72	241	3215
Ceará	362	3.12	2.04	0	11	985.45	277.51	507	2083
Rio Grande do Norte	226	2.14	1.80	0	9	818.38	267.28	321	2045
Paraíba	281	2.76	2.14	0	11	861.51	337.13	275	2510
Pernambuco	388	3.41	2.27	0	16	999.52	375.05	414	5000
Alagoas	151	3.66	2.45	0	13	945.11	315.34	404	3173
Sergipe	113	3.04	2.08	0	12	935.62	315.76	417	2428
Bahia	690	3.45	2.20	0	19	980.20	297.30	351	3210
Minas Gerais	1306	3.26	2.23	0	13	897.89	295.17	207	2243
Espírito Santo	192	3.37	2.30	0	14	1005.91	351.99	565	2907
Rio de Janeiro	538	3.86	3.61	0	43	1041.79	788.29	459	8743
São Paulo	1879	3.10	2.30	0	26	953.47	415.41	192	7314
Paraná	655	3.21	2.29	0	18	934.00	398.13	328	5281
Santa Catarina	495	2.78	2.37	0	27	862.48	317.26	329	2062
Rio Grande do Sul	826	2.73	2.17	0	18	816.65	337.65	305	4273
Mato Grosso do Sul	111	3.86	2.60	0	14	1148.48	546.47	486	4306
Mato Grosso	209	2.77	2.02	0	8	861.24	310.93	260	1857
Goiás	388	3.26	2.49	0	12	872.33	342.43	217	2174
Distrito Federal	51	3.00	1.97	0	8	1142.47	409.61	689	2695

Fonte: IBGE - Amostra do Censo Demográfico 2010

4.3. Aplicação da Imputação Múltipla para os dados faltantes de escolaridade

No caso do Censo, os respondentes não foram inquiridos sobre outras características dos indivíduos que faleceram no domicílio, além de sexo e idade. A ausência de informação sobre a escolaridade é intencional. Logo, não se trata de um problema comum de dado faltante, mas que pode ser tratado com metodologias de imputação. Como a ausência do dado é inerente ao desenho da pesquisa do Censo 2010, ela pode ser comparada à perda de informações devido à censura.

Com relação ao mecanismo de perda, o pressuposto é de que os dados faltantes de escolaridade são ignoráveis (MAR). A hipótese é que a escolaridade faltante está associada

à idade, ao sexo, e a outros atributos individuais que não são observáveis, a um conjunto de características do domicílio e da área geográfica no qual ele se insere (área de ponderação, município e UF). Assim, supõe-se que não há seletividade por nível de escolaridade da observação com dado faltante, uma vez que a informação não foi recolhida para todas as unidades de análise com registro de óbito.

Conforme discutido no Capítulo 3, a adoção de mais de um método ou modelo de imputação é importante para detectar a sensibilidade da imputação a cada um. Por outro lado, um modelo adequado para a distribuição preditiva dos valores faltantes, dados os valores observados, torna a variância pequena devido ao fato de que o conhecimento dos dados observados implica no quase conhecimento dos valores faltantes (Rubin, 1978, p. 22). Para identificação da consistência das imputações foram aplicados métodos de imputação diferenciados.

A primeira IM foi realizada por *hot deck* no *Stata* 14.0. No processo, as células com dado faltante são preenchidas com dados observados, dentro do estrato definido. Neste caso, os estratos foram definidos pela área de ponderação, pelo sexo e por grupos quinquenais de idade. As unidades observadas usadas para preenchimento das observações faltantes são selecionadas estocasticamente por meio do método de aproximação Bayesiana por amostragem *bootstrap* (ABB)¹². O primeiro passo é extrair uma amostra por *bootstrap* com reposição das unidades observadas. Então, as unidades com dados faltantes são completadas com seleção aleatória e com reposição a partir da amostra *bootstrap* (Stata.Corp, 2015a). Como o método de seleção é estocástico, foram realizadas 5 repetições do método para representar a variabilidade amostral.

Com esse procedimento, supõem-se que o *missing* é do tipo MAR, de forma que todas as variáveis que são importantes na determinação da probabilidade de ausência do dado devem ser consideradas. A consideração da área de ponderação, idade e sexo é bastante restritiva, mas a inclusão de outras variáveis implicaria estratos menores, o que comprometeria a aplicação do método. De qualquer forma, este é um exercício preliminar importante para comparação com os demais métodos de imputação adotados, que consideram um conjunto grande de variáveis.

¹² O método é conhecido como ABB e foi descrito no Capítulo 3.

A segunda IM envolve a escolha de um modelo que descreva a relação da escolaridade em função de um conjunto de covariáveis X . O primeiro passo é escolher o modelo para descrever $Pr(\theta|X, Y_{obs})$, ressaltando que o Y_{obs} são os dados de escolaridade dos sobreviventes. Foram estimadas regressões logísticas ordenadas, considerando o conjunto de covariáveis X descritas no QUADRO 2 e adotando como variável dependente a escolaridade categorizada.

Após a estimação dos modelos, foram obtidos os parâmetros θ , que chamamos de θ^* . A etapa subsequente consiste na estimação de um valor para Y_{mis} a partir de sua distribuição condicional e do valor estimado para θ^* . Em seguida, o processo é repetido m vezes criando m extrações da distribuição conjunta a *posteriori* de (Y_{mis}, θ) . Foram realizadas 5 imputações ($m=5$), totalizando 5 vetores com as escolaridades imputadas. O número de imputações é importante para incorporar a variabilidade amostral e ele deve variar em função do peso do dado faltante. Como a perda é relativamente baixa (TABELA 4), é possível alcançar estimativas válidas com m pequeno¹³.

TABELA 4: Percentual de perda de informação de escolaridade por idade e sexo, estado de São Paulo, 2010

sexo	idade	obitos	vivos	total	perda
		a	b	$c=a+b$	$a/c*100$
Mulheres	25	103	161605	161708	0.064
	30	151	154697	154848	0.098
	35	157	141267	141424	0.111
	40	261	133593	133854	0.195
	45	334	125663	125997	0.265
	50	453	111626	112079	0.404
	55	522	92014	92536	0.564
Homens	25	316	159881	160197	0.197
	30	334	148731	149065	0.224
	35	399	134051	134450	0.297
	40	488	126143	126631	0.385
	45	630	116346	116976	0.539
	50	781	101970	102751	0.760
	55	902	83978	84880	1.063

Fonte: IBGE - Amostra do Censo Demográfico 2010 para São Paulo

¹³ Essa discussão foi realizada no Capítulo 3.

4.4. Aplicação o algoritmo EM para o dado faltante de escolaridade

A imputação usando algoritmo EM foi realizada usando um modelo de regressão multivariada normal (MVN) no *Stata* 14.0. O método MVN utiliza *data augmentation* (DA) com iterações via MCMC. As estimativas são obtidas por máxima verossimilhança, porém acrescenta-se um componente aleatório para criar uma amostra *bootstrap*.

Primeiro se estimam os valores esperados para os parâmetros que descrevem a relação entre as variáveis independentes e a escolaridade, partindo de uma distribuição uniforme, por meio do algoritmo EM. A distribuição uniforme assume que todos os valores dos parâmetros são igualmente prováveis. Em seguida, partindo de amostras *bootstrap* são obtidos os valores esperados para os dados faltantes. O método é iterativo e as imputações são realizadas por método de *Monte Carlo* via cadeias de *Markov* (MCMC). Quando os valores dos parâmetros convergem para uma distribuição estacionária, as observações faltantes são preenchidas. Não há uma regra de parada que garanta a distribuição estacionária. Deve-se ainda atentar para a possibilidade de dependência serial. Mas como a taxa de convergência depende da fração de dados faltantes, que neste caso é baixa, isso não deve ser um problema grande.

Dentre as vantagens, destaca-se a possibilidade de realizar imputações para subgrupos dentro da amostra (Stata.Corp, 2015b). Vale ressaltar que o método é apontado entre os mais eficazes para tratamento de dados faltantes (Schafer, 1997).

4.5. Seleção das variáveis dos modelos

A seleção das variáveis utilizadas no modelo tem um papel importante mesmo que elas não estabeleçam relação de causa e efeito com a variável imputada. Um critério importante e óbvio é a incorporação de fatores que estejam altamente correlacionados com dado faltante. Quanto mais forte essa correlação, menor será a variância dos dados faltantes preditos. O segundo critério a ser considerado é que as variáveis escolhidas estejam relacionadas à ausência do dado, para garantir MCAR (Rubin, 1978).

Como a amostra do Censo é suficientemente grande para modelos mais inclusivos, a estratégia é incluir todas as variáveis disponíveis que possam guardar relação com a

escolaridade, mas também com a mortalidade, já que essa última é seletiva e a imputação será feita exclusivamente com o grupo de falecidos. Assim, futuramente as variáveis também podem ser incluídas no modelo de análise da mortalidade. Isso é importante, uma vez que modelos utilizados para imputação mais restritivos em relação ao modelo de análise podem promover IM com estimativas inválidas, dependendo das hipóteses adotadas no modelo de análise que se mostrem irrealistas. Mas quando o contrário ocorre, IM tende a produzir estimativas válidas (Schafer, 1997). Além disso, as variáveis do modelo não devem necessariamente representar uma relação de causa e efeito, mas devem preservar importantes aspectos da distribuição conjunta das variáveis envolvidas, conforme discutido no Capítulo 3. As variáveis foram selecionadas com base neste contexto e são apresentadas no QUADRO 2.

Quanto a variável dependente, as imputações podem ser realizadas usando a escolaridade quantitativa discreta, medida em anos de estudo e variando de 0 a 21, ou categorizada, de acordo com as especificidades dos modelos adotados e qualidade proporcionada por cada tipo. Dada a forma como o quesito de escolaridade foi abordado na amostra do Censo, há perda de informação ao transformar a variável para o tipo quantitativa discreta. No modelo de imputação por MNV, o ideal é trabalhar com anos de escolaridade, dada a suposição de normalidade da variável dependente. Para o modelo de imputação múltipla (IM) foram testadas duas categorizações da escolaridade, tendo por base as categorias do MS/SIM. Uma categorização, considera 4 grupos de escolaridade: 3 ou menos, 4 a 7, 8 a 11, 12 ou mais anos de escolaridade. Um segundo agrupamento restringe a escolaridade a 3 categorias: 7 ou menos, 8 a 11, 12 anos ou mais de escolaridade.

QUADRO 2: Variáveis selecionadas para as regressões

Nível/classificação	Nome da variável	Descrição	Medida/agrupamento
Dependente	educsim2	escolaridade agrupada com base nas categorias disponíveis no SIM (MS)	1: <=7 anos 2: >=8 e <=11 3: >=12
	educsim3	escolaridade agrupada com base nas categorias disponíveis no SIM (MS)	1: <=3 2: >=4 e <=7 3: >=8 e <=11 4: >=12
	educsim4	escolaridade do responsável pelo domicílio	1: nenhum 2: >=1 e <=7 3: >=8 & <=11

Nível/classificação	Nome da variável	Descrição	Medida/agrupamento
			4: >= 12
	educ	escolaridade	anos de estudo
	educimp	escolaridade imputada	educ, educsim2, educsim3
Indivíduo	idgrup2	idade agrupada	1: 25-29 anos 2: 30-34 anos 3: 35-39 anos 4: 40-44 anos 5: 45-49 anos 6: 50-54 anos 7: 55-59 anos
	sexo	sexo	1 se homem, 0 se mulher
Características do responsável pelo domicílio	ch_sexo	sexo do responsável pelo domicílio	1 se homem, 0 se mulher
	ch_educsim2	escolaridade do responsável pelo domicílio	1: <=7 anos 2: >=8 e <=11 3: >=12
	ch_educsim3	escolaridade do responsável pelo domicílio	1: <=3 2: >=4 e <=7 3: >=8 e <=11 4: >=12
	ch_educsim4	escolaridade do responsável pelo domicílio	1: nenhum\ 2: >=1 e <=7 3: >=8 & <=11 4: >= 12
	ch_educ	escolaridade do responsável pelo domicílio	anos de estudo
	ch_idgrup5	idade do responsável pelo domicílio	1: <30 anos 2: >=30 e <39 3: >=40 e <49 4: >=50 e <=59 5: >=60
	ch_idgrup4	idade do responsável pelo domicílio	1: <25 anos 2: 25-29 anos 3: 30-34 anos 4: 35-39 anos 5: 40-44 anos 6: 45-49 anos 7: 50-54 anos 8: 55-59 anos 9: >= 60 anos
	ch_conjuge	estado civil do responsável pelo domicílio	1 se tem cônjuge, 0 se não
	ch_fisico	responsável pelo domicílio tem pelo menos uma deficiência total ou grande?	1 se sim, 0 se não
		ch_cor	raça/cor do responsável pelo domicílio

Nível/classificação	Nome da variável	Descrição	Medida/agrupamento
			3: amarela 4: parda 5: indígena
	iddif	diferença entre a idade do responsável do domicílio e a do indivíduo (adulto entre 25 e 59 anos)	idade simples
	iddif2	diferença entre grupo etário do responsável do domicílio e do indivíduo (adulto entre 25 e 59 anos)	cada unidade de diferença representa diferença entre 0 e 4 anos de idade
	iddif3	diferença entre grupo etário do responsável do domicílio e do indivíduo (adulto entre 25 e 59 anos)	1: diferença > - 5 & < 5 anos 2: chefe mais jovem 5 ou mais anos 3: chefe mais velho 5 ou mais anos
	iddif4	diferença entre grupo etário do responsável do domicílio e do indivíduo (adulto entre 25 e 59 anos)	1: diferença > - 10 & < 10 anos 2: responsável mais jovem entre 10 e 19 anos 3: responsável 20 anos ou mais jovem que o indivíduo 4: responsável mais velho entre 10 e 19 anos 5: responsável 20 anos ou mais velho que o indivíduo
Domicílio	ses	<p>indicador de condição socioeconômica do domicílio composto pelas variáveis:</p> <ul style="list-style-type: none"> * existência de tv no domicílio * existência de carro no domicílio * existência de moto no domicílio * existência de máquina de lavar no domicílio * existência de geladeira no domicílio * existência de água canalizada em pelo menos 1 cômodo * existência de computador com internet no domicílio * existência de celular no domicílio * material permanente predominante (alvenaria) * se moradia é adequada (ter abastecimento de água por rede geral, esgotamento sanitário por rede geral ou fossa séptica, coleta de lixo direta ou indireta e, no máximo, dois moradores por dormitório) * presença de pelo menos 1 morador do domicílio é beneficiário de programa de transferência de renda (bolsa família; PETI ou outro) 	4 categorias - quartis
	rd_d	razão dependência (≤ 14 ou ≥ 60)/(> 15 e ≤ 59)	1 se > 1 , 0 se ≤ 1
	morteinf_d	ocorrência de óbito infantil no domicílio	1 se sim, 0 se não

Nível/classificação	Nome da variável	Descrição	Medida/agrupamento
	brasilnato_d	domicílios compostos exclusivamente por estrangeiros ou naturalizados ou tem brasileiro nato?	1 se tem algum brasileiro nato; 0 se só tem estrangeiros ou naturalizados
Comunidade	perc_subnor2	proporção de pessoas vivendo em áreas subnormais dentro da área de ponderação	0: nenhum; 1: >0 e <50; 2: ≥50 e < 90; 3: ≥90
	rm	região metropolitana	1 se sim, 0 se não
	sitres	situação do domicílio	1 se urbana, 0 se rural
	sample_educ	escolaridade (educ) de indivíduo selecionado aleatoriamente (amostragem <i>bootstrap</i>)	
	sample_educsim2	escolaridade (educsim2) de indivíduo selecionado aleatoriamente (amostragem <i>bootstrap</i>)	1: ≤7 anos 2: ≥8 e ≤11 3: ≥12
	sample_educsim3	escolaridade (educsim3) de indivíduo selecionado aleatoriamente (amostragem <i>bootstrap</i>)	1: ≤3 2: ≥4 e ≤7 3: ≥8 & ≤11 4: ≥12

O indicador de nível socioeconômico SES foi criado usando Análise de Componentes Principais (ACP). Como as variáveis utilizadas são binárias, as componentes foram estimadas com base em matriz de correlação tetracórica (QUADRO 3), que mede correlação linear de uma distribuição normal bivariada. Embora o coeficiente de correlação tetracórico tenha uma variabilidade maior que o do *Pearson*, a amostra do Censo é grande, o que deve compensar esse problema. A partir da primeira componente foi criada a variável “ses” composta por quatro categorias definidas pelos quartis.

QUADRO 3: Matriz de correlações tetracóricas das variáveis componentes do indicador socioeconômico para domicílio

	adequada	material	canal	auto	banheiro	moto	maquina	tv	geladeira	celular	internet	social_d
adequada	1											
material	0.300	1										
canal	0.274	0.268	1									
auto	0.304	0.271	0.090	1								
banheiro	0.394	0.534	0.466	0.386	1							
moto	0.058	0.056	0.037	0.206	0.154	1						
maquina	0.301	0.347	0.076	0.483	0.455	0.124	1					
tv	0.138	0.250	0.200	0.210	0.437	0.118	0.316	1				
geladeira	0.239	0.366	0.298	0.440	0.596	0.201	0.562	0.600	1			
celular	0.124	0.206	0.108	0.395	0.328	0.231	0.335	0.313	0.432	1		
internet	0.397	0.330	0.110	0.538	0.419	0.143	0.601	0.306	0.474	0.468	1	
social_d	-0.312	-0.219	-0.064	-0.305	-0.185	-0.087	-0.300	-0.062	-0.147	-0.074	-0.272	1

4.6. Validade dos dados imputados

Para avaliar a validade dos dados imputados, foi realizado um exercício de simulação de dados faltantes. O primeiro passo é extrair uma subamostra aleatória de tamanho n^* da amostra de sobreviventes de tamanho n do Censo, dentre os indivíduos vivos. O segundo passo é tornar a informação de escolaridade artificialmente faltante para as n^* unidades da subamostra. O terceiro passo é imputar informação de escolaridade para estas n^* observações, com base nos dados restantes, compostos por $n-n^*$ unidades. Por fim, a escolaridade imputada pode ser comparada à observada para avaliar a qualidade da imputação.

Os critérios de avaliação foram baseados em parte do trabalho de Collins *et al* (2001). Em uma simulação de dados faltantes, eles elegem quatro medidas de avaliação para comparação de métodos de imputação. A primeira medida utilizada é o viés padronizado (*VP*), que é uma medida de viés bruto. O *VP* é dado pela diferença do valor médio de uma estimativa com dados imputações e a obtida com os dados observados medida em percentual do erro padrão. Para isso, a referida diferença é dividida pelo erro padrão das estimativas com dados imputados. Os valores de *VPs* que excedem 40-50% do erro padrão (em qualquer direção) indicam que as imputações têm impactos adversos sobre a eficiência, a cobertura e as taxas de erro das estimativas. Entretanto, na simulação dos autores a amostra é vertiginosamente menor ($n=500$) e os percentuais de *missing* muito elevados (25% e 50%), o que deve produzir erros padrão muito maiores que os encontrados neste trabalho.

As demais medidas utilizadas dizem respeito aos intervalos de confiança produzidos. Os autores sugerem comparar a cobertura e a extensão dos intervalos de confiança, sendo os intervalos preferíveis os menores intervalos.

Outra medida proposta, mas que não foi analisada, é a raiz do erro quadrático médio (*REQM*). O erro quadrático médio é uma medida de precisão que combina viés e eficiência. Considera-se para a análise a sua raiz para que sua medida fique na mesma escala do parâmetro.

4.7. Modelo analítico para mortalidade

Dada a disponibilidade dos microdados do censo, foram adotados modelos estatísticos para estimar as taxas de mortalidade por educação, após realizadas as imputações de educação para os indivíduos reportados como falecidos no censo. O modelo mais adequado é o de Poisson (Coleman, 1964) que permite estimar as mortes em função do risco de morte e do tempo de exposição. O modelo assume que existe uma quantidade de taxa de incidência na qual o evento ocorre. Essa taxa de incidência pode ser multiplicada por um tempo de exposição para obter o número de eventos ocorridos, sendo as exposições não sobrepostas independentes entre si.

O modelo de regressão para as taxas de mortalidade é dado por:

$$\log (\text{morte}) = \beta \cdot X_i + \log [\text{exposição (pessoas ano)}]$$

em que a variável dependente é o logaritmo da taxa de mortalidade. Os coeficientes são estimados por máxima verossimilhança e as variáveis independentes são idade, sexo e escolaridade.

Para a estimação do modelo de regressão, além do número de óbitos, é preciso definir o tempo de exposição. No caso dos microdados censitários, cada sobrevivente ao final do período de referência contribui com um ano-pessoa. Já os indivíduos mortos contribuem com a fração do ano em que sobreviveram, com base no mês em que o óbito ocorreu. Como não há informação do dia dos óbitos, foi considerado o meio do mês como data de referência, pressupondo que os óbitos estejam linearmente distribuídos dentro do mês de ocorrência. Portanto, as estimativas do modelo de Poisson calculadas apenas com os dados imputados do censo demográfico correspondem às taxas médias de mortalidade dos 12 meses anteriores à data de referência do censo demográfico.

Para fins de comparação, as taxas de mortalidade obtidas com o Censo foram calculadas usando as escolaridades imputadas para os mortos e usando a escolaridade do responsável pelo domicílio. O objetivo é avaliar se há diferenciais no gradiente educacional da mortalidade quando usamos a escolaridade de um representante do domicílio e usando a escolaridade do indivíduo morto, mesmo que imputada.

No caso das taxas de mortalidade estimadas com dados do SIM, não foram utilizados microdados, mas dados agregados por educação, idade e sexo. O número de óbitos corresponde aos eventos ocorridos entre 1 de julho de 2012 e 31 de agosto de 2013. Esse período foi selecionado para corresponder aos mesmos meses do período de referência do Censo. O cálculo do tempo de exposição seguiu uma lógica similar à estratégia para os microdados do censo. Os dados do SIM incluem a data completa do óbito, o que permite o cálculo exato do tempo de contribuição de cada indivíduo falecido no período de referência. Como o SIM só tem informação sobre óbitos, foi necessário estimar a contribuição dos vivos com base em dados do Censo. O primeiro passo foi calcular as taxas de crescimento observadas entre agosto de 2010 e julho de 2012, sendo a população de 2012 aquela projetada por idade e sexo pelo IBGE. No segundo passo, tomou-se a população estimada pelo IBGE para 1º de julho de 2013 para o Tribunal de Contas da União, que foi projetada para 31 de julho de 2013, com base nas referidas taxas de crescimento. Por fim, foi aplicada a mesma estrutura por idade, sexo e escolaridade observada no Censo de 2010 na população de 31 de julho de 2013, considerando que três anos é um tempo pequeno o suficiente para que não tenha havido mudanças expressivas nesta estrutura.

4.8. Correção do nível dos óbitos no Censo

Após apresentação das medidas de mortalidade, foi selecionada aquela que se desponta como mais adequada na análise de consistência dos resultados. Elas foram comparadas com aquelas obtidas no SIM. Os possíveis erros de sub ou sobre enumeração dos óbitos no Censo foram então corrigidos usando o nível de óbitos do SIM. Queiroz (2012) destaca que os registros do SIM para São Paulo cobrem praticamente 100% dos óbitos.

A correção foi realizada por meio de um índice de correção para cada grupo de idade, sexo e escolaridade. O índice é o quociente entre as taxas de mortalidade obtidas com SIM e as censitárias. Tais índices foram aplicados a cada unidade com óbitos nos microdados do censo, de modo que os coeficientes do Modelo de Poisson e as taxas de mortalidade pudessem ser reestimadas.

5. RESULTADOS DAS IMPUTAÇÕES DA ESCOLARIDADE

Este capítulo apresenta os principais resultados obtidos com os métodos de imputação adotados. Todos os métodos foram aplicados nos dados da amostra do estado de São Paulo, que é maior que a do município, para garantir melhores resultados nas imputações.

É importante ressaltar que as imputações devem produzir estimativas válidas, sendo mais importante a distribuição assumida para os dados faltantes, comparativamente à produção de um valor particular para o cada um. O que se pretendeu é utilizar a imputação para criar uma distribuição de escolaridade para um grupo particular da amostra, onde todos os indivíduos com dados faltantes experimentam um evento em comum - o óbito.

A seção 5.1 apresenta a análise de consistência das imputações por *hotdeck*. Embora esse não seja um método mais adequado para o caso, seus resultados ajudam elucidar os ganhos observados quando se consideram modelos mais apropriados para representar a distribuição da escolaridade dos indivíduos mortos. Tais modelos seriam a imputação múltipla e a imputação por modelo multivariado normal (MVN), cujas análises de consistências são apresentadas, respectivamente, nas seções 5.2 e 5.3. Na seção 5.4 encontram-se as distribuições da escolaridade imputada por IM e MVN e do responsável pelo domicílio. Na seção 5.5 as distribuições analisadas em 5.4 são comparadas com aquelas obtidas com dados do SIM. Em 5.6 são apresentadas algumas considerações finais sobre as imputações para destacar algumas decisões tomadas em relação a análise dos modelos de mortalidade presentes no Capítulo 6.

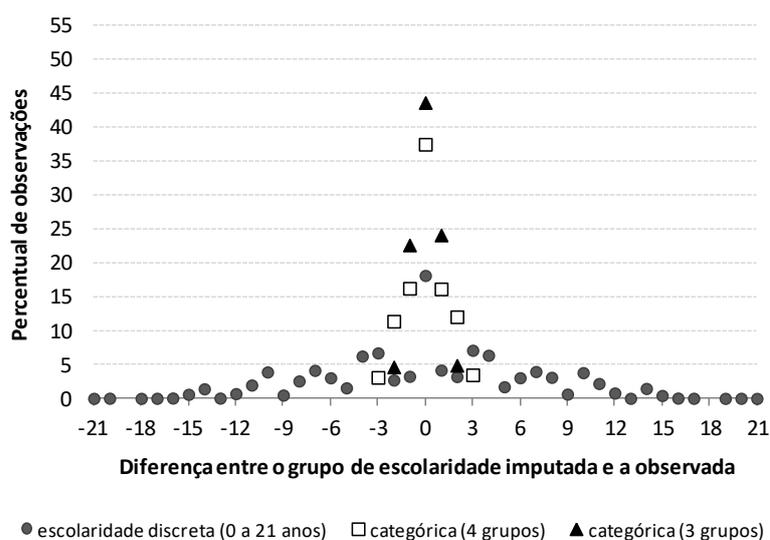
5.1. Imputação por *hotdeck*: consistência dos resultados

Nas imputações por *hotdeck*, os estratos foram definidos por área de ponderação, sexo e idade. Como o número de variáveis de controle é pequeno, o método produz estimativas

coincidentes com a informação observada, em média, em 18% dos casos¹⁴, considerando o número de imputações $m = 5$.

Por outro lado, essa estimativa tende a aumentar quando a escolaridade é agrupada, sendo a relação inversa ao número de categorias utilizado (FIGURA 1). O comportamento das distâncias entre o dado observado e o estimado segue uma distribuição de frequência normal. As distâncias mais elevadas têm frequências substancialmente menores, comparativamente à moda da distribuição, que é zero. Quando se utilizam 4 categorias de escolaridade, em média, 37% dos valores imputados¹⁵ coincidem com o valor observado. Esse percentual alcança uma média de 43%¹⁶ quando se agrupam os anos de estudo em 3 categorias.

FIGURA 1: Proporção de observações segundo a diferença entre escolaridade observada e imputada, *hotdeck* com 5 repetições, subamostra de 10000 pessoas, estado de São Paulo, 2010



Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo

Embora as estimativas obtidas com escolaridade com maior número de categorias sejam superiores em termos de precisão, há a desvantagem em aumentar a concentração da amostra nos estratos de menor escolaridade. Mais da metade da população adulta (25 a 59

¹⁴ Combinação das 5 imputações, com erro padrão de 0,3.

¹⁵ Erro padrão = 0,48.

¹⁶ Erro padrão = 0,49.

anos) do estado e do de São Paulo tem no máximo ensino fundamental completo; para a capital esse grupo representa de 47%. Ademais, as imputações não visam produzir um valor particular para o dado faltante. Este é um aspecto não controlável, uma vez que não há como saber se o dado imputado coincide de fato com a escolaridade do indivíduo falecido. A questão está na validade distribuição da escolaridade para este grupo de indivíduos, minimizando possíveis vieses.

Neste aspecto, as imputações procedidas pelo *hotdeck* não parecem causar problemas graves. Com base na subamostra utilizada como teste e na escolaridade discreta, as médias se mantêm estatisticamente iguais para os dados imputados e os observados (TABELA 5). Para as imputações com escolaridade categórica, algumas proporções são estatisticamente diferentes, mas as precisões das estimativas são elevadas com intervalos de confiança pequenos. Apesar destas diferenças, as distribuições da escolaridade imputadas e observadas seguem padrões muito semelhantes (TABELA 6 e TABELA 7). Por outro lado, quando analisados os vieses padronizados (VP), alguns resultados superam o limite de 0,4 ou 0,5 (para ambas as direções) sugerido por Collins et al (2001).

TABELA 5: Escolaridade média observada e imputada, *hotdeck* com escolaridade discreta, subamostra de 10000 pessoas, estado de São Paulo, 2010

Origem da estimativa	Escolaridade		
	Média	Erro padrão	Intervalo de confiança 95%
Imputada	7.85	0.0495	7.7565 7.9507
Observada	7.77	0.0480	7.6724 7.8606
VP	1.76		
Extensão do intervalo	0.19		

Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo

TABELA 6: Distribuição dos indivíduos segundo escolaridade imputada e observada, *hotdeck* com 3 categorias de escolaridade, subamostra de 10000 pessoas, estado de São Paulo, 2010

Anos de estudo	Escolaridade imputada			Escolaridade observada				Qualidade da imputação	
	Distribuição %	Erro padrão	Intervalo de confiança	Distribuição %	Erro padrão	Intervalo de confiança 95%	VP	Extensão do intervalo	
<= 7	37.40	0.49	36.44 38.35	39.41	0.47	38.49 40.33	-4.14	1.91	
>=8 e <=11	43.62	0.50	42.64 44.60	41.54	0.47	40.61 42.47	4.16	1.96	
>=12	18.99	0.39	18.21 19.76	19.05	0.38	18.31 19.79	-0.15	1.55	
Total	100.00			100.00					

Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo

TABELA 7: Distribuição dos indivíduos segundo escolaridade imputada e observada, *hotdeck* com 4 categorias de escolaridade, subamostra de 10000 pessoas, estado de São Paulo, 2010

Anos de estudo	Escolaridade imputada			Escolaridade observada				Qualidade da imputação	
	Distribuição %	Erro padrão	Intervalo de confiança	Distribuição %	Erro padrão	Intervalo de confiança 95%	VP	Extensão do intervalo	
<=3	26.22	0.44	25.35 27.09	27.46	0.43	26.62 28.30	-2.79	1.74	
>=4 e <=7	12.16	0.33	11.51 12.80	11.96	0.31	11.35 12.57	0.62	1.29	
>=8 e <=11	42.49	0.50	41.52 43.47	41.54	0.47	40.61 42.47	1.91	1.95	
>=12	19.13	0.40	18.35 19.90	19.05	0.38	18.31 19.79	0.20	1.55	
Total	100.00			100.00					

Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo

Estes resultados são esperados, uma vez que o método envolve extração aleatória da escolaridade, levando em consideração a distribuição observada por área de ponderação, sexo e idade. Como os doadores de informação são os vivos, prevalece a distribuição de escolaridade observada entre eles. Isso torna o *hotdeck* um método não atrativo para essa análise. Sua apresentação serve apenas de base para comparação com os métodos da IM e MVN, que por hipótese são mais adequados e produzem resultados superiores.

5.2. Imputação Múltipla com regressão logística ordenada: consistência dos resultados

Para a imputação múltipla por regressão, foi adotado o modelo logístico para dados ordenados, com a variável dependente escolaridade categorizada. Nesta seção são

apresentados os resultados usando 3 e 4 categorias de escolaridade. Existem três grupos de modelos estimados: o primeiro grupo, modelos IM1 a IM3 (TABELA 8) que incluem variáveis do indivíduo, localidade e escolaridade de outros indivíduos; o segundo grupo – modelos IM4 a IM6, inclui variáveis do responsável pelo domicílio e usam escolaridade em 3 categorias (TABELA 9); e o terceiro grupo, modelos IM7 a IM9, que usam as mesmas variáveis dos modelos IM4 a IM6, porém com escolaridade em 4 categorias (TABELA 10).

O primeiro modelo estimado (IM1) levou em consideração aspectos do nível do indivíduo, características do domicílio e da localidade, representados pelas variáveis expostas no QUADRO 2. O objetivo foi gerar um modelo base para comparar com os demais modelos que incluem escolaridade do vizinho e/ou do responsável e outras características do responsável pelo domicílio.

No IM1, todas as variáveis são significativas, exceto a ocorrência de mortalidade infantil no domicílio. O modelo IM2 inclui a escolaridade do vizinho, que se apresenta significativamente associada à escolaridade do indivíduo. No entanto, a grande diferença é observada quando se adiciona a escolaridade do responsável do domicílio no modelo IM3. As razões de chance são mais expressivas em relação à escolaridade do vizinho e o *pseudo R²* cresce consideravelmente.

Nos segundo e terceiro grupos de modelos, as demais características do responsável pelo domicílio são significativas. Responsáveis do sexo feminino, de idades mais avançadas, mais jovens que os indivíduos não chefes, brancos (em relação aos pretos, pardos e indígenas) representam vantagens para a educação dos indivíduos. Entretanto, o incremento no *pseudo R²* após inclusão destas variáveis não é expressivo, se comparado ao observado com a introdução da escolaridade do chefe.

TABELA 8: Modelos de regressão logística ordenada para imputação, 3 categorias de escolaridade, IM1 a IM3

Variáveis/categorias	Modelos					
	IM1		IM2		IM3	
	OR	Erro padrão	OR	Erro padrão	OR	Erro padrão
1.sexo	0.8530	0.0042	0.8751	0.0043	0.9117	0.0046
idgrup2						
30 a 34	0.6998	0.0046	0.7231	0.0048	0.6492	0.0046
35 a 39	0.4804	0.0035	0.5173	0.0038	0.4381	0.0034
40 a 45	0.3579	0.0028	0.3948	0.0032	0.3298	0.0027
45 a 49	0.2950	0.0025	0.3275	0.0029	0.2711	0.0024
50 a 54	0.2261	0.0022	0.2588	0.0025	0.2083	0.0020
55 a 59	0.1687	0.0019	0.1958	0.0022	0.1559	0.0017
educsampl2						
>=8 e <=11			1.2792	0.0070		
>=12			2.0392	0.0142		
ch_educsim2						
>=8 e <=11					2.5811	0.0141
>=12					14.6852	0.1337
ses						
2ºQ	2.0475	0.0130	2.0296	0.0131	1.8309	0.0123
3ºQ	7.4644	0.0451	7.1369	0.0436	4.4523	0.0289
4ºQ	9.9758	0.1013	9.4900	0.0969	5.3842	0.0579
1.morte_d	0.8877	0.0170	0.8825	0.0170	1.0294	0.0211
1.rd_d	1.0395	0.0091	1.0301	0.0090	1.2097	0.0114
1.morteinf_d	0.8296	0.0988	0.8452	0.1006	0.7022	0.0886 *
1.brasilnato_d	0.5546	0.0511	0.6000	0.0549	0.8263	0.0701
perc_subnor2						
>0 e <50	0.6684	0.0040	0.7076	0.0042	0.7856	0.0048
≥50 e < 90	0.3945	0.0088	0.4578	0.0102	0.5244	0.0120
≥90	0.3353	0.0126	0.3722	0.0143	0.4581	0.0175
1.sitres	1.6591	0.0186	1.6591	0.0186	1.5052	0.0167
1.rm	1.5365	0.0087	1.5365	0.0087	1.3623	0.0079
/cut1	-0.3394	0.0928	-0.0123	0.0924	0.2195	0.0856
/cut2	2.0817	0.0928	2.4417	0.0924	3.0033	0.0857
Pseudo R ²	0.1296		0.1371		0.2097	
n	979946		974817		979946	

**P-valor ≥0.14 para todos os modelos

Demais variáveis: p-valor ≤ 0.001

Fonte: Estimativas com base em IBGE - Censo Demográfico 2010, amostra estado de São Paulo

TABELA 9: Modelos de regressão logística ordenada para imputação, 3 categorias de escolaridade, IM4 a IM6

Variáveis/categorias	Modelos						P-valor
	IM4		IM5		IM6		
	OR	Erro padrão	OR	Erro padrão	OR	Erro padrão	
l.sexo	0.7070	0.0046	0.7222	0.0047	0.7224	0.0047	0.0000
idgrup2							
30 a 34	0.6261	0.0062	0.6375	0.0063	0.6402	0.0063	0.0000
35 a 39	0.4231	0.0042	0.4410	0.0044	0.4441	0.0044	0.0000
40 a 45	0.2624	0.0040	0.2780	0.0042	0.2822	0.0043	0.0000
45 a 49	0.2070	0.0030	0.2201	0.0032	0.2233	0.0033	0.0000
50 a 54	0.1239	0.0026	0.1352	0.0029	0.1408	0.0030	0.0000
55 a 59	0.0866	0.0019	0.0967	0.0021	0.1000	0.0021	0.0000
educsample2							
>=8 e <=11			1.2026	0.0068	1.1966	0.0068	0.0000
>=12			1.5632	0.0112	1.5632	0.0112	0.0000
ch_sexo							
feminino	1.2426	0.0085	1.2322	0.0084	1.2135	0.0082	0.0000
ch_educsim2							
>=8 e <=11	3.5317	0.0210	2.8970	0.0535	1.1966	0.0068	0.0000
>=12	20.8247	0.2040	15.4993	0.3860	1.5632	0.0112	0.0000
ch_idgrup5							
30 a 39	1.3870	0.0162			1.3988	0.0165	0.0000
<40 a 49	1.9948	0.0324			2.0097	0.0327	0.0000
<50 a 59	3.0562	0.0670			3.0674	0.0670	0.0000
>= 60	5.1051	0.1468			4.7744	0.1359	0.0000
iddif4							
mais jovem entre 10 e 19 anos	1.2978	0.0153	1.2979	0.0154			0.0000
20 anos mais jovem ou mais	1.5588	0.0465	1.5736	0.0465			0.0000
mais velho entre 10 e 19 anos	0.7161	0.0075	0.7228	0.0076			0.0000
20 anos mais velho ou mais	1.0294	0.0221	1.0203	0.0219			≥0.1780
ch_educsim2#ch_idgrup5							
<=7#30 a 39			1.1051	0.0233			0.0000
<=7#<40 a 49			1.5740	0.0367			0.0000
<=7#<50 a 59			2.5231	0.0688			0.0000
<=7#>= 60			4.5561	0.1489			0.0000
>=8 e <=11#30 a 39			1.4487	0.0188			0.0000
>=8 e <=11#<40 a 49			2.0918	0.0364			0.0000
>=8 e <=11#<50 a 59			3.0819	0.0707			0.0000
>=8 e <=11#>= 60			4.9002	0.1498			0.0000
>=12#30 a 39			1.5640	0.0358			0.0000
>=12#<40 a 49			2.3265	0.0607			0.0000
>=12#<50 a 59			3.6430	0.1116			0.0000
>=12#>= 60			3.7139	0.1490			0.0000
ch_educsim2#iddif4							
<=7#mais jovem entre 10 e 19 anos					1.7470	0.0322	0.0000
<=7#20 anos mais jovem ou mais					4.1859	0.2062	0.0000
<=7#mais velho entre 10 e 19 anos					1.0288	0.0141	0.0380
<=7#20 anos mais velho ou mais					1.6419	0.0375	0.0000
>=8 e <=11#mais jovem entre 10 e 19 anos					1.2336	0.0170	0.0000
>=8 e <=11#20 anos mais jovem ou mais					1.3227	0.0440	0.0000
>=8 e <=11#mais velho entre 10 e 19 anos					0.6736	0.0079	0.0000
>=8 e <=11#20 anos mais velho ou mais					0.8516	0.0200	0.0000
>=12#mais jovem entre 10 e 19 anos					0.9752	0.0227	0.2800
>=12#20 anos mais jovem ou mais					0.8149	0.0481	0.0010
>=12#mais velho entre 10 e 19 anos					0.4762	0.0084	0.0000
>=12#20 anos mais velho ou mais					0.4571	0.0138	0.0000
l.ch_conjuge	1.1673	0.0065	1.1746	0.0066	1.1656	0.0066	0.0000
ch_fisico							
pelo menos 1 deficiência total ou grande	0.8917	0.0090	0.8879	0.0090	0.8770	0.0090	0.0000
ch_cor2							
preta	0.7365	0.0073	0.7441	0.0074	0.7405	0.0074	0.0000
amarela	1.4419	0.0320	1.4252	0.0316	1.4309	0.0315	0.0000
parda	0.6882	0.0038	0.6974	0.0039	0.6951	0.0039	0.0000
indígena	0.6867	0.0492	0.6931	0.0497	0.6935	0.0499	0.0000
ses							
2ºQ	1.6615	0.0115	1.6601	0.0116	1.6607	0.0117	0.0000
3ºQ	3.5663	0.0240	3.5158	0.0239	3.5334	0.0242	0.0000
4ºQ	4.2862	0.0473	4.1965	0.0465	4.2199	0.0467	0.0000
l.morte_d	0.8973	0.0186	0.8958	0.0186	0.8935	0.0186	0.0000
l.rd_d	0.9135	0.0089	0.9092	0.0088	0.9227	0.0090	0.0000
l.brasilnato_d	0.8225	0.0700	0.8660	0.0737	0.8594	0.0722	0.02/ 0.07/0.05
perc_subnor2							
>0 e <50	0.8350	0.0052	0.8575	0.0053	0.8563	0.0053	0.0000
≥50 e < 90	0.6374	0.0149	0.6917	0.0162	0.6922	0.0164	0.0000
≥90	0.5921	0.0228	0.6262	0.0246	0.6284	0.0249	0.0000
l.sitres	1.4843	0.0169	1.4407	0.0165	1.4349	0.0166	0.0000
l.rm	1.3602	0.0080	1.3164	0.0077	1.3206	0.0078	0.0000
/cut1	0.7362	0.0863	0.7917	0.0876	1.2375	0.0856	0.0000
/cut2	3.6568	0.0864	3.7296	0.0877	4.1853	0.0857	0.0000
Pseudo R ²	0.2370		0.2400		0.2424		
n	979737		974612		974612		

Fonte: Estimativas com base em IBGE - Censo Demográfico 2010, amostra estado de São Paulo

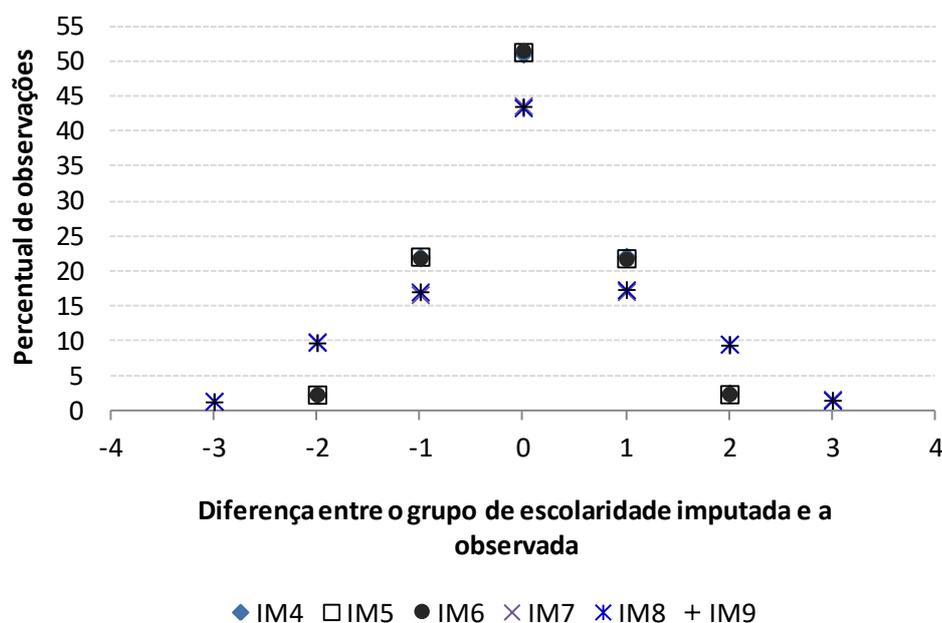
TABELA 10: Modelos de regressão logística ordenada para imputação, 4 categorias de escolaridade, IM7 a IM9

Variáveis/categorias	Modelos						
	IM7		IM8		IM9		P-valor
	OR	Erro padrão	OR	Erro padrão	OR	Erro padrão	
l.sexo	0.7202	0.0045	0.7340	0.0046	0.7347	0.0046	
idgrup2							
30 a 34	0.6214	0.0060	0.6334	0.0061	0.6342	0.0061	0.000
35 a 39	0.4186	0.0040	0.4332	0.0042	0.4348	0.0042	0.000
40 a 45	0.2601	0.0038	0.2735	0.0040	0.2762	0.0040	0.000
45 a 49	0.2121	0.0030	0.2240	0.0031	0.2259	0.0032	0.000
50 a 54	0.1309	0.0026	0.1406	0.0028	0.1449	0.0029	0.000
55 a 59	0.0937	0.0019	0.1043	0.0021	0.1073	0.0022	0.000
educsample3							
>=4 e <=7			1.1016	0.0090	1.0993	0.0091	0.000
>=8 e <=11			1.2157	0.0073	1.2093	0.0073	0.000
>=12			1.5527	0.0116	1.5531	0.0116	0.000
ch_sexo							
feminino	1.2297	0.0081	1.2211	0.0081	1.2042	0.0079	0.000
ch_educlim3							
>=4 e <=7	1.9881	0.0139	2.1038	0.0609	2.4807	0.0302	0.000
>=8 e <=11	4.3032	0.0285	3.9762	0.0976	6.1691	0.0675	0.000
>=12	26.0963	0.2707	21.9621	0.6553	43.5429	0.6313	0.000
ch_idgrup5							
30 a 39	1.4299	0.0159			1.4484	0.0163	0.000
<40 a 49	2.0407	0.0316			2.0743	0.0322	0.000
<50 a 59	3.0721	0.0642			3.1247	0.0654	0.000
>= 60	5.1257	0.1413			4.8789	0.1338	0.000
iddif4							
mais jovem entre 10 e 19 anos	1.2817	0.0143	1.2864	0.0144			0.000
20 anos mais jovem ou mais	1.4737	0.0409	1.4887	0.0413			0.000
mais velho entre 10 e 19 anos	0.7157	0.0071	0.7189	0.0071			0.000
20 anos mais velho ou mais	1.0310	0.0211	1.0203	0.0210			>=0.137
ch_educlim3#ch_idgrup5							
<=3#30 a 39			1.3024	0.0354			0.000
<=3#<40 a 49			1.8444	0.0523			0.000
<=3#<50 a 59			2.8812	0.0900			0.000
<=3#>= 60			5.0220	0.1796			0.000
>=4 e <=7#30 a 39			1.2273	0.0275			0.000
>=4 e <=7#<40 a 49			1.6862	0.0417			0.000
>=4 e <=7#<50 a 59			2.6133	0.0738			0.000
>=4 e <=7#>= 60			4.9405	0.1685			0.000
>=8 e <=11#30 a 39			1.4653	0.0185			0.000
>=8 e <=11#<40 a 49			2.1167	0.0356			0.000
>=8 e <=11#<50 a 59			3.0788	0.0681			0.000
>=8 e <=11#>= 60			4.8993	0.1443			0.000
>=12#30 a 39			1.5918	0.0364			0.000
>=12#<40 a 49			2.3601	0.0609			0.000
>=12#<50 a 59			3.6031	0.1083			0.000
>=12#>= 60			3.6717	0.1439			0.000
ch_educlim3#iddif4							
<=3#mais jovem entre 10 e 19 anos					1.8037		0.000
<=3#20 anos mais jovem ou mais					3.9490		0.000
<=3#mais velho entre 10 e 19 anos					1.0522		0.001
<=3#20 anos mais velho ou mais					1.6916		0.000
>=4 e <=7#mais jovem entre 10 e 19 anos					1.4529		0.000
>=4 e <=7#20 anos mais jovem ou mais					2.5421		0.000
>=4 e <=7#mais velho entre 10 e 19 anos					0.7985		0.000
>=4 e <=7#20 anos mais velho ou mais					1.2015		0.000
>=8 e <=11#mais jovem entre 10 e 19 anos					1.2098		0.000
>=8 e <=11#20 anos mais jovem ou mais					1.2882		0.000
>=8 e <=11#mais velho entre 10 e 19 anos					0.6719		0.000
>=8 e <=11#20 anos mais velho ou mais					0.8640		0.000
>=12#mais jovem entre 10 e 19 anos					0.9642		0.117
>=12#20 anos mais jovem ou mais					0.7730		0.000
>=12#mais velho entre 10 e 19 anos					0.4770		0.000
>=12#20 anos mais velho ou mais					0.4592		0.000
l.ch_conjuge	1.1504		1.1569	0.0062	1.1489		0.000
ch_fisico							
1 deficiência total ou grande	0.9103		0.9080	0.0087	0.8969		0.000
ch_cor2							
preta	0.7559		0.7629	0.0072	0.7595		0.000
amarela	1.4252		1.4151	0.0313	1.4222		0.000
parda	0.7142		0.7231	0.0038	0.7211		0.000
indígena	0.7164		0.7288	0.0479	0.7285		0.000
ses							
2°Q	1.5724		1.5717	0.0101	1.5726		0.000
3°Q	3.2402		3.1985	0.0206	3.2114		0.000
4°Q	3.9011		3.8275	0.0417	3.8435		0.000
l.morte_d	0.9095		0.9071	0.0178	0.9044		0.000
l.r_d	0.9259		0.9244	0.0084	0.9364		0.000
l.brasilnato_d	0.8635		0.8979	0.0769	0.8989		0.086 / >=0.207
pere_subnor2							
>0 e <=0	0.8442		0.8662	0.0052	0.8651		0.000
>=0 e <=90	0.6694		0.7203	0.0156	0.7211		0.000
>=90	0.6153		0.6715	0.0241	0.6753		0.000
l.sitres	1.4410		1.3948	0.0143	1.3869		0.000
l.rm	1.3260		1.2841	0.0073	1.2892		0.000
/cut1	0.1879	0.0864	0.3450	0.0896	0.7120		
/cut2	0.9151	0.0864	1.0746	0.0896	1.4484		
/cut3	3.8257	0.0865	4.0014	0.0897	4.3806		
Pseudo R ²		0.2023		0.2048		0.2067	
n		979737		973826		973826	

Fonte: Estimativas com base em IBGE - Censo Demográfico 2010, amostra estado de São Paulo

Com relação ao teste para consistência das imputações, a IM (modelos IM4 a IM9) eleva a concentração de observações em torno do valor observado da escolaridade, em relação ao *hotdeck* (FIGURA 2). A adoção de um modelo e o controle por mais variáveis apontam uma melhoria na distribuição obtida para Y_{mis} . As distâncias entre a escolaridade imputada e a observada seguem a mesma distribuição observada com a imputação por *hotdeck*, porém com dispersão menor. As proporções de categorias imputadas iguais às observadas são consideravelmente maiores: 43,5% e 55,5% para imputações com 4 e 3 categorias de escolaridade, respectivamente.

FIGURA 2: Proporção de observações segundo a diferença entre escolaridade observada e imputada, imputação múltipla com regressão logística ordenada (IM1 a IM9), subamostra de 10000 pessoas, São Paulo, 2010



Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo

Tendo em vista os resultados da comparação entre os modelos IM1 a IM9, IM6 e IM9, foram adotados IM6 e IM9 para realizar as imputações. Com base neles, as distribuições da escolaridade imputadas são bem consistentes, pois são muito semelhantes à observada segundo sexo e idade (TABELA 11 a TABELA 14).

Por outro lado, os resultados do viés padronizado (VP) se mostram elevados. Seguindo critérios de Collins *et al* (2001), valores maiores que 0,4 ou 0,5 podem causar viés significativos nas estimativas obtidas com imputação.

TABELA 11: Distribuição dos indivíduos segundo escolaridade imputada e observada por sexo, IM com regressão logística com 3 categorias de escolaridade (IM6), para uma subamostra de 10000 pessoas, estado de São Paulo, 2010

Sexo	Anos de estudo	Observada				Imputada				VP	Extensão do intervalo
		Proporção	Erro padrão	Intervalo de confiança 95%		Proporção	Erro padrão	Intervalo de confiança 95%			
Mulheres	<=7	37.18	0.67	35.86	38.50	35.37	1.20	32.83	37.90	* -1.5180	5.07
	>=8 e <=11	42.93	0.69	41.58	44.29	44.09	1.22	41.51	46.67	* 0.9403	5.16
	>=12	19.88	0.56	18.79	20.97	20.55	0.88	18.77	22.33	* 0.7567	3.57
Homens	<=7	39.26	0.71	37.87	40.64	35.79	1.48	32.75	38.82	* -2.3411	6.07
	>=8 e <=11	43.26	0.72	41.85	44.66	45.34	1.60	42.04	48.64	* 1.3066	6.60
	>=12	17.48	0.55	16.41	18.56	18.87	1.34	16.03	21.71	* 1.0341	5.68

*Proporção imputada dentro do intervalo de confiança da observada

Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo

TABELA 12: Distribuição dos indivíduos segundo escolaridade imputada e observada por sexo, IM com regressão logística com 4 categorias de escolaridade (IM9), subamostra de 10000 pessoas, estado de São Paulo, 2010

Sexo	Anos de estudo	Observada				Imputada				Qualidade da	
		Proporção	Erro padrão	Intervalo de confiança 95%		Proporção	Erro padrão	Intervalo de confiança 95%		VP	Extensão do intervalo
Mulheres	<=3	25.00	0.60	23.81	26.18	24.06	1.17	21.48	26.64	* -0.80	5.17
	>=4 e <=7	12.19	0.46	11.29	13.08	11.03	0.80	9.29	12.78	-1.43	3.49
	>=8 e <=11	42.93	0.69	41.58	44.29	43.72	1.12	41.40	46.05	* 0.71	4.65
	>=12	19.88	0.56	18.79	20.97	21.18	0.88	19.37	23.00	* 1.47	3.63
Homens	<=3	26.55	0.64	25.30	27.80	25.02	1.33	22.25	27.79	* -1.15	5.54
	>=4 e <=7	12.71	0.48	11.76	13.65	11.23	1.00	9.12	13.33	* -1.48	4.20
	>=8 e <=11	43.26	0.72	41.85	44.66	45.49	1.55	42.24	48.74	* 1.44	6.49
	>=12	17.48	0.55	16.41	18.56	18.26	0.94	16.42	20.10	* 0.83	3.68

*Proporção imputada dentro do intervalo de confiança da observada

Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo

TABELA 13: Distribuição dos indivíduos segundo escolaridade imputada e observada por idade, IM com regressão logística com 3 categorias de escolaridade (IM6), subamostra de 10000 pessoas, estado de São Paulo, 2010

Idade	Anos de estudo	Observada			Imputada			Qualidade da imputação			
		Proporção	Erro padrão	Intervalo de confiança 95%	Proporção	Erro padrão	Intervalo de confiança 95%	VP	Extensão do intervalo		
25-29	<=7	20.23	0.96	18.34	22.11	22.17	1.89	18.05	26.28	1.03	8.24
	>=8 e <=11	54.56	1.19	52.22	56.89	49.05	1.87	45.23	52.87	-2.95	7.64
	>=12	25.21	1.04	23.18	27.25	28.79	1.88	24.80	32.77	1.90	7.97
30-34	<=7	28.28	1.11	26.11	30.45	28.57	2.26	23.66	33.48	* 0.13	9.82
	>=8 e <=11	51.66	1.23	49.25	54.07	47.13	2.22	42.48	51.77	* -2.04	9.28
	>=12	20.06	0.98	18.13	21.99	24.31	1.46	21.43	27.18	2.90	5.74
35-39	<=7	36.75	1.24	34.32	39.17	35.83	2.17	31.43	40.22	* -0.42	8.79
	>=8 e <=11	45.47	1.28	42.97	47.97	45.69	2.63	40.09	51.29	* 0.08	11.19
	>=12	17.78	0.98	15.86	19.70	18.48	1.80	14.79	22.17	* 0.39	7.38
40-44	<=7	40.47	1.28	37.97	42.98	41.00	2.33	36.30	45.71	* 0.23	9.42
	>=8 e <=11	40.95	1.28	38.44	43.46	43.14	2.20	38.76	47.51	* 0.99	8.75
	>=12	18.58	1.01	16.59	20.56	15.86	1.60	12.68	19.04	* -1.70	6.37
45-49	<=7	44.81	1.37	42.11	47.50	41.60	3.20	34.63	48.56	* -1.00	13.94
	>=8 e <=11	39.16	1.35	36.52	41.80	43.19	3.32	35.90	50.49	1.21	14.59
	>=12	16.03	1.01	14.04	18.02	15.21	2.05	10.93	19.49	* -0.40	8.55
50-54	<=7	52.76	1.41	49.99	55.53	51.71	3.06	45.33	58.09	* -0.34	12.76
	>=8 e <=11	31.31	1.31	28.73	33.88	36.88	2.72	31.35	42.40	2.05	11.05
	>=12	15.93	1.04	13.90	17.96	11.42	1.84	7.64	15.19	-2.45	7.55
55-59	<=7	58.57	1.59	55.45	61.68	55.50	3.86	47.29	63.72	* -0.79	16.43
	>=8 e <=11	27.73	1.44	24.90	30.55	35.15	4.27	25.64	44.65	1.74	19.01
	>=12	13.71	1.11	11.53	15.88	9.35	1.98	5.32	13.38	-2.20	8.06

*Proporção imputada dentro do intervalo de confiança da observada

Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo

TABELA 14: Distribuição dos indivíduos segundo escolaridade imputada e observada por idade, IM com regressão logística com 4 categorias de escolaridade (IM9), subamostra de 10000 pessoas, estado de São Paulo, 2010

Idade	Anos de estudo	Observada				Imputada				Qualidade da	
		Proporção	Erro padrão	Intervalo de confiança 95%		Proporção	Erro padrão	Intervalo de confiança 95%		VP	Extensão do intervalo
25-29	<=3	11.92	0.78	10.40	13.44	13.37	1.38	10.45	16.29	* 1.05	5.84
	>=4 e <=7	8.31	0.66	7.01	9.61	8.57	1.26	5.80	11.33	* 0.21	5.53
	>=8 e <=11	54.56	1.19	52.22	56.89	47.89	1.87	44.01	51.76	-3.56	7.76
	>=12	25.21	1.04	23.18	27.25	30.18	1.83	26.32	34.04	2.72	7.72
30-34	<=3	17.40	0.93	15.57	19.23	18.16	1.76	14.39	21.93	* 0.43	7.54
	>=4 e <=7	10.88	0.77	9.38	12.38	10.01	1.31	7.26	12.76	* -0.66	5.50
	>=8 e <=11	51.66	1.23	49.25	54.07	47.97	2.23	43.26	52.69	-1.66	9.43
	>=12	20.06	0.98	18.13	21.99	23.86	1.56	20.74	26.98	2.43	6.24
35-39	<=3	24.87	1.11	22.70	27.04	23.73	2.55	17.99	29.48	* -0.45	11.49
	>=4 e <=7	11.88	0.83	10.25	13.50	12.19	1.90	7.95	16.42	0.16	8.47
	>=8 e <=11	45.47	1.28	42.97	47.97	44.64	3.29	37.01	52.27	* -0.25	15.26
	>=12	17.78	0.98	15.86	19.70	19.43	2.40	13.99	24.88	* 0.69	10.89
40-44	<=3	28.95	1.18	26.63	31.26	28.23	1.89	24.47	31.98	* -0.38	7.51
	>=4 e <=7	11.53	0.83	9.90	13.16	12.30	1.65	8.84	15.75	* 0.47	6.91
	>=8 e <=11	40.95	1.28	38.44	43.46	43.38	3.06	36.56	50.21	* 0.80	13.66
	>=12	18.58	1.01	16.59	20.56	16.09	2.38	10.69	21.49	-1.04	10.80
45-49	<=3	30.08	1.27	27.59	32.56	29.60	2.45	24.50	34.71	* -0.19	10.21
	>=4 e <=7	14.73	0.98	12.81	16.65	12.38	1.69	8.90	15.85	-1.39	6.95
	>=8 e <=11	39.16	1.35	36.52	41.80	43.80	2.39	38.97	48.62	1.94	9.65
	>=12	16.03	1.01	14.04	18.02	14.23	1.81	10.49	17.97	* -1.00	7.48
50-54	<=3	37.39	1.37	34.71	40.07	37.20	3.30	29.88	44.52	-0.06	14.64
	>=4 e <=7	15.37	1.02	13.37	17.37	12.57	2.06	8.13	17.01	-1.36	8.88
	>=8 e <=11	31.31	1.31	28.73	33.88	37.10	2.62	31.70	42.50	2.21	10.80
	>=12	15.93	1.04	13.90	17.96	13.13	1.77	9.52	16.74	-1.58	7.22
55-59	<=3	40.60	1.58	37.50	43.71	40.99	3.10	34.58	47.40	* 0.12	12.82
	>=4 e <=7	17.96	1.24	15.54	20.39	13.04	2.33	8.09	17.98	-2.12	9.89
	>=8 e <=11	27.73	1.44	24.90	30.55	36.99	2.63	31.76	42.22	3.52	10.46
	>=12	13.71	1.11	11.53	15.88	8.99	1.96	4.84	13.14	-2.41	8.30

*Proporção imputada dentro do intervalo de confiança da observada

Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo

Ao contrário do método *hotdeck*, os modelos da IM são estimados levando em conta características dos domicílios, localidade, responsáveis pelos domicílios, que abrangem os indivíduos vivos e falecidos. Se os modelos são bem especificados e as imputações consistentes, elas devem produzir uma distribuição da escolaridade dos falecidos válida, e não necessariamente a escolaridade verdadeira para cada observação. A hipótese é de que a distribuição entre vivos e falecidos seja diferente, já que se espera que a escolaridade tenha efeito sobre a probabilidade de morte.

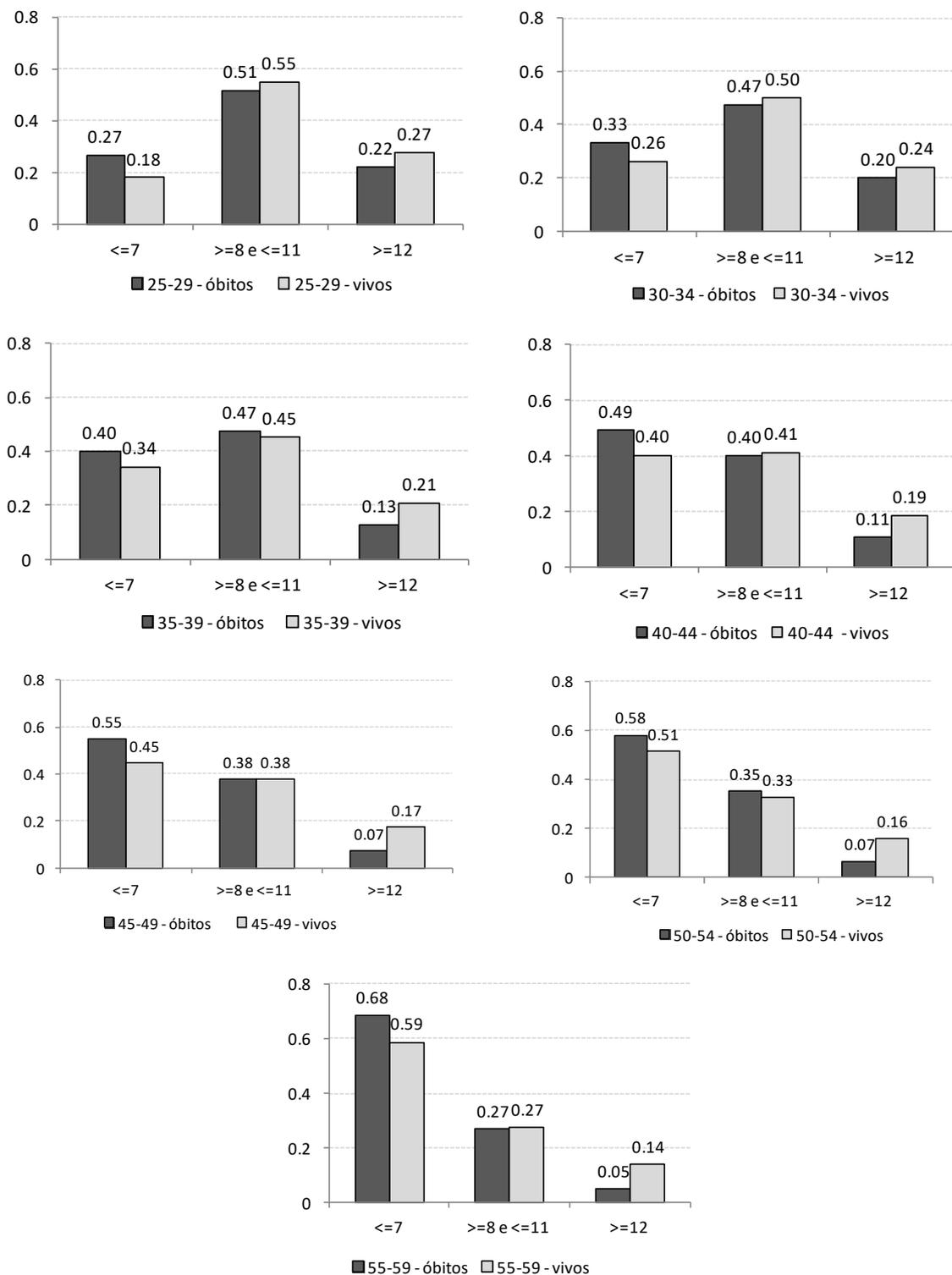
Neste aspecto, as imputações por IM6 e IM9 produzem distribuições ligeiramente diferenciadas por *status* de sobrevivência entre as mulheres (FIGURA 3 e FIGURA 5).

Para os homens, as diferenças são maiores, sobretudo entre os grupos extremos de escolaridade (FIGURA 4 e FIGURA 6). Os grupos intermediários tendem a se manter semelhantes entre vivos e mortos. Estes resultados devem resultar em diferenciais de mortalidade por escolaridade menos expressivos.

O que se deve analisar é se as imputações neste caso foram apropriadas para representar a distribuição de escolaridade dos dados faltantes, ou se, ao se basear nas informações de sobreviventes, ela não tenha sido capaz de captar as especificidades da distribuição dos óbitos.

Na seção 5.3 são apontados os resultados da imputação por modelo multivariado normal. Partindo de uma distribuição uniforme são produzidas estimativas por meio do algoritmo EM, que são utilizadas para produzir imputações por meio de método de Monte Carlo via cadeias de *Markov* (MCMC), conforme descrito no Capítulo 4.

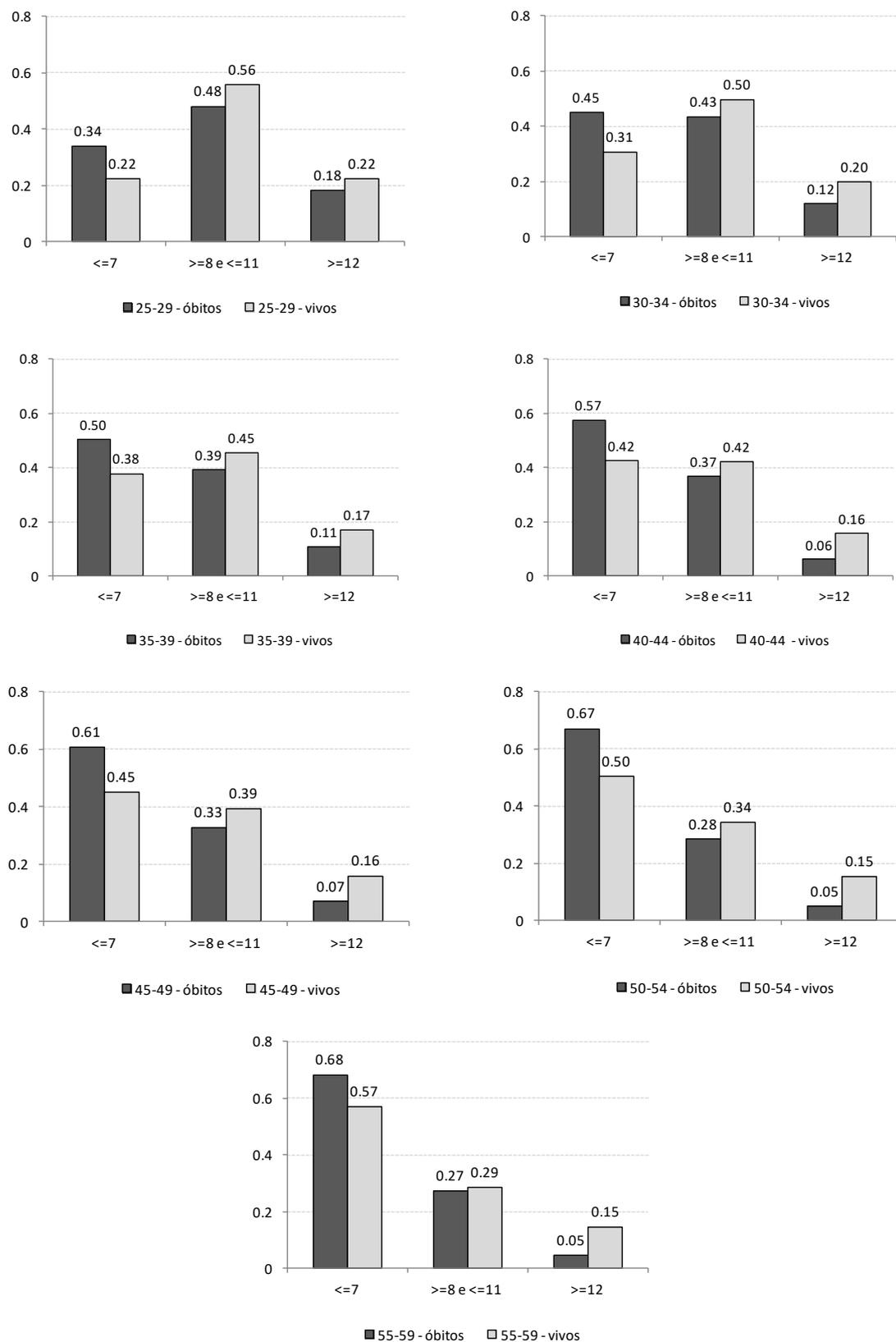
FIGURA 3: Distribuição das mulheres falecidas e vivas por escolaridade segundo a idade, IM modelo IM6, 3 categorias de escolaridade, estado de São Paulo, 2010



As diferenças são significativas apenas para os 2 grupos extremos de escolaridade

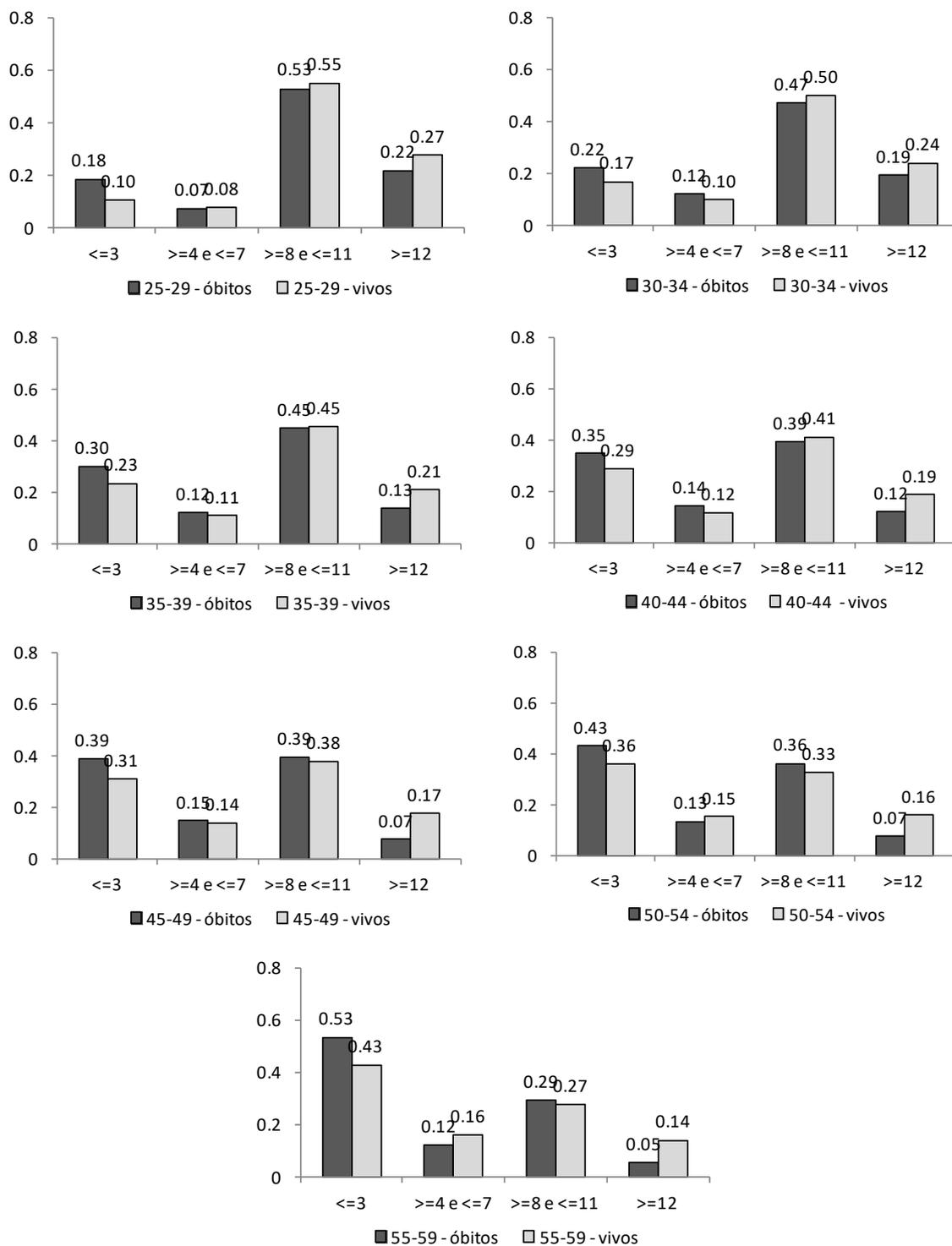
Fonte: com base em dados do IBGE - Censo Demográfico 2010, amostra estado de São Paulo

FIGURA 4: Distribuição dos homens falecidos e vivos por escolaridade segundo a idade, IM modelo IM6, 3 categorias de escolaridade, estado de São Paulo, 2010



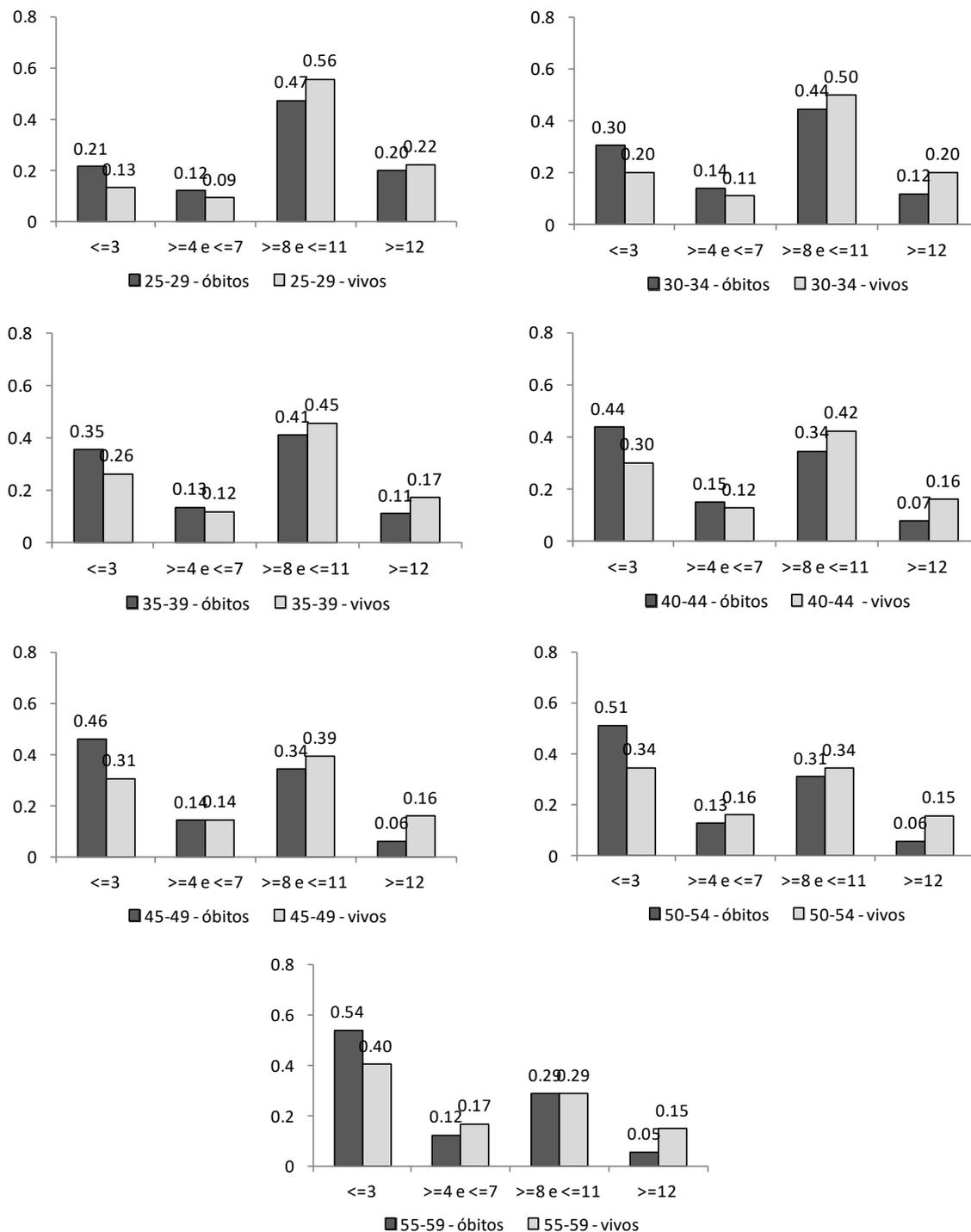
As diferenças são significativas apenas para os 2 grupos extremos de escolaridade
 Fonte: com base em dados do IBGE - Censo Demográfico 2010, amostra estado de São Paulo

FIGURA 5: Distribuição das mulheres falecidas e vivas por escolaridade segundo idade, IM modelo IM9, 4 categorias de escolaridade, estado de São Paulo, 2010



As diferenças são significativas apenas para o grupo de maior escolaridade a partir do grupo etário 30-34 anos
 Fonte: com base em dados do IBGE - Censo Demográfico 2010, amostra estado de São Paulo

FIGURA 6: Distribuição dos homens falecidos e vivos por escolaridade segundo idade, IM modelo IM9, 4 categorias de escolaridade, estado de São Paulo, 2010



As diferenças são significativas apenas para os 2 grupos extremos de escolaridade, a partir do segundo grupo etário
 Fonte: com base em dados do IBGE - Censo Demográfico 2010, amostra estado de São Paulo

5.3. Imputação por regressão multivariada normal (MVN) com algoritmo EM e MCMC: consistência dos resultados

Para o modelo MVN, a variável dependente escolaridade foi medida em anos de estudo, variando entre 0 e 21 anos. Para gerar estimativas por meio de algoritmo EM, foram utilizadas as variáveis independentes de IM6 e IM9, alterando apenas a escolaridade do vizinho e do chefe, para que fossem medidas na mesma escala da variável dependente. Embora essa forma de medir a escolaridade apresente limitações pelo número de informações perdidas, os benefícios do método podem compensar possíveis vieses adicionais gerados pela perda. O percentual de perda é muito pequeno no contexto da literatura sobre imputação. São 6705 perdas geradas, que correspondem a 0,37% da amostra (sobreviventes) do estado de São Paulo. Essas perdas ocorrem porque algumas categorias de escolaridade consideradas no censo não permitem convertê-las para anos de estudo.

Um primeiro ganho observado com o MVN está na queda acentuada dos valores dos VP em relação ao método IM por regressão logística ordenada. Embora alguns vieses da média de escolaridade imputada ainda excedam os limites de Collins *et al* (2001), eles apresentam uma queda substancial em relação ao viés causado nas distribuições observadas por IM6 e IM9 (TABELA 15 e TABELA 16).

TABELA 15: Escolaridade média com dados imputados e observados, por sexo, imputação por MVN, subamostra de 10000 pessoas, estado de São Paulo, 2010

Sexo	Observada				Imputada				Qualidade	
	Média	Erro padrão	Intervalo de confiança 95%		Média	Erro padrão	Intervalo de confiança 95%		VP	Extensão do intervalo
Mulher	8.03	0.05	7.94	8.13	8.08	0.08	7.92	8.24	0.60	0.32
Homens	7.73	0.05	7.63	7.83	7.98	0.11	7.76	8.19	2.29	0.43

Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo

TABELA 16: Escolaridade média com dados imputados e observados, por grupos etários, imputação por MNV, subamostra de 10000 pessoas, estado de São Paulo, 2010

Grupo etário	Observada				Imputada				Qualidade da imputação		
	média	Erro padrão	Intervalo de confiança 95%		média	Erro padrão	Intervalo de confiança 95%		VP	Extensão do intervalo	
25-29	9.90	0.11	9.68	10.12	9.77	0.18	9.28	10.26	*	-0.73	0.98
30-34	8.85	0.14	8.57	9.13	8.88	0.24	8.00	9.75	*	0.11	1.75
35-39	7.97	0.17	7.64	8.29	8.12	0.20	7.64	8.60	*	0.75	0.96
40-44	7.60	0.19	7.24	7.96	7.47	0.21	7.01	7.92	*	-0.62	0.91
45-49	7.41	0.20	7.03	7.80	7.34	0.31	6.26	8.43	*	-0.22	2.17
50-54	6.01	0.21	5.60	6.42	6.01	0.22	5.57	6.46	*	0.01	0.89
55-59	5.55	0.24	5.08	6.03	5.58	0.26	5.04	6.12		0.10	1.09

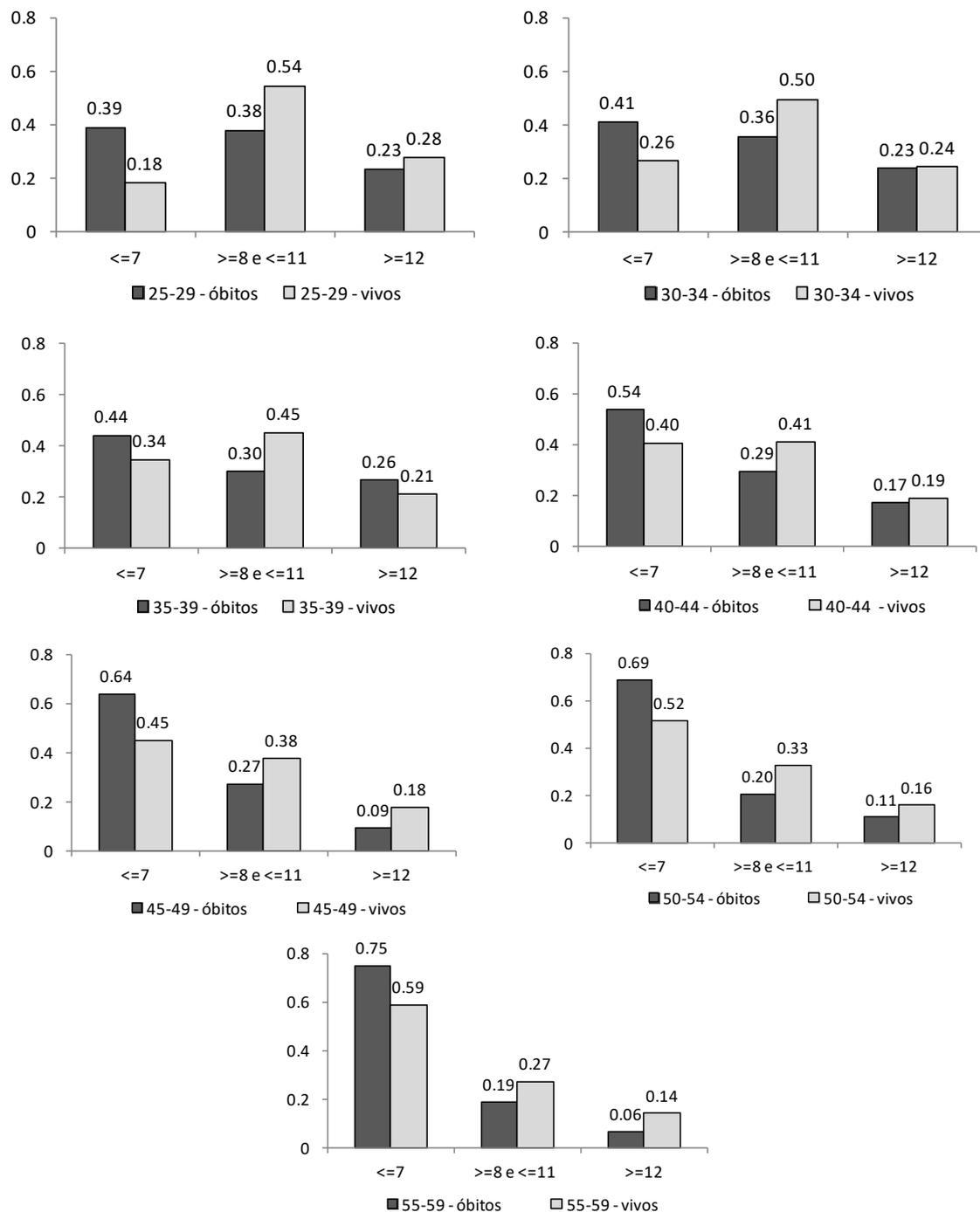
*Média imputada dentro do intervalo de confiança da observada

Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo

Com relação à distribuição de escolaridade por *status* de sobrevivência, foram realizados agrupamentos segundo categorias usadas em IM6 e IM9 para fins de comparabilidade. Fica evidente que o método de MVN produz diferenciais mais expressivos entre sobreviventes e falecidos. Há uma concentração muito maior de óbitos nos grupos de escolaridade mais baixa em relação aos vivos, sendo grande parte das diferenças significativas (FIGURA 7 a FIGURA 10). Isso deve refletir nas medições de mortalidade, produzindo gradiente educacional mais expressivo.

Na seção 5.4, as distribuições de escolaridade imputadas são comparadas à observada para os responsáveis pelos domicílios. Como a escolaridade do responsável tem um peso grande nos modelos de imputação, o objetivo é avaliar sua influência sobre a distribuição resultante do tratamento dos dados faltantes.

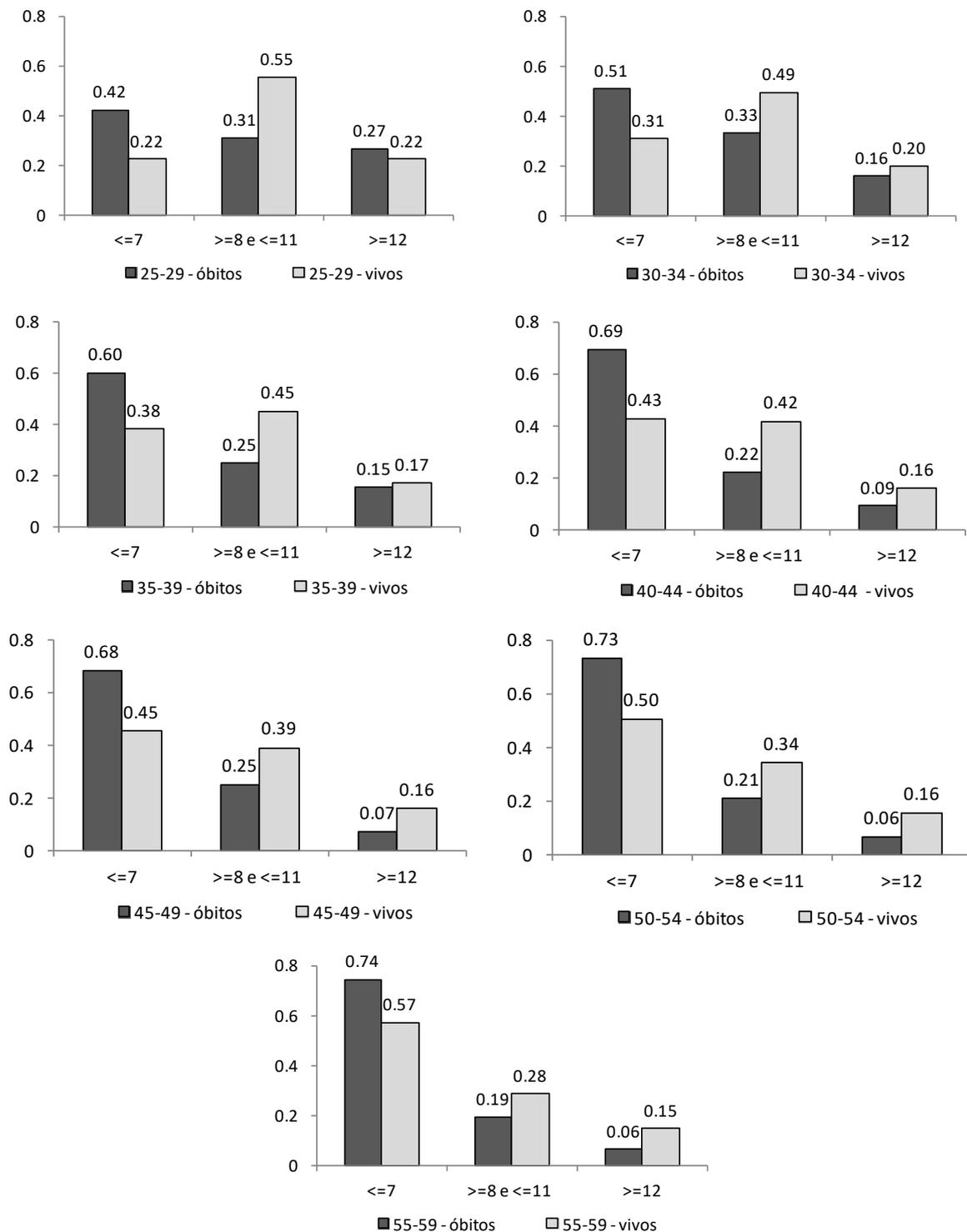
FIGURA 7: Distribuição das mulheres falecidas e vivas por escolaridade segundo idade, imputação MVN, 3 categorias de escolaridade, estado de São Paulo, 2010



As diferenças não são significativas a 5% apenas para as idades abaixo de 40 anos para o grupo mais elevado de escolaridade

Fonte: com base em dados do IBGE - Censo Demográfico 2010, amostra estado de São Paulo

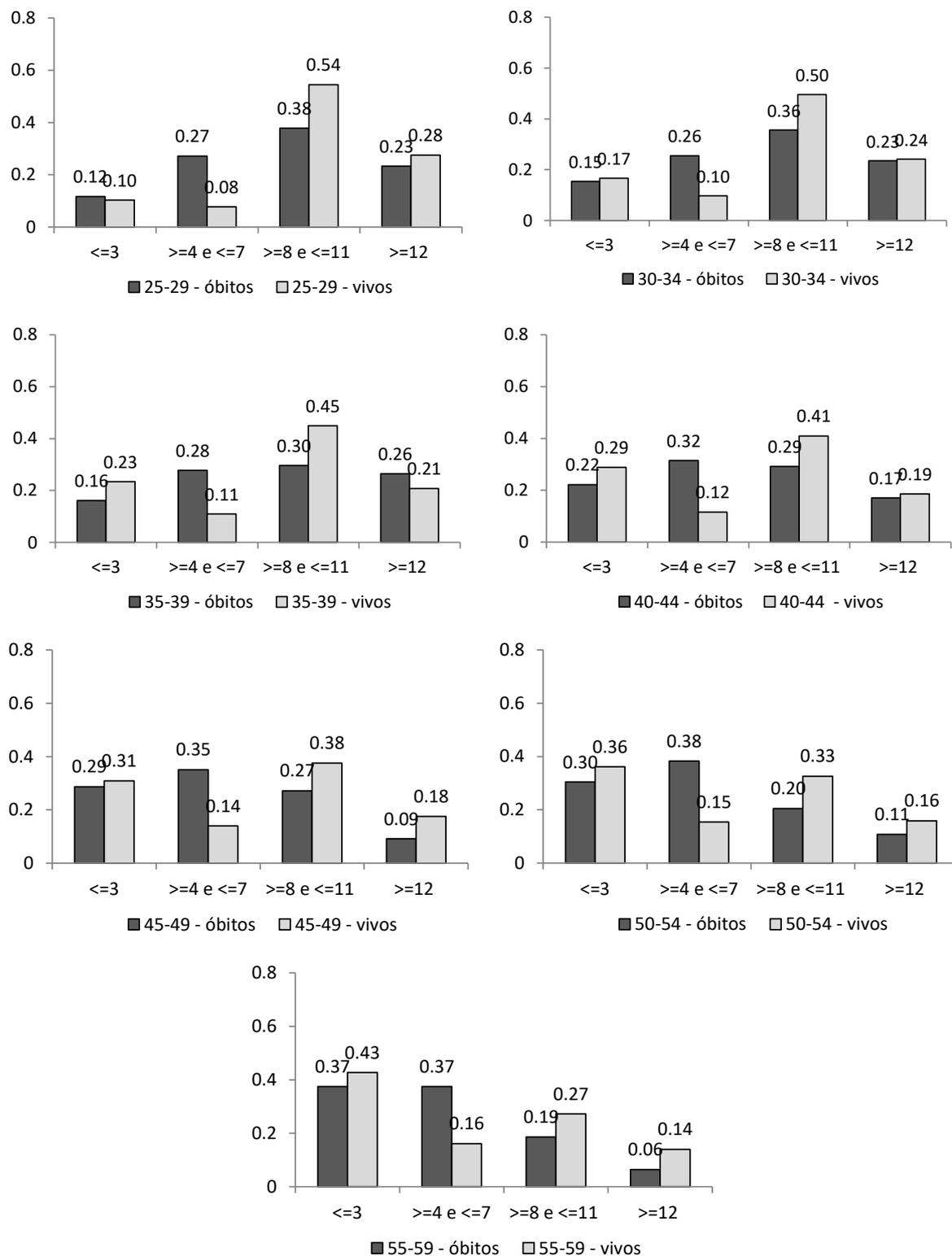
FIGURA 8: Distribuição dos homens falecidos e vivos por escolaridade segundo idade, imputação MVN, 3 categorias de escolaridade, estado de São Paulo, 2010



Todas as diferenças são significativas a 5%

Fonte: com base em dados do IBGE - Censo Demográfico 2010, amostra estado de São Paulo

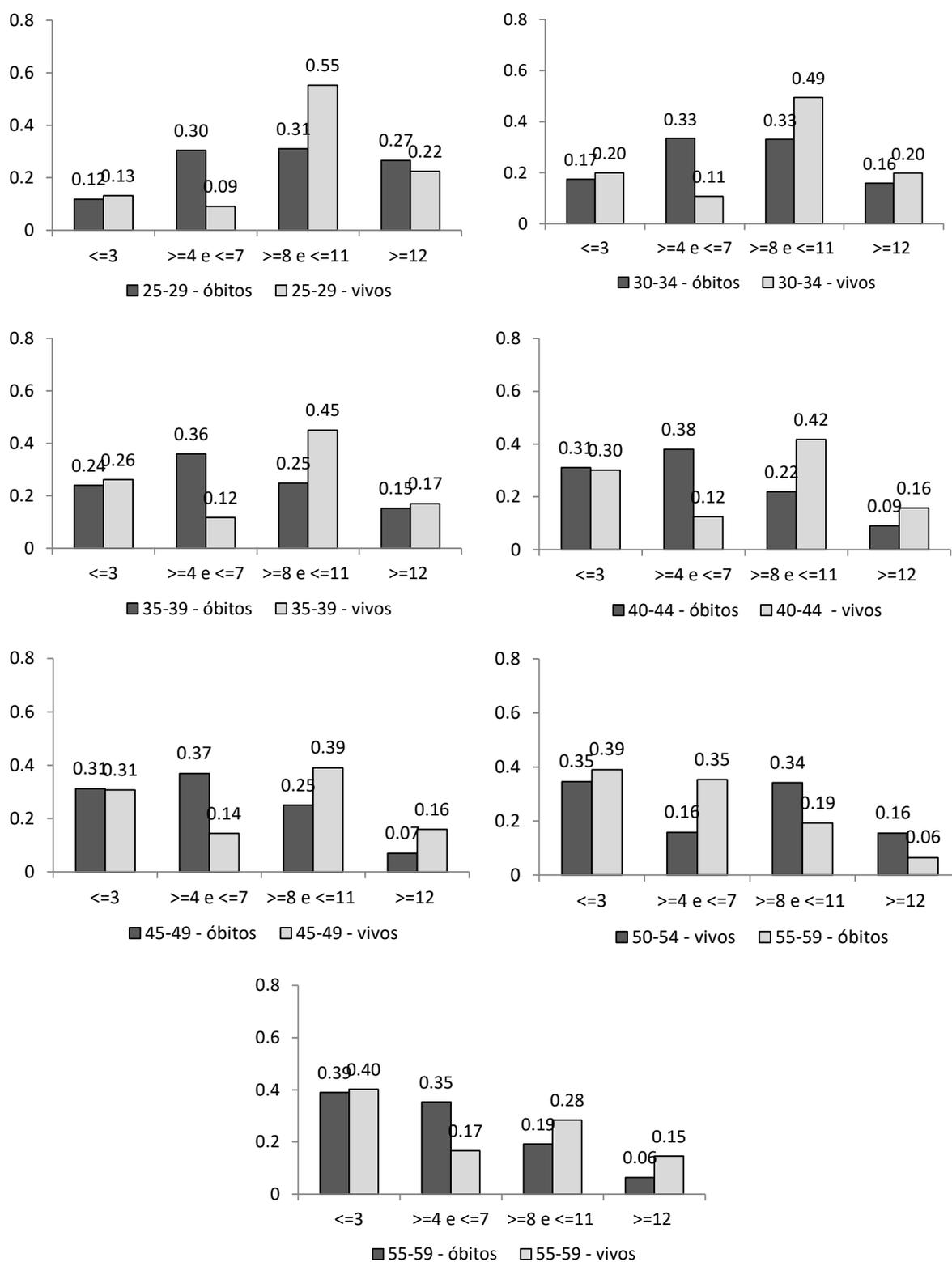
FIGURA 9: Distribuição das mulheres falecidas e vivas por escolaridade segundo idade, imputação MVN, 4 categorias de escolaridade, estado de São Paulo, 2010



Diferenças significativas a 5% para grupos etários intermediários, para grupo de maior escolaridade a partir dos 45 anos e para grupo de menor escolaridade de 35 a 44 anos e 50 anos ou mais.

Fonte: com base em dados do IBGE - Censo Demográfico 2010, amostra estado de São Paulo

FIGURA 10: Proporção de homens falecidos e vivos por escolaridade segundo idade, imputação MVN, 4 categorias de escolaridade, estado de São Paulo, 2010



Diferenças significativas a 5% para grupos etários intermediários e para grupo de maior escolaridade a partir dos 40 anos.
 Fonte: com base em dados do IBGE - Censo Demográfico 2010, amostra estado de São Paulo

5.4. Distribuições das escolaridades do responsável pelo domicílio e das imputadas por IM e MVN

O peso da escolaridade do chefe de domicílio teve grande influência na distribuição obtida por meio dos Modelos de IM6 e IM9. Estimativas apresentadas na TABELA 17 e na TABELA 18 indicam distribuições de escolaridade imputada e do chefe muito semelhantes, sobretudo quando consideramos escolaridade com 4 categorias.

Esses resultados apontam que a imputação por MVN produz uma distribuição menos influenciada pela escolaridade do chefe uma vez que, partindo de uma distribuição não informativa, utiliza-se o algoritmo EM e iterações por MCMC para ajustar os valores imputados. Além disso, as estimativas MVN produzem intervalos de confiança menores que IM, o que é preferível segundo critérios de Collins et al (2001). Em parte, isso se explica pelo fato da IM utilizar repetições¹⁷, o que acrescenta certo nível de ruído nos dados imputados.

TABELA 17: Distribuição percentual dos óbitos segundo escolaridade observada e imputada por IM6 e MVN para os responsáveis pelo domicílio, segundo sexo, 3 categorias de escolaridade, estado de São Paulo, 2010

Sexo/escolaridade	IM6				MVN				Responsável pelo domicílio					
	Proporção	Erro padrão	Intervalo de confiança 95%		Extensão do intervalo	Proporção	Erro padrão	Intervalo de confiança 95%		Extensão do intervalo	Proporção	Erro padrão	Intervalo de confiança 95%	
<i>Mulheres</i>														
≤7	53.99	1.3249	51.33	56.64	5.31	62.86	1.1094	60.66	65.00	4.35	52.90	1.1218	50.70	55.10
≥8 e ≤11	36.85	1.4858	33.74	39.96	6.21	25.18	0.9966	23.28	27.19	3.91	37.76	1.0895	35.65	39.92
≥ 12	9.16	0.8576	7.39	10.94	3.55	11.96	0.7450	10.57	13.50	2.93	9.34	0.6539	8.13	10.70
n (100%)	1923					1839					1981			
<i>Homens</i>														
≤7	58.11	0.9920	56.11	60.11	3.99	65.11	0.8084	63.51	66.68	3.17	60.52	0.7879	58.96	62.05
≥8 e ≤11	34.15	1.2499	31.41	36.89	5.47	25.05	0.7349	23.64	26.52	2.88	31.92	0.7514	30.47	33.41
≥ 12	7.74	0.6911	6.23	9.24	3.01	9.84	0.5051	8.89	10.87	1.98	7.56	0.4261	6.76	8.44
n (100%)	3495					3490					3850			

Fonte: com base em dados do IBGE - Censo Demográfico 2010, amostra estado de São Paulo

¹⁷ Neste caso, 5 repetições.

TABELA 18: Distribuição percentual dos óbitos segundo escolaridade observada e imputada por IM6 e MVN para os responsáveis pelo domicílio, segundo o sexo, estado de São Paulo, 2010

Sexo/escolaridade	IM6				MVN				Responsável pelo domicílio					
	Proporção	Erro padrão	Intervalo de confiança 95%		Extensão do intervalo	Proporção	Erro padrão	Intervalo de confiança 95%		Extensão do intervalo	Proporção	Erro padrão	Intervalo de confiança 95%	
<i>Mulheres</i>														
≤3	39.77	1.4657	36.74	42.80	6.07	27.92	1.0300	25.95	29.99	4.04	37.15	1.0859	35.02	39.28
≥4 e ≤7	12.86	1.0511	10.65	15.07	4.42	34.93	1.0946	32.82	37.11	4.29	15.75	0.8186	14.14	17.36
≥8 e ≤11	37.69	1.8073	33.67	41.71	8.04	25.18	0.9966	23.28	27.19	3.91	37.76	1.0895	35.62	39.90
≥12	9.68	0.9854	7.57	11.79	4.22	11.96	0.7450	10.57	13.50	2.93	9.34	0.6539	8.06	10.62
n (100%)	1923					1839					1981			
<i>Homens</i>														
≤3	43.59	1.1957	41.03	46.14	5.11	32.24	0.7928	30.71	33.81	3.11	44.26	0.8006	42.69	45.83
≥4 e ≤7	13.17	1.0076	10.87	15.47	4.60	32.87	0.7968	31.33	34.45	3.12	16.26	0.5948	15.09	17.43
≥8 e ≤11	35.05	1.0284	32.93	37.17	4.24	25.05	0.7349	23.64	26.52	2.88	31.92	0.7514	30.45	33.40
≥12	8.20	0.6195	6.90	9.49	2.59	9.84	0.5051	8.89	10.87	1.98	7.56	0.4261	6.72	8.39
n (100%)	3495					3490					3850			

Fonte: com base em dados do IBGE - Censo Demográfico 2010, amostra estado de São Paulo

Na seção 5.5, essas distribuições são comparadas à observada para o município de São Paulo com base em dados do MS/SIM. Conforme discutido, para a capital a cobertura da informação sobre escolaridade tem aumentado na década de 2010 nos dados do SIM, sendo o percentual de perda observado entre agosto de 2012 a julho de 2013 de 7,7% entre adultos de 25 a 59 anos.

5.5. Distribuições da escolaridade imputada, do responsável pelo domicílio e a observada no SIM

As distribuições de escolaridade do SIM foram estimadas com dados completos, onde as observações faltantes foram preenchidas por meio de método de *hotdeck*. Essa distribuição foi tomada como base para avaliação das imputações realizadas.

É importante ressaltar que a distribuição observada no SIM pode sofrer distorções dada a forma como a informação é coletada, conforme pontuado no capítulo 4. Além disso, os estados brasileiros ainda lidam com os dois modelos de declaração de óbito: o anterior a 2011 e o novo. A sistematização dos dados requer compatibilização das formas diferentes de preenchimento da informação da escolaridade, o que se torna mais uma fonte de erros na informação.

Para fins de comparação, foram estimadas as distribuições adicionais para o município de São Paulo, para comparação com o SIM. Quanto à educação do chefe, a distribuição originada do censo se diferencia da observada no SIM. As proporções de óbitos em cada faixa de escolaridade se mostram distintas, mesmo com intervalos de confiança grandes. As distâncias ficam maiores quando se adotam 4 categorias de escolaridade (TABELA 19 e TABELA 20). A distribuição educacional dos responsáveis pelos domicílios não representa a distribuição de escolaridade dos óbitos enumerados no censo, se considerarmos a distribuição do SIM como referência.

TABELA 19: Distribuição percentual dos óbitos por escolaridade do responsável pelo domicílio e observada no SIM, 3 categorias de escolaridade, estado e município de São Paulo, 2010

Anos de estudo	SIM	estado de São Paulo				município de São Paulo				
	2012/13	Proporção	Proporção	Erro padrão	Intervalo de confiança 95%	Proporção	Erro padrão	Intervalo de confiança 95%	Extensão do intervalo	
<i>Mulheres</i>										
≤ 7	54.37	62.86	1.11	60.66	65.00	55.74	3.19	49.40	61.89 *	12.49
≥ 8 e ≤ 11	30.31	25.18	1.00	23.28	27.19	26.23	2.82	21.06	32.15 *	11.09
≥ 12	15.32	11.96	0.75	10.57	13.50	18.03	2.47	13.67	23.41 *	9.74
n (100%)	6437	1898				239				
<i>Homens</i>										
≤ 7	58.73	65.11	0.81	63.51	66.68	60.34	2.26	55.82	64.69 *	8.86
≥ 8 e ≤ 11	30.45	25.05	0.73	23.64	26.52	27.29	2.06	23.44	31.52 *	8.08
≥ 12	10.82	9.84	0.51	8.89	10.87	12.37	1.52	9.67	15.68 *	6.01
n (100%)	12351	3477				468				

*Proporção do SIM dentro do intervalo de confiança

Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo; Sistema de Informações sobre Mortalidade (SIM)

TABELA 20: Distribuição percentual dos óbitos por escolaridade do responsável pelo domicílio e observada no SIM, 4 categorias de escolaridade, estado e município de São Paulo, 2010

Anos de estudo	SIM	estado de São Paulo				município de São Paulo					
	2012/13	Proporção	Proporção	Erro padrão	Intervalo de confiança 95%	Proporção	Erro padrão	Intervalo de confiança 95%	Extensão do intervalo		
<i>Mulheres</i>											
≤3	26.22	37.15	1.09	35.02	39.28	30.95	2.92	25.52	36.97	*	11.45
≥4 e ≤7	28.18	15.75	0.82	14.14	17.36	9.13	1.82	6.12	13.39		7.27
≥8 e ≤11	30.46	37.76	1.09	35.62	39.90	44.84	3.14	38.77	51.07		12.30
≥ 12	15.13	9.34	0.65	8.06	10.62	15.08	2.26	11.15	20.08	*	8.94
n (100%)	6437	1981				252					
<i>Homens</i>											
≤3	25.91	44.26	0.80	42.69	45.83	32.27	2.09	28.31	36.50		8.19
≥4 e ≤7	32.86	16.26	0.59	15.09	17.43	17.33	1.69	14.26	20.91		6.65
≥8 e ≤11	30.43	31.92	0.75	30.45	33.40	39.64	2.19	35.44	44.00		8.57
≥ 12	10.81	7.56	0.43	6.72	8.39	10.76	1.38	8.32	13.79	*	5.47
n (100%)	12351	3850				502					

*Proporção do SIM dentro do intervalo de confiança

Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo; Sistema de Informações sobre Mortalidade (SIM)

Como a distribuição dos óbitos obtida por imputação múltipla com modelos IM6 e IM9 produzem proporções de escolaridade próximas às dos responsáveis pelos domicílios, elas também se distanciam da observada no SIM (TABELA 20 e TABELA 21). As consequências desta distância são amenizadas com a escolaridade em 3 categorias, dado que o primeiro grupo de escolaridade passa a concentrar grande parte da população do estado e do município. Para os homens, as diferenças são não significativas.

Ao desagregar a escolaridade em 4 grupos, as diferenças passam a ser significativas. As imputações subestimam os óbitos entre 4 e 7 anos de escolaridade em favor dos grupos adjacentes.

TABELA 21: Distribuição percentual dos óbitos por escolaridade imputada por modelo IM6 e observada no SIM, 3 categorias de escolaridade, estado e município de São Paulo, 2010

Anos de estudo	SIM	estado de São Paulo				município de São Paulo				
	2012/13	Proporção	Proporção	Erro padrão	Intervalo de confiança 95%	Proporção	Erro padrão	Intervalo de confiança 95%	Extensão do intervalo	
<i>Mulheres</i>										
≤7	54.37	53.99	1.32	51.33	56.64 *	42.52	3.36	35.88	49.16	13.29
≥8 e ≤11	30.31	36.85	1.49	33.74	39.96	44.20	3.87	36.33	52.07	15.74
≥ 12	15.32	9.16	0.86	7.39	10.94	13.28	3.34	5.94	20.62 *	14.68
n (100%)	6437	1923				238				
<i>Homens</i>										
≤7	58.73	58.11	0.99	56.11	60.11 *	52.32	3.46	44.85	59.79 *	14.94
≥8 e ≤11	30.45	34.15	1.25	31.41	36.89	37.17	3.32	30.00	44.35 *	14.34
≥ 12	10.82	7.74	0.69	6.23	9.24	10.51	1.84	6.72	14.30 *	7.58
n (100%)	12351	3495				453				

*Proporção do SIM dentro do intervalo de confiança

Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo; Sistema de Informações sobre Mortalidade (SIM)

TABELA 22 Distribuição percentual dos óbitos por escolaridade imputada por modelo IM6 e observada no SIM, estado e município de São Paulo, 2010

Anos de estudo	SIM	estado de São Paulo				município de São Paulo				
	2012/13	Proporção	Proporção	Erro padrão	Intervalo de confiança 95%	Proporção	Erro padrão	Intervalo de confiança 95%	Extensão do intervalo	
<i>Mulheres</i>										
≤3	26.22	39.77	1.47	36.74	42.80	30.89	4.46	21.11	40.68	19.57
≥4 e ≤7	28.18	12.86	1.05	10.65	15.07	12.76	4.58	1.43	24.10	22.67
≥8 e ≤11	30.46	37.69	1.81	33.67	41.71	41.95	4.38	32.62	51.28	18.66
≥ 12	15.13	9.68	0.99	7.57	11.79	14.39	2.40	9.63	19.15 *	9.53
n (100%)	6437	1923				238				
<i>Homens</i>										
≤3	25.91	43.59	1.20	41.03	46.14	37.65	2.75	32.04	43.26	11.22
≥4 e ≤7	32.86	13.17	1.01	10.87	15.47	13.63	2.58	7.86	19.40	11.54
≥8 e ≤11	30.43	35.05	1.03	32.93	37.17	38.85	3.37	31.52	46.17	14.65
≥ 12	10.81	8.20	0.62	6.90	9.49	9.87	1.84	6.02	13.72 *	7.71
n (100%)	12351	3495				453				

*Proporção do SIM dentro do intervalo de confiança

Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo; Sistema de Informações sobre Mortalidade (SIM)

As imputações por MVN produzem distribuições para o município de São Paulo estatisticamente iguais às observadas no SIM (TABELA 23 e TABELA 24). Embora os

intervalos de confiança das estimativas do censo sejam grandes, eles são menores se comparados aos produzidos por IM. A grande diferença observada na distribuição com 4 categorias de escolaridade no caso da IM se reduzem substancialmente com os resultados de MVN. Tomando-se a distribuição do SIM como referência, as estimativas por MVN são as mais consistentes e parecem ser mais adequadas para representar a distribuição de escolaridade dos óbitos no censo.

TABELA 23: Distribuição percentual dos óbitos por escolaridade imputada por MVN e observada no SIM, 3 categorias de escolaridade, estado e município de São Paulo, 2010

Anos de estudo	SIM 2012/13	estado de São Paulo				município de São Paulo				
	Proporção	Proporção	Erro padrão	Intervalo de confiança 95%		Proporção	Erro padrão	Intervalo de confiança 95%		Extensão do intervalo
<i>Mulheres</i>										
≤7	54.37	62.86	1.11	60.66	65.00	55.74	3.19	49.40	61.89	* 12.49
≥8 e ≤11	30.31	25.18	1.00	23.28	27.19	26.23	2.82	21.06	32.15	* 11.09
≥ 12	15.32	11.96	0.75	10.57	13.50	18.03	2.47	13.67	23.41	* 9.74
n (100%)	6437	1898				239				
<i>Homens</i>										
≤7	58.73	65.11	0.81	63.51	66.68	60.34	2.26	55.82	64.69	* 8.86
≥8 e ≤11	30.45	25.05	0.73	23.64	26.52	27.29	2.06	23.44	31.52	* 8.08
≥ 12	10.82	9.84	0.51	8.89	10.87	12.37	1.52	9.67	15.68	* 6.01
n (100%)	12351	3477				468				

*Proporção do SIM dentro do intervalo de confiança

Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo; Sistema de Informações sobre Mortalidade (SIM)

TABELA 24: Distribuição percentual dos óbitos por escolaridade imputada por MVN e observada no SIM, 4 categorias de escolaridade, estado e município de São Paulo, 2010

Anos de estudo	SIM	estado de São Paulo				município de São Paulo				Extensão do intervalo
	2012/13	Proporção	Proporção	Erro padrão	Intervalo de confiança 95%	Proporção	Erro padrão	Intervalo de confiança 95%		
<i>Mulheres</i>										
≤3	26.22	27.92	1.03	25.95	29.99 *	21.72	2.65	16.96	27.38 *	10.42
≥4 e ≤7	28.18	34.93	1.09	32.82	37.11	34.02	3.04	28.31	40.23	11.92
≥8 e ≤11	30.46	25.18	1.00	23.28	27.19	26.23	2.82	21.06	32.15 *	11.09
≥12	15.13	11.96	0.75	10.57	13.50	18.03	2.47	13.67	23.41 *	9.74
n (100%)	6437	1898				239				
<i>Homens</i>										
≤3	25.91	32.24	0.79	30.71	33.81	26.44	2.04	22.63	30.63 *	8.00
≥4 e ≤7	32.86	32.87	0.80	31.33	34.45 *	33.90	2.19	29.74	38.32 *	8.58
≥8 e ≤11	30.43	25.05	0.73	23.64	26.52	27.29	2.06	23.44	31.52 *	8.08
≥12	10.81	9.84	0.51	8.89	10.87 *	12.37	1.52	9.67	15.68 *	6.01
n (100%)	12351	3477				468				

*Proporção do SIM dentro do intervalo de confiança

Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo; Sistema de Informações sobre Mortalidade (SIM)

5.6. Considerações finais

A imputação por MVN se desdobra como um método mais adequado para a proposta da tese, uma vez que os resultados da sua validação são superiores. Além disso, a distribuição de escolaridade dos óbitos se assemelha às observadas no SIM. Uma desvantagem do método é a perda de escolaridade entre os adultos sobreviventes, pela impossibilidade de converter algumas categorias de escolaridade do censo em anos de estudo. No entanto, a magnitude não é substancial, conforme ilustrado na TABELA 25, de modo que mais de 99,5% das observações possuem informações sobre escolaridade. É importante observar que, embora tenha se adotado as mesmas variáveis independentes para os modelos IM6 e MVN, no segundo o percentual de perda entre os óbitos é diferente pelo processo de simulação envolver queima de observações que não se encaixem na distribuição estacionária visada pelo método. Adicionalmente, o diagnóstico da MVN não aponta nenhum problema de tendência nas iterações, tendo a convergência sido alcançada com 500 iterações.

TABELA 25: Distribuição dos óbitos de pessoas entre 25 e 59 anos, estado e município de São Paulo 2010, 2012-2013

Origem	óbitos		sobreviventes		Total		
	UF-SP	Município São Paulo	UF-SP	Município São Paulo	UF-SP	Município São Paulo	
Amostra Censo - sem perda	<i>n</i>	5831	754	1791565	279329	1797396	280083
	%	100.00	100.00	100.00	100.00	100.00	100.00
Amostra Censo - com perda da imputação IM6	<i>n</i>	5418	691	1791565	279329	1796983	280020
	%	92.92	91.64	100.00	100.00	99.98	99.98
Amostra Censo - com perda da imputação MVN	<i>n</i>	5329	707	1788888	278369	1794217	279076
	%	91.39	93.77	99.85	99.66	99.82	99.64

Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo; MS/SIM microdados de 2012 e 2013

Com relação às categorias de escolaridade, dado que os resultados apontam consistência maior para número menor de agrupamentos, serão adotadas três categorias para os modelos de análise da mortalidade: ≤ 7 anos; ≥ 8 e ≤ 11 anos e ≥ 12 anos de escolaridade. Para fins práticos, os grupos poderão ser chamados, respectivamente, como grupo de baixa, média e alta escolaridade. Além da qualidade das imputações, o uso de três categorias é um benefício para os adultos mais jovens o peso do grupo de escolaridade ≤ 3 anos é pequeno, o que comprometeria a qualidade das estimativas ao proporcionar graus de liberdade pequenos, sobretudo ao considerar o município de São Paulo, cuja amostra é menor que a do estado.

6. MORTALIDADE POR NÍVEIS DE ESCOLARIDADE

Neste capítulo apresentam-se as estimativas de mortalidade usando as escolaridades imputadas, a escolaridade do responsável pelo domicílio e a registrada no SIM. Na seção 6.1 e 6.2 são descritas, respectivamente, as estimativas de mortalidade feminina e masculina para o município de São Paulo usando as três fontes de escolaridade referidas. O objetivo é investigar diferenciais provenientes do uso da escolaridade imputada e do responsável pelo domicílio e comparar com as registradas pelo SIM. O pressuposto é de que as estimativas com a escolaridade imputada represente melhor os diferenciais de mortalidade por escolaridade, pelo menos aquelas estimadas por MVN, tendo por base os resultados discutidos no capítulo 5.

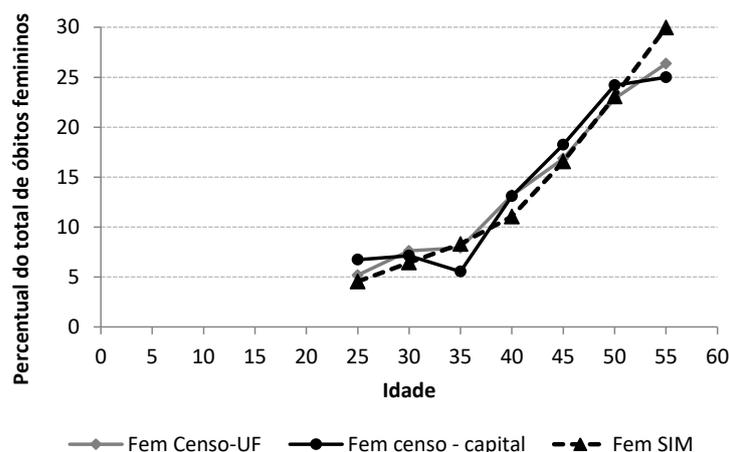
Na seção 6.3, as estimativas de mortalidade por níveis de escolaridade imputada são rerepresentadas com as devidas correções para sub ou sobreenumeração. Neste caso, a correção foi aplicada apenas as taxas de mortalidade calculadas com base na escolaridade imputada pelo melhor método. De posse das taxas ajustadas, o objetivo foi avaliar o gradiente educacional da mortalidade por idade e o seu comportamento segundo o sexo. A seção 6.4 traz discussões dos resultados *vis a vis* as evidências existentes para o Brasil e outras populações e os principais arcabouços teóricos discutidos no Capítulo 2.

6.1. Mortalidade feminina no município de São Paulo: estimativas de diferenciais por escolaridade imputada, por escolaridade do chefe e escolaridade do SIM

A distribuição dos óbitos por sexo é semelhante entre os dados do Censo e do SIM. Entre 64 e 67% dos óbitos ocorreram entre os homens, indicando grande diferencial por sexo (TABELA 24). Quanto à distribuição etária dos óbitos, observam-se algumas diferenças em relação à distribuição etária. Conforme ilustrado na FIGURA 11, no SIM há uma concentração maior de óbitos femininos no último grupo etário, o que pode ser explicado pela subenumeração de óbitos no censo ocorridos em domicílios unipessoais. Outra diferença é o percentual mais baixo de óbitos no censo entre as idades 35-39 anos para o município de São Paulo, em relação ao registrado no SIM. Essa diferença pode estar

relacionada a erros amostrais, subenumeração de óbitos no domicílio, ou ao erro de declaração de idade no Censo. Não foi realizada avaliação deste aspecto, mas no caso dos falecidos a idade é declarada pelo sobrevivente, o que torna essa informação suscetível a erro. Isso pode ter efeitos sobre as estimativas de mortalidade por idade.

FIGURA 11: Distribuição percentual dos óbitos de pessoas entre 25 e 59 anos segundo grupos etários, populações femininas do estado e município de São Paulo, 2010, 2012-2013



Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo; MS/SIM microdados de 2012 e 2013

Os coeficientes do Modelo de Poisson apontam um gradiente da mortalidade mais expressivo com os dados do SIM e imputados por MVN (TABELA 26). O modelo em que se considera a escolaridade do chefe e as imputações por IM6 não aponta diferenciais significativos entre a mortalidade do grupo de menor escolaridade e o intermediário. O grupo de maior escolaridade apresenta uma mortalidade que equivale, respectivamente, a 56% e 42% da taxa do grupo de menor escolaridade, dependendo do método. Por outro lado, as estimativas com base no SIM e nas imputações por MVN indicam uma razão maior, igual a 34 e 37% entre a taxa de mortalidade dos grupos extremos (maior e menor) de escolaridade para o município de São Paulo. Além disso, o grupo de escolaridade intermediária possui uma nítida vantagem em relação ao grupo de menor escolaridade (sua taxa é cerca de 50% da taxa de mortalidade das pessoas com escolaridade inferior a 8 anos) e uma desvantagem em relação ao grupo escolaridade mais alta, como deveríamos esperar.

TABELA 26: Resultado dos coeficientes e razões das taxas de incidência (IRR) dos modelos de Poisson, população feminina do município de São Paulo, 2010, 2012-2013

Variáveis	Escolaridade do responsável pelo domicílio - município SP				Escolaridade imputada por modelo IM6 - município SP				Escolaridade imputada MNV - município SP				Dados do SIM - município de	
	morte	IRR	Erro padrão	Intervalo de confiança 95%	IRR	Erro padrão	Intervalo de confiança 95%	IRR	Erro padrão	Intervalo de confiança 95%	IRR	Erro padrão	IRR	Erro padrão
Idade														
30 a 34	1.12	0.38	0.58	2.17	0.95	0.33	0.48	1.89	1.04	0.35	0.53	2.01	1.40	0.11 *
35 a 39	0.97	0.35	0.48	1.97	0.91	0.33	0.45	1.86	0.85	0.31	0.42	1.74	1.93	0.14 *
40 a 45	2.52	0.75	1.40	4.53 *	2.18	0.66	1.20	3.95 *	2.05	0.62	1.14	3.71 *	2.67	0.19 *
45 a 49	3.73	1.06	2.14	6.50 *	3.17	0.91	1.80	5.58 *	2.91	0.84	1.66	5.11 *	4.18	0.28 *
50 a 54	5.47	1.50	3.19	9.37 *	4.57	1.28	2.65	7.90 *	4.24	1.17	2.46	7.30 *	6.18	0.40 *
55 a 59	6.79	1.86	3.97	11.62 *	5.67	1.58	3.28	9.78 *	5.10	1.41	2.97	8.78 *	9.18	0.58 *
grupos de escolaridade														
>=8 e <=11	1.16	0.16	0.89	1.52	0.94	0.14	0.70	1.26	0.51	0.07	0.39	0.68 *	0.50	0.01 *
>=12	0.56	0.11	0.39	0.82 *	0.42	0.12	0.23	0.75 *	0.34	0.07	0.24	0.50 *	0.37	0.01 *
_cons	0.001	0.000	0.000	0.001 *	0.001	0.000	0.000	0.001 *	0.001	0.000	0.001	0.002 *	0.001	0.000 *
ln(anopes)	1.00				1.00		0.00	0.00	1.00		0.00	0.00	1.00	

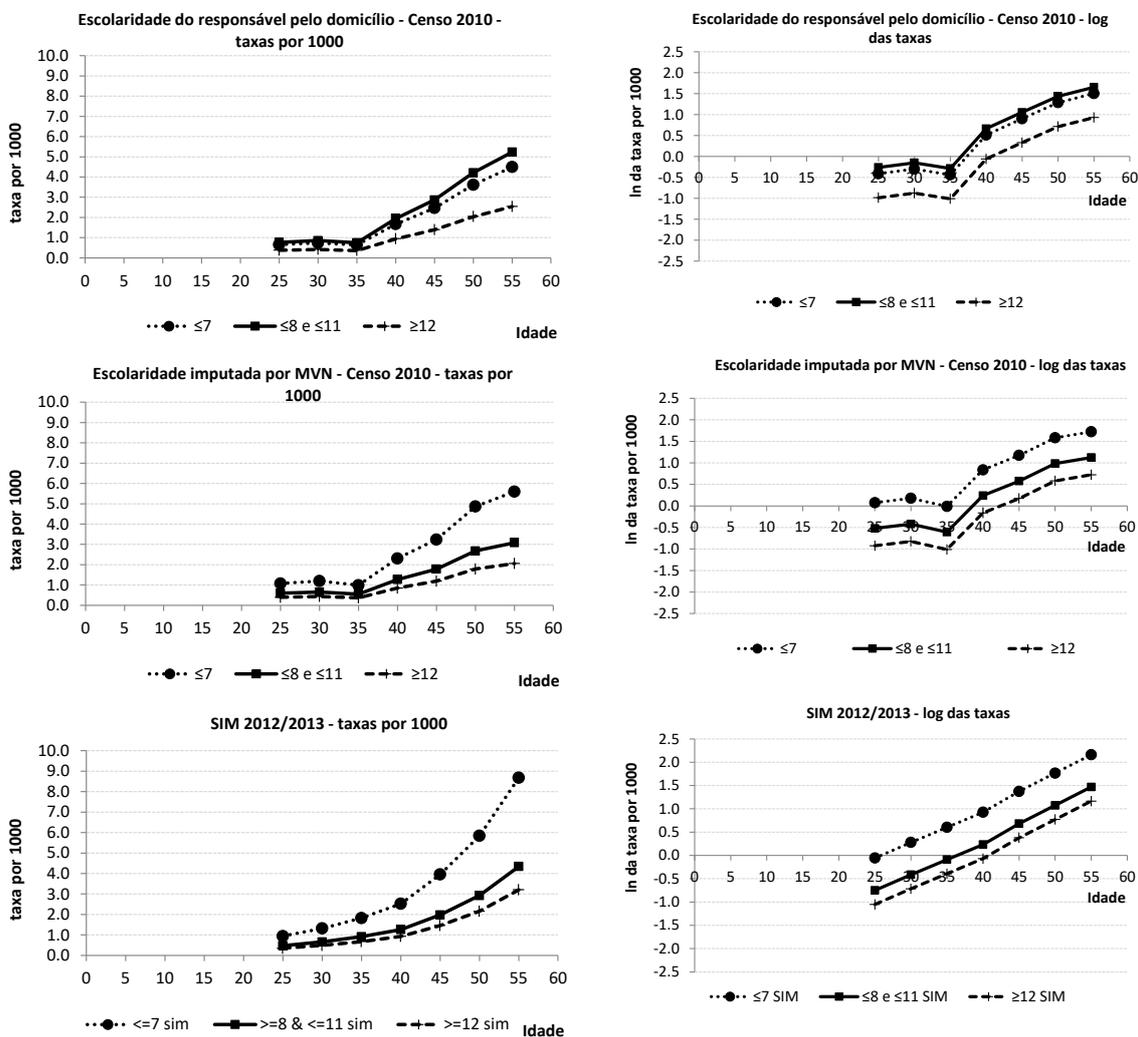
* P-valor ≤ 0.01 P-valor ≥ 0.26

Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo; MS/SIM microdados de 2012 e 2013

Como os modelos que envolvem imputação por MVN exibem coeficientes mais próximos aos do SIM, para simplificação da análise, as taxas por IM6 não foram apresentadas. Quando se comparam as taxas de mortalidade por idade estimadas a partir dos coeficientes dos modelos visualizam-se os diferentes padrões segundo a fonte da escolaridade (FIGURA 12). O padrão do gradiente educacional da mortalidade alcançado com as escolaridades imputadas se aproxima mais do observado a partir do SIM. Para os dois primeiros grupos de escolaridade, as taxas de mortalidade estimadas por meio de imputações se aproximam mais das do SIM, se comparadas com a escolaridade do chefe do domicílio.

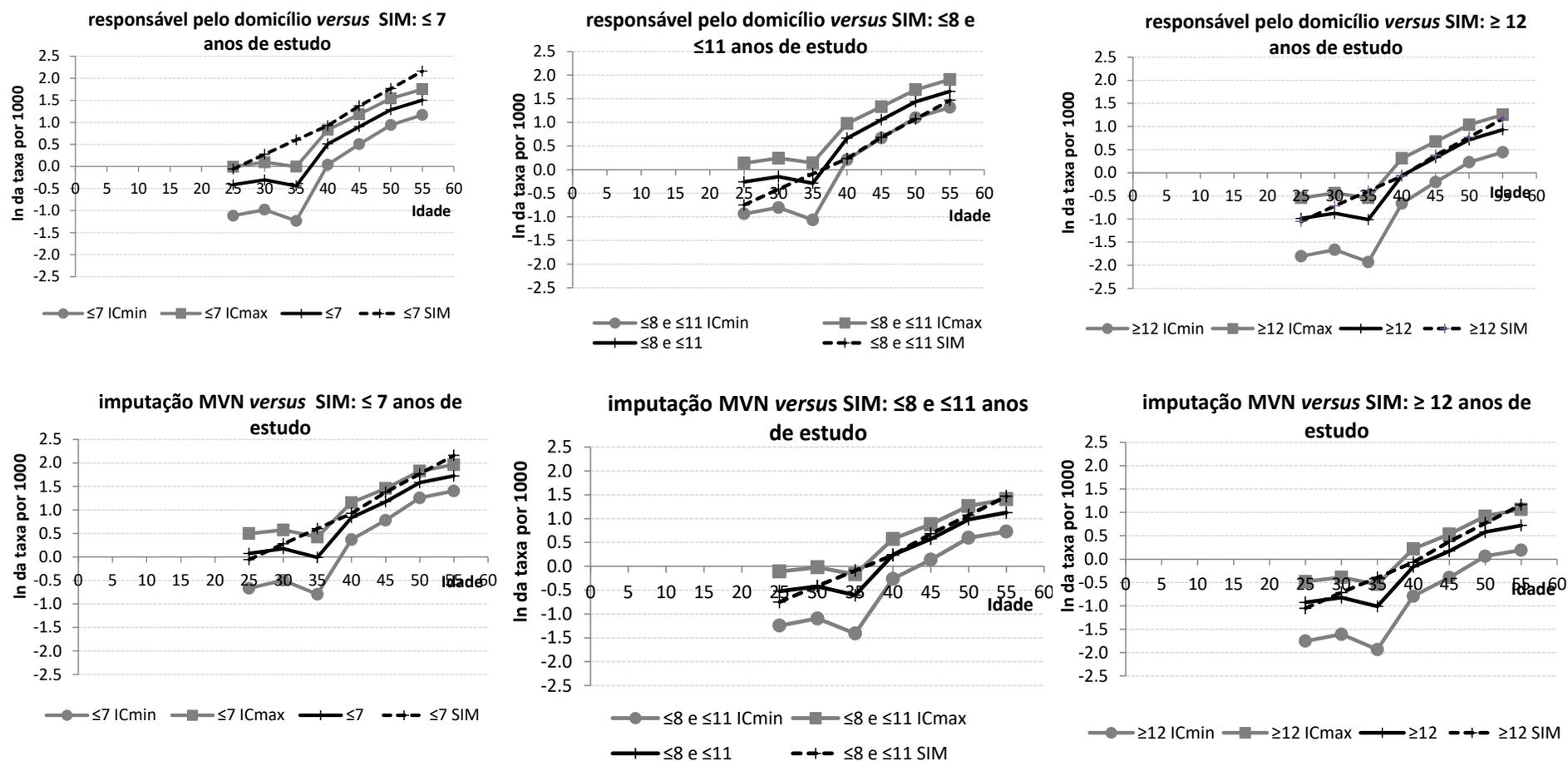
As taxas de mortalidade por idade para a população com 7 anos ou menos de estudo obtidas com o SIM, ficam fora do intervalo de confiança das taxas estimadas para indivíduos com chefes com esta escolaridade (FIGURA 13). Para o grupo entre 8 e 11 anos de escolaridade, o intervalo de confiança inclui as datas do SIM, porém ficam próxima do limite inferior do intervalo.

FIGURA 12: Taxas de mortalidade (por mil) segundo escolaridade do responsável pelo domicílio, escolaridade imputada por MVN e observada no SIM, população feminina do município de São Paulo, 2010, 2012-2013



Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo; MS/SIM microdados de 2012 e 2013

FIGURA 13: Taxas de mortalidade (por mil) e intervalos de confiança estimados (IC) segundo escolaridade do responsável pelo domicílio, escolaridade imputada por MVN e observada no SIM, população feminina do município de São Paulo, 2010, 2012-2013

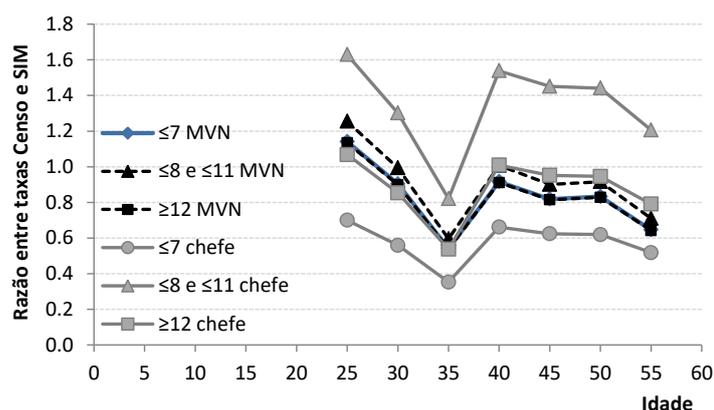


Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo; MS/SIM microdados de 2012 e 2013

Os resultados da imputação por MVN também são superiores aos encontrados com a escolaridade do chefe do domicílio em termos de cobertura das taxas de mortalidade. Considerando que os registros do SIM cobrem 100% dos óbitos de São Paulo, ambas as taxas calculadas com base no Censo (chefe do domicílio ou imputadas) sofrem, em média, com mesma magnitude de sub ou sobre estimação. Mas curiosamente as taxas obtidas com escolaridade do chefe sofrem oscilações muito maiores, conforme ilustrado na FIGURA 14. As taxas de mortalidade entre mulheres de média escolaridade são expressivamente sobre estimadas, atingindo um nível 60% maior que a taxa do SIM para as mulheres de 25 a 29 anos. A exceção ocorre para as idades de 35 a 39 anos, para as quais a proporção de óbitos é subestimada. Opostamente, as taxas entre mulheres de baixa escolaridade são extremamente subestimadas.

No caso das taxas de mortalidade estimadas com MVN, as oscilações são menores e podem estar relacionadas a sub e sobre enumeração dos óbitos do censo enquanto as do chefe a problemas de sua utilização para representar indivíduos falecidos em um contexto de mudança na cobertura educacional no Brasil, conforme é discutido na seção 6.4.

FIGURA 14: Razão entre as taxas de mortalidade estimadas pelo Censo, por escolaridade do responsável pelo domicílio e escolaridade imputada por MVN, e as taxas do SIM, população feminina município de São Paulo, 2010, 2012-2013

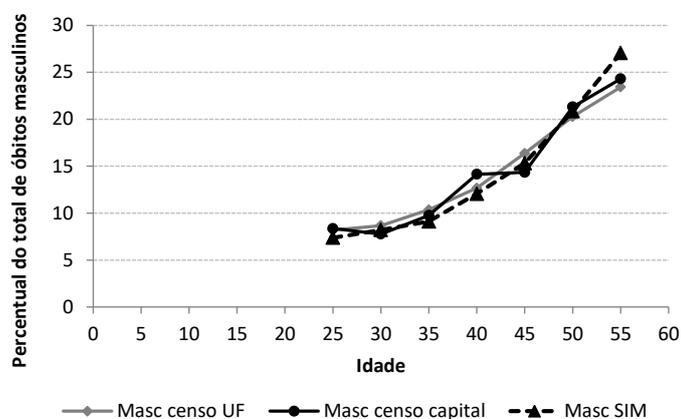


Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo; MS/SIM microdados de 2012 e 2013

6.2. Mortalidade masculina no município de São Paulo: estimativas de diferenciais por escolaridade imputada, por escolaridade do chefe e escolaridade do SIM

A distribuição dos óbitos masculinos aponta um peso ligeiramente maior das idades entre 40-45 anos para a capital em relação ao estado de São Paulo e à observada no SIM. Isso tem consequências sobre as taxas estimadas para estas idades.

FIGURA 15: Distribuição percentual dos óbitos de pessoas entre 25 e 59 anos segundo grupos etários, populações masculinas do estado e município de São Paulo, 2010, 2012-2013



Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo; MS/SIM microdados de 2012 e 2013

Os coeficientes dos Modelos de Poisson para os homens são mais expressivos e apontam um gradiente da mortalidade mesmo quando consideramos imputação múltipla ou escolaridade dos chefes (TABELA 27). Contudo, os diferenciais são bem maiores nos modelos com imputação por MVN e com dados do SIM, o que é consistente com os resultados apresentados para as mulheres.

TABELA 27: Resultado dos coeficientes e razões das taxas de incidência (IRR) dos modelos de Poisson, população masculina do município de São Paulo, 2010, 2012-2013

Variáveis	Escolaridade do responsável pelo domicílio - município SP			Escolaridade imputada por modelo IM6 - município SP			Escolaridade imputada MNV - município SP			Dados do SIM - município de São Paulo		
	Coef	IRR	Erro Padrão (coef)	Coef	IRR	Erro Padrão (coef)	Coef	IRR	Erro Padrão (coef)	Coef	IRR	Erro Padrão (coef)
idgrup2												
30 a 34	0.0254	1.0257	0.2224 **	0.0305	1.0310	0.2267 **	-0.0794	0.9237	0.2055 **	0.1125	1.1191	0.0457 *
35 a 39	0.3600	1.4333	0.2103 **	0.3257	1.3850	0.2156 **	0.1906	1.2100	0.2550 **	0.2702	1.3102	0.0446 *
40 a 45	0.8162	2.2620	0.1947 *	0.6252	1.8687	0.2061 *	0.5934	1.8101	0.3548 *	0.5996	1.8213	0.0422 *
45 a 49	0.9523	2.5917	0.1942 *	0.8778	2.4056	0.2013 *	0.7296	2.0743	0.4046 *	0.9603	2.6126	0.0405 *
50 a 54	1.4628	4.3179	0.1822 *	1.2787	3.5920	0.1927 *	1.2105	3.3551	0.6151 *	1.3627	3.9066	0.0388 *
55 a 59	1.8083	6.1001	0.1791 *	1.4870	4.4239	0.1934 *	1.5210	4.5766	0.8263 *	1.8052	6.0809	0.0377 *
educimp												
>=8 e <=11	-0.2151	0.8065	0.0953 *	-0.4653	0.6279	0.1518 *	-1.0368	0.3546	0.0379 *	-0.7904	0.4537	0.0204 *
>=12	-1.1271	0.3240	0.1501 *	-1.3444	0.2607	0.2098 *	-1.4139	0.2432	0.0346 *	-1.4703	0.2299	0.0301 *
_cons	-6.1210	0.0022	0.1621 *	-5.9345	0.0026	0.1855 *	-5.5911	0.0037	0.0006 *	-5.6780	0.0034	0.0349 *
ln(anopes)	1.0000	1.0000		1.0000	1.0000		1.0000	1.0000		1.0000	1.0000	

* P-valor ≤0.01

** P-valor =0.09

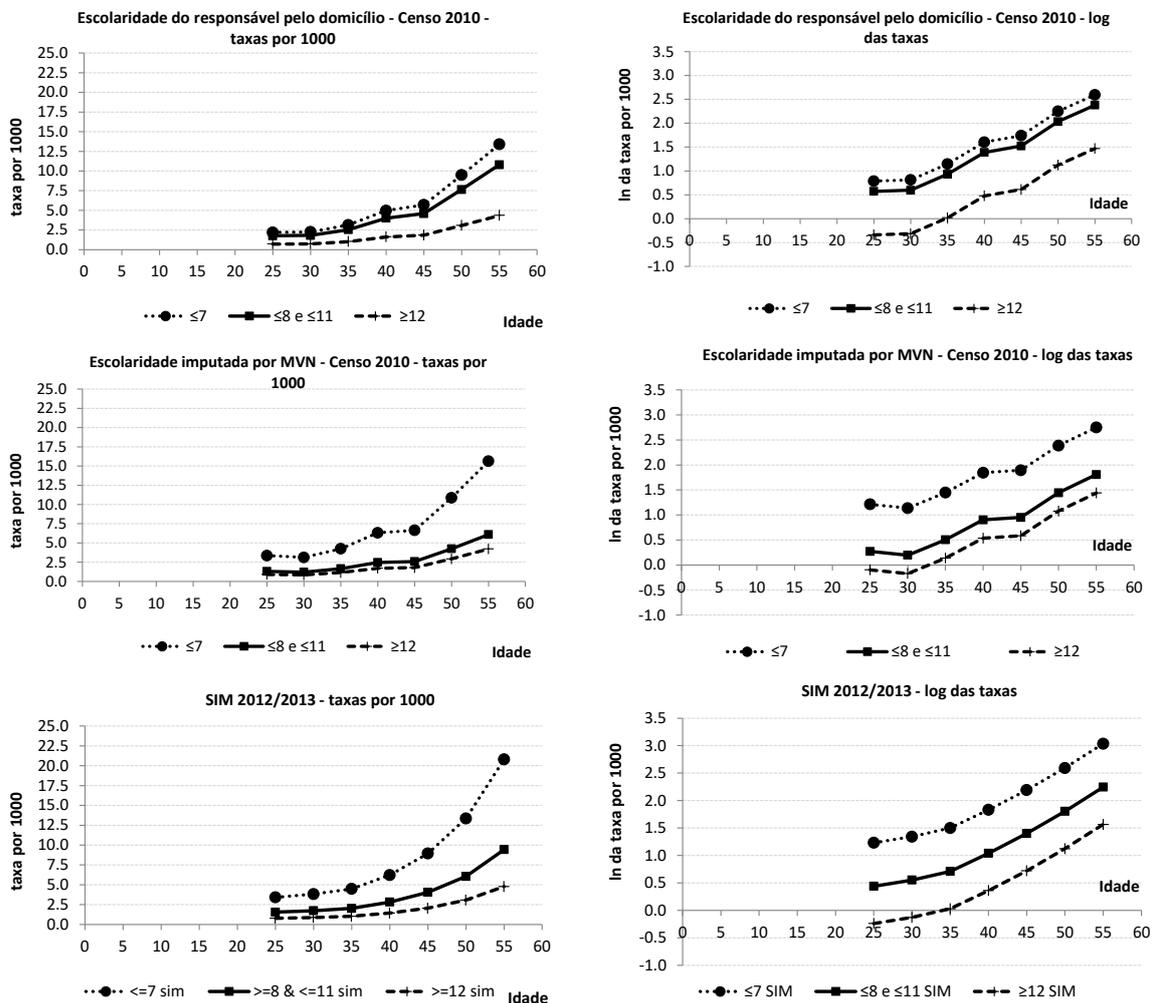
P-valor ≥ 0.13

Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo; MS/SIM microdados de 2012 e 2013

O padrão das taxas de mortalidade masculinas do Censo, mesmo quando se utilizam as escolaridades imputadas por MVN, se diferencia da observada no SIM. No caso da MVN, observa-se um gradiente educacional menos acentuado, com taxas para escolaridade média mais próximas das observadas no SIM, especialmente para o grupo de alta escolaridade (FIGURA 16).

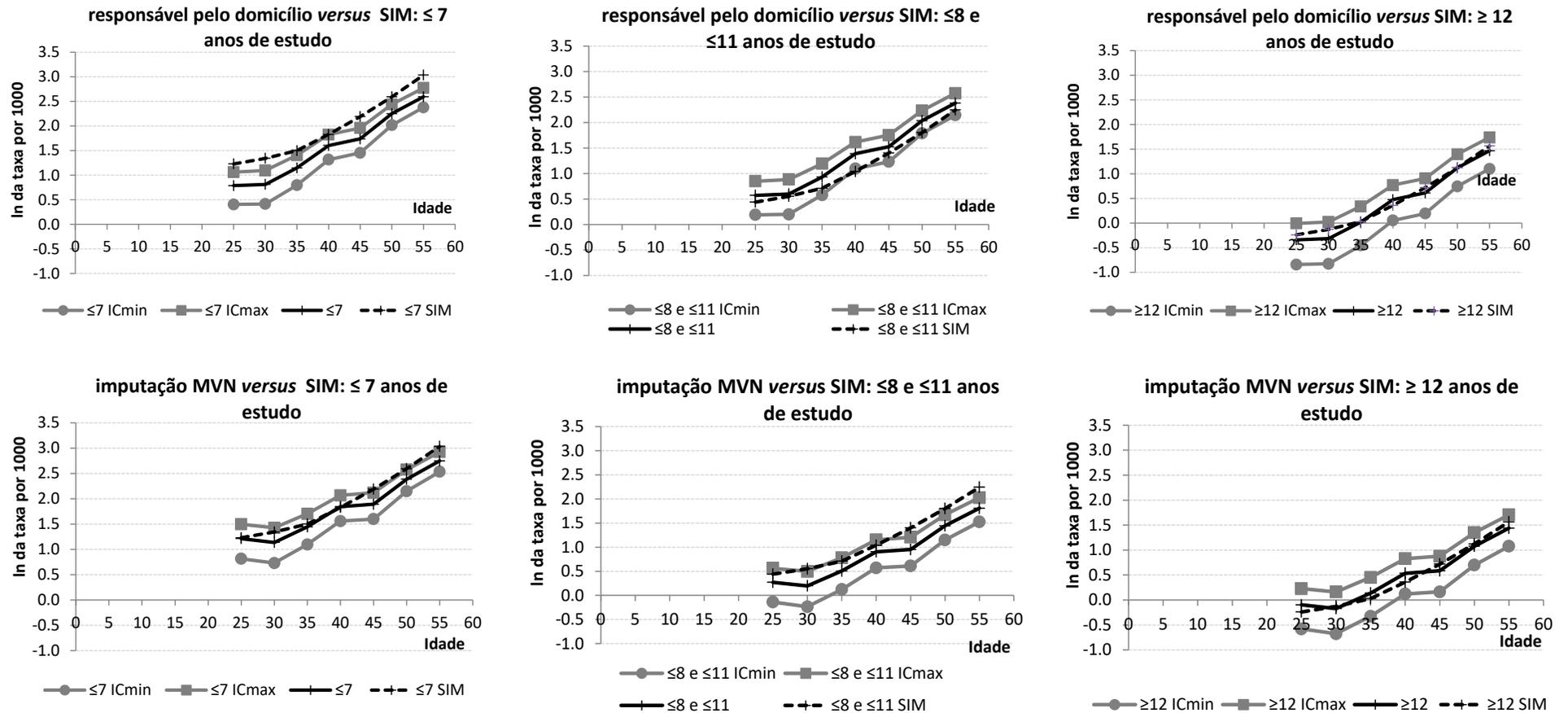
De qualquer forma, os diferenciais de escolaridade por imputação se apresentam mais próximos dos observadas no SIM, sobretudo para o grupo de menor escolaridade (FIGURA 17). As razões entre as taxas do Censo e SIM para a população masculina são bem mais expressivas quando a escolaridade do chefe é adotada (FIGURA 18).

FIGURA 16: Taxas de mortalidade (por mil) segundo escolaridade do responsável pelo domicílio, escolaridade imputada por MVN e observada no SIM, população masculina do município de São Paulo, 2010, 2012-2013



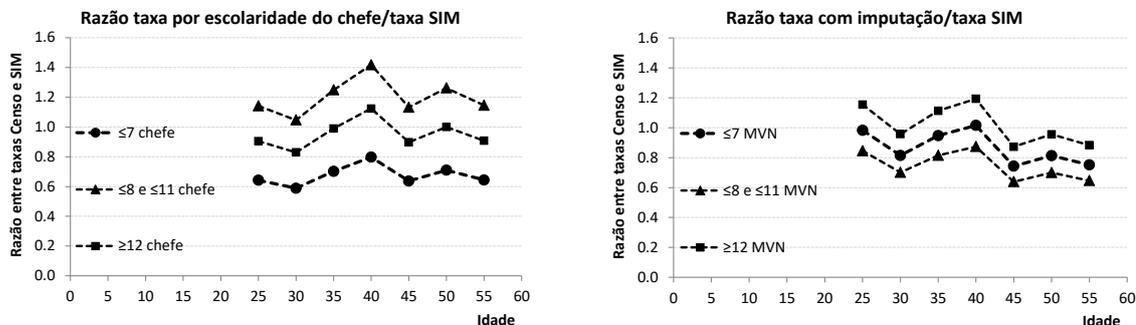
Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo; MS/SIM microdados de 2012 e 2013

FIGURA 17: Taxas de mortalidade (por mil) e intervalos de confiança estimados (IC) segundo escolaridade do responsável pelo domicílio, escolaridade imputada por MVN e observada no SIM, população masculina do município de São Paulo, 2010, 2012-2013



Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo; MS/SIM microdados de 2012 e 2013

FIGURA 18: Razão entre as taxas de mortalidade estimadas pelo Censo, por escolaridade do responsável pelo domicílio e escolaridade imputada por MVN, e as taxas do SIM, população masculina do município de São Paulo, 2010, 2012-2013



Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo; MS/SIM microdados de 2012 e 2013

Diante dos resultados discutidos na presente seção e na 6.1, os modelos de imputação por MVN representam melhor os diferenciais por escolaridade da mortalidade. A seção 6.3 apresenta os resultados considerando esses modelos e com correção dos óbitos pelos níveis observados no SIM.

6.3. Diferenciais de mortalidade por nível de escolaridade por idade e sexo: resultados com correção para sub e sobre enumeração dos óbitos

Nesta seção encontram-se as estimativas de mortalidade por nível de escolaridade tanto para o estado quanto para o município de São Paulo obtidas por MVN, com correção de nível, usando o SIM como referência de cobertura de óbitos. O objetivo é investigar os padrões do gradiente educacional por idade, sexo e localidade. A comparação entre estado e capital permite iniciar uma investigação sobre variações do gradiente educacional por região geográfica. Mesmo que a capital esteja inserida no estado, trata-se de uma região com dinâmicas demográfica e socioeconômica específicas que a difere dos demais municípios.

Os resultados dos modelos de Poisson indicam um gradiente educacional da mortalidade mais expressivo entre os homens e entre a população da capital. Os diferenciais de mortalidade entre os níveis médio e elevado de escolaridade são tímidos entre os homens e

inexistente entre as mulheres do estado de São Paulo, embora haja um acentuado diferencial do menor nível de escolaridade para os demais (TABELA 28 e TABELA 29).

TABELA 28: Resultado dos coeficientes e razões das taxas de incidência (IRR) dos modelos de Poisson, com níveis de mortalidade ajustados por idade sexo e educação, população feminina do estado e do município de São Paulo, 2010, 2012-2013

Variáveis	Escolaridade imputada MNV UF SP					Escolaridade imputada MNV município de SP				
	Coef	IRR	Erro Padrão (coef)	Intervalo de confiança 95% (coef)		Coef	IRR	Erro Padrão (coef)	Intervalo de confiança 95% (coef)	
idgrup2										
30 a 34	0.6366	1.8901	0.1358	0.3706	0.9027 *	0.3549	1.4260	0.3564	-0.3436	1.0533
35 a 39	0.9721	2.6434	0.1298	0.7176	1.2265 *	0.4630	1.5889	0.3535	-0.2299	1.1559
40 a 45	1.1734	3.2331	0.1270	0.9246	1.4223 *	1.0005	2.7197	0.3261	0.3614	1.6396 *
45 a 49	1.5409	4.6688	0.1228	1.3003	1.7815 *	1.4213	4.1423	0.3110	0.8118	2.0308 *
50 a 54	1.8871	6.6004	0.1204	1.6511	2.1232 *	1.8195	6.1690	0.3019	1.2277	2.4113 *
55 a 59	2.3021	9.9950	0.1188	2.0692	2.5349 *	2.1120	8.2645	0.2994	1.5252	2.6987 *
educimp										
>=8 e <=11	-0.6669	0.5133	0.0516	-0.7680	-0.5659 *	-0.6277	0.5338	0.1360	-0.8943	-0.3611 *
>=12	-0.6242	0.5357	0.0665	-0.7546	-0.4938 *	-0.9947	0.3698	0.1781	-1.3438	-0.6457 *
_cons	-7.0802	0.0008	0.1146	-7.3048	-6.8556 *	-7.0039	0.0009	0.2873	-7.5671	-6.4408 *
ln(anopes)	1.0000	1.0000				1.0000	1.0000			

* P-valor ≤ 0.01

P-valor ≥ 0.19

Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo; MS/SIM microdados de 2012 e 2013

TABELA 29: Resultado dos coeficientes e razões das taxas de incidência (IRR) dos modelos de Poisson, com níveis de mortalidade ajustados por idade sexo e educação, população masculina do estado e do município de São Paulo, 2010, 2012-2013

Variáveis	Escolaridade imputada MNV UF SP					Escolaridade imputada MNV município de SP				
	Coef	IRR	Erro Padrão (coef)	Intervalo de confiança 95% (coef)		Coef	IRR	Erro Padrão (coef)	Intervalo de confiança 95% (coef)	
idgrup2										
30 a 34	0.1851	1.2034	0.0765	0.0352	0.3350 *	0.0754	1.0783	0.2178	-0.3516	0.5023
35 a 39	0.3095	1.3628	0.0753	0.1619	0.4571 *	0.2830	1.3271	0.2103	-0.1291	0.6951
40 a 45	0.4496	1.5677	0.0736	0.3052	0.5939 *	0.6192	1.8575	0.1987	0.2297	1.0087 *
45 a 49	0.9666	2.6290	0.0681	0.8331	1.1001 *	0.9088	2.4814	0.1927	0.5312	1.2865 *
50 a 54	1.1836	3.2661	0.0672	1.0520	1.3152 *	1.3486	3.8519	0.1833	0.9894	1.7078 *
55 a 59	1.4974	4.4699	0.0661	1.3679	1.6268 *	1.7603	5.8142	0.1788	1.4099	2.1107 *
educimp										
>=8 e <=11	-0.7093	0.4920	0.0366	-0.7810	-0.6375 *	-0.8285	0.4367	0.0981	-1.0207	-0.6363 *
>=12	-0.8845	0.4129	0.0542	-0.9908	-0.7782 *	-1.4491	0.2348	0.1424	-1.7281	-1.1700 *
_cons	-5.7017	0.0033	0.0601	-5.8195	-5.5840 *	-5.6978	0.0034	0.1650	-6.0212	-5.3743 *
ln(anopes)	1.0000	1.0000				1.0000	1.0000			

* P-valor ≤ 0.01

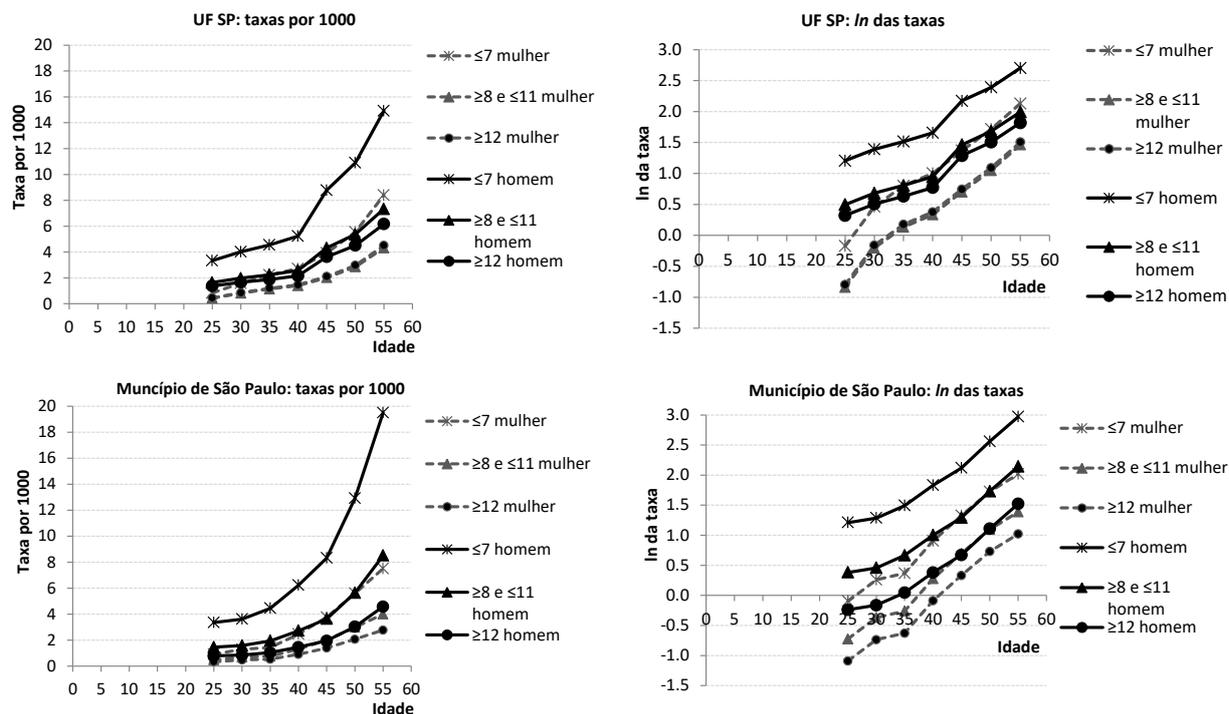
P-valor ≥ 0.17

Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo; MS/SIM microdados de 2012 e 2013

Para o município, há um claro gradiente educacional para ambos os sexos. Há uma já esperada sobremortalidade masculina. Os homens de baixa e média escolaridade apresentam curvas de mortalidade por idade que se aproximam, respectivamente, às de média e alta escolaridade femininas (FIGURA 19). Os diferenciais por sexo decrescem com a idade e com o nível de escolaridade (FIGURA 20). Para o município de São Paulo, os diferenciais por idade são menores e o gradiente da educação mais marcado. As estimativas das taxas, erro padrão e intervalos de confiança são apresentadas entre a TAB A 1 a TAB A 4 do Apêndice A.

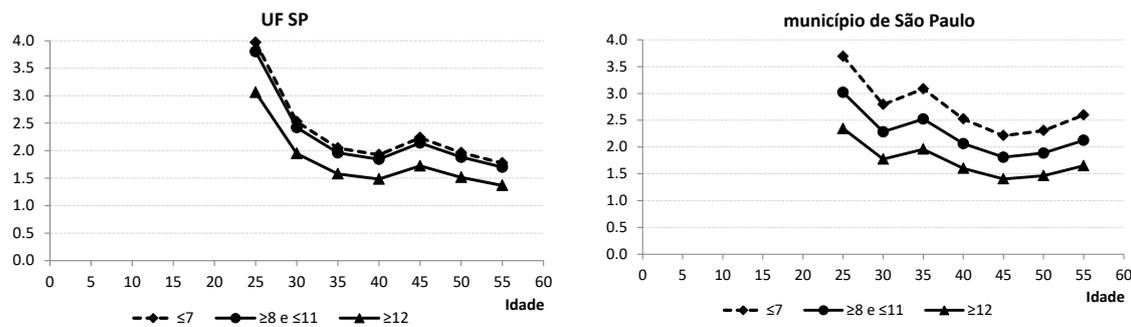
Conforme observados nos modelos de Poisson, além da sobremortalidade em relação às mulheres, os homens também apresentam um diferencial por escolaridade mais elevado. A mortalidade masculina do grupo de menor escolaridade é aproximadamente 4 vezes maior que a grupo de maior escolaridade para o município de São Paulo; para as mulheres a razão é de 2,7 (FIGURA 21).

FIGURA 19: Taxas de mortalidade por idade e sexo segundo níveis de escolaridade, estado e município de São Paulo, 2010



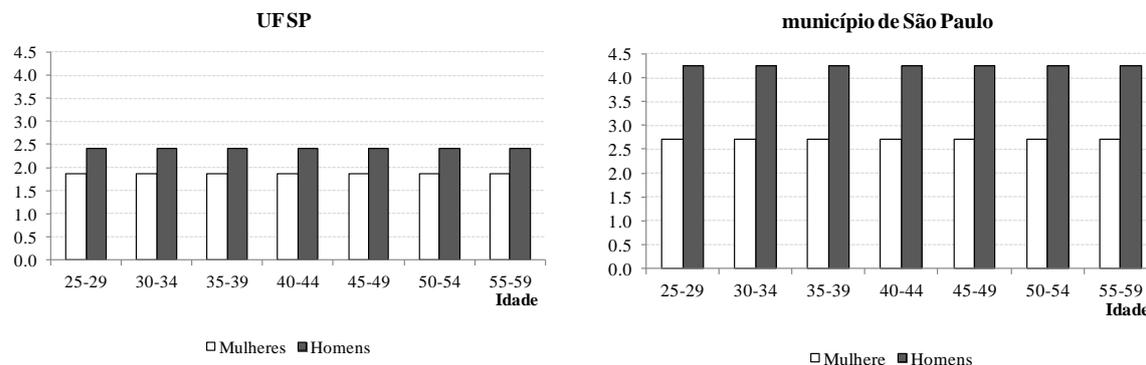
Fonte: IBGE - Censo Demográfico 2010, amostra estado e município de São Paulo

FIGURA 20: Razão entre as taxas de mortalidade masculinas e femininas por idade e sexo, segundo níveis de escolaridade, estado e município de São Paulo, 2010



Fonte: IBGE - Censo Demográfico 2010, amostra estado e município de São Paulo

FIGURA 21: Razão entre as taxas de mortalidade dos grupos de menor e maior escolaridade segundo idade e sexo, estado e município de São Paulo, 2010



Fonte: IBGE - Censo Demográfico 2010, amostra estado e município de São Paulo

6.4. Discussão

Um dos aspectos de interesse do trabalho foi comparar a mortalidade por escolaridade imputada, representando a educação dos indivíduos mortos, e a escolaridade dos responsáveis pelos domicílios. Se a escolaridade do chefe representasse bem a dos indivíduos, de fato ela poderia ser utilizada e nenhum esforço de estimar mortalidade dos indivíduos se justificaria. Essa seria uma hipótese razoável em um cenário em que a cobertura educacional entre gerações se mantivesse estável.

No entanto, o Brasil tem experimentado um aumento do acesso à educação desde a década de 1990, com a instituição da cobertura universal do ensino primário. Entre 1986 e 2008, a escolaridade média da população entre 7 e 25 anos passou de 4 para 6 anos de estudo. Isso não indica necessariamente uma mudança no processo de estratificação, uma vez que, no mesmo período, a escolaridade dos indivíduos manteve uma relação direta com a escolaridade dos chefes de domicílio. Há um sinal de convergência para séries menos elevadas, mas para séries mais elevadas uma estratificação persiste (Rios-Neto & Guimarães, 2010).

Nota-se que a mortalidade por escolaridade do chefe não apresentou diferenças entre o grupo de menor e média escolaridade, mas apenas para o grupo de alta escolaridade em relação aos demais. Além disso, foram exatamente os grupos de baixa e média de escolaridade que se distanciaram mais da mortalidade do SIM, sobretudo no caso das

mulheres. O grupo de menor escolaridade tem suas taxas subestimadas enquanto o de média escolaridade sobre-estimada. Ao usar a escolaridade do chefe, o denominador das taxas para baixa escolaridade pode estar inflado, já que o aumento da cobertura educacional permite que muitos indivíduos de domicílios com chefes de baixa escolaridade tenham alcançado uma escolaridade maior que a deles. Entre os indivíduos com ensino superior, as taxas estimadas foram muito semelhantes às obtidas com imputação. Silva et al (2016), usando dados do censo para estimar diferenciais de mortalidade por escolaridade do chefe do domicílio, reconhecem que a existência de algum grau de mobilidade intergeracional de educação poderia subestimar diferenciais de mortalidade por nível educacional.

Diante deste contexto, a escolaridade imputada se despointa como mais adequada para estimar diferenciais educacionais da mortalidade. Em resumo, existe um gradiente educacional da mortalidade, sendo os diferenciais maiores entre os homens.

Gomes *et al* (2013; 2015) também encontram uma associação negativa entre nível de escolaridade e mortalidade para pessoas de 60 anos para o município de São Paulo. Em outro estudo, os autores evidenciam um efeito da escolaridade maior para os homens (Gomes *et al*, 2013). No trabalho de Perez (2010) e Perez e Turra (2008), em que são estimadas taxas de mortalidade feminina, são encontrados diferenciais expressivos por escolaridade para idades acima de 15 e acima de 40 anos. Entretanto, a comparação da magnitude da mortalidade é dificultada por terem sido adotados intervalos de idade e escolaridade diferentes ou formas diferentes de medição da mortalidade entre os trabalhos conhecidos para o Brasil.

Contrariamente, Silva *et al* (2016) medem as probabilidades de morte e expectativa de vida para a população entre 15 e 60 anos por escolaridade do chefe do domicílio e encontra um gradiente educacional de expectativa de vida mais acentuado entre a população feminina.

No trabalho de Sandoval e Turra (2015), são estimados diferenciais de mortalidade por escolaridade para o Chile para 2001-2003, com medidas e intervalos de idade comparáveis com os adotados neste trabalho. Os grupos de escolaridade adotados são ligeiramente distintos: ≥ 8 , ≥ 9 e ≤ 12 , ≥ 13 anos de escolaridade. Como exemplo, os homens entre 55-59 anos apresentam taxas de 13,1 (baixa escolaridade), 8,5 (média) e 3,7 por mil (alta escolaridade) enquanto em São Paulo elas são, respectivamente, 14,9; 7,3 e 6,17. Embora

essas populações sejam diferentes, sendo o Chile um país latino americano em desenvolvimento suas estimativas podem ser uma fonte para detectar taxas irrealistas. Os diferenciais entre as duas populações diminuem quando se considera as estimativas para o município de São Paulo. Para o Chile, no entanto, os diferenciais entre os grupos de menor e maior escolaridade são iguais entre os sexos, atingindo 3,7. No caso de São Paulo, sobretudo na capital, os diferenciais masculinos são mais expressivos que o das mulheres (4,2 e 2,7, respectivamente).

Outro ponto importante é o comportamento do diferencial de mortalidade por escolaridade entre os grupos etários. A partir da literatura apresentada no capítulo 2, era esperado um diferencial de mortalidade por escolaridade decrescente com idade. Isso é o que ocorre, por exemplo, no caso do Chile no referido trabalho de Sandoval e Turra (2015). Uma possível explicação seria que a exposição dos indivíduos à baixa condição socioeconômica nas idades jovens levaria a uma seleção dos mais saudáveis dentro do grupo. Em consequência, sua experiência de saúde nas idades mais avançadas se igualaria à dos indivíduos socialmente favorecidos (Lauderdale, 2001). Em uma abordagem epidemiológica, House et al (1994) indica que a associação entre os fatores de risco e as características socioeconômicas enfraquecem com idade. Nos Estados Unidos, uma explicação complementar seria a existência de programas públicos de saúde específico para idoso (*Medicare*), que minimizaria os efeitos dos fatores socioeconômicos (Preston & Elo, 1995).

No entanto, entre os adultos de 25 a 59 anos de São Paulo não foram encontradas variações dos diferenciais. Não sabemos como seria o comportamento a partir dos 60 anos ou em outras regiões do país. Essa pode ser uma especificidade do estado e município de São Paulo. Para explorar esses resultados torna-se necessário entender a forma com a transição da mortalidade tem ocorrido no estado e como as medidas de período podem influenciar os resultados. Lauderdale (2001) destaca a importância de se analisar as relações entre período, idade e coorte para entender o comportamento dos diferenciais de mortalidade por escolaridade. Além da transição da mortalidade existe a transição educacional no país. O aumento do acesso à escolaridade tem elevado o nível educacional das coortes mais jovens (Rios-Neto e Guimarães, 2010). Segundo Liberatos et al (1988), sob um cenário de mudança, deve-se pensar na possibilidade de variação do peso da escolaridade no *status* socioeconômico dos indivíduos. A relação da educação com a renda e com a ocupação

pode mudar entre gerações, de forma que um nível de educação elevado tenha significado social diferente em cada uma.

7. CONSIDERAÇÕES FINAIS

O objetivo principal da tese foi apresentar estimativas de mortalidade por níveis de escolaridade para o estado e o município de São Paulo usando dados do Censo. Para isso lançou-se mão de estratégias de imputação para obter escolaridade dos indivíduos que sofreram óbito. O objetivo da imputação não é, no entanto, estabelecer valores individuais para a escolaridade, mas construir uma distribuição válida para a escolaridade dos mortos. Os métodos aqui utilizados, IM e MVN com algoritmo EM, são destacados na literatura como os mais apropriados e com resultados superiores sobre os demais para imputação de dados faltantes. A análise de consistência e validação dos métodos apontou resultados satisfatórios, tendo o método de MVN se despontado como o melhor.

No entanto, isso não afasta a necessidade de se avançar nas estratégias de imputação e de avaliação das mesmas para o alcance de resultados melhores. Trata-se de uma sugestão de investigação inicial, com resultados que podem ser vistos como otimistas, uma vez que a escolaridade imputada apresentou avanços na análise em relação às informações já disponíveis, como a escolaridade dos responsáveis pelos domicílios, por exemplo. Mas para que se possa aplicar os métodos para todo o Brasil é necessário testá-los para as diferentes regiões e identificar como os resultados se comportam diante de diferentes populações, com diferentes características daquelas observadas para São Paulo.

No que tange à discussão dos diferenciais de mortalidade por escolaridade encontrados, é necessário avançar. Além de um aprofundamento dos aspectos teóricos explicativos existentes na literatura, é fundamental que se avance nas especificidades brasileira em relação à composição e transição da cobertura educacional e à estrutura de mortalidade por causas de morte. Conforme discutido em alguns pontos do referencial teórico, o modo como a mortalidade interage com os aspectos socioeconômicos individuais varia entre causas de morte. Como a estrutura de causas de morte depende da idade e do sexo, o conhecimento de sua distribuição com relação a essas variáveis é fundamental para o entendimento do gradiente educacional da mortalidade.

Outro aspecto importante ao se discutir variações regionais do gradiente educacional é o conhecimento da estrutura de oferta de serviços de saúde em cada localidade. As abordagens econômica e psicossocial apontam esse aspecto como um importante mediador entre características socioeconômicas e saúde.

A despeito de todas as limitações, o trabalho trouxe uma contribuição em relação à discussão do uso do Censo para estudos de diferenciais de mortalidade. Embora exista uma expectativa de melhoria da cobertura e qualidade das informações de escolaridade do SIM diante das mudanças recentes na declaração de óbito, o Censo pode ser uma importante fonte para estudos de diferenciais socioeconômicos de mortalidade. Além da escolaridade, a imputação pode ser vista como um método para contornar a escassez de informações individuais dos mortos. As chances de êxito com esses métodos no Censo é grande devido ao tamanho da amostra. Além disso, ele oferece um número considerável de características do nível do domicílio e indivíduos sobreviventes que o compõem, que podem complementar os estudos sobre diferenciais de mortalidade por escolaridade.

REFERÊNCIAS

- Adler N. E, Boyce T., Chesney M. A, Cohen S., Folkman. S, Kahn R.L., Syme S.L. Adler. Socioeconomic status and health. The challenge of the gradient. **Am Psychol.** 49(1):15-24, 1994.
- Allison P. D. Missing Data. Millsap R. E., In: Maydeu-Olivares A. **The SAGE Handbook of Quantitative Methods in Psychology.** SAGE Publications. p 72-89, 2009
- Antonovsky A. Social class life expectancy and overall mortality. **Milbank Mem. Fund Q.** 45(2, pt. 1):31–73, 1967.
- Barros R., Sawyer D., Unequal opportunity to survive, education and regional disparities in Brazil. Ipea: Brasília. Texto de Discussão, 1993.
- Becker, G. S. A theory of the allocation of time. **Economic Journal**, New York, 75(299), p. 493–517, Sept. 1965.
- Berhane Y., Hogberg U., Byass P., Wall S. Gender, literacy, and survival among ethiopian adults, 1987–96. **Bulletin of the World Health Organization**, New York, 80(9), p. 714-720, Sept. 2002.
- Blane D., Smith G. D, Bartley M. Social selection: what does it contribute to social class differences in health? **Sociology of Health & Illness.** 15(1), 1993.
- Bloor M.; Samphier M.; Prior L. Artefact explanations of inequalities in health: an assessment of the evidence. **Sociology of Health an Illness**, 15(1), 1993.
- Brasil. Ministério da Saúde. Monitoramento da acurácia dos sistemas de informações sobre mortalidade e nascidos vivos. In: Anais 3a Expoepi – Mostra Nacional de Experiências Bem-sucedidas em Epidemiologia, Prevenção e Controle de Doenças. Brasília: Ministério da Saúde; 2004. p. 173-80.
- Caldwell, J. C. Mass education as a determinant of mortality decline. In: Caldwell, J. C.; Santow, G. (ed.). Selected readings in the cultural, social and behavioural determinants of health. Canberra: The Australian National University: Health Transition Centre. 1989. p.101-109.
- Camargos V. P., Cesar C. C.; Caiaffa W. T., Xavier C. C., Proietti F. A. Imputação múltipla e análise de casos completos em modelos de regressão logística: uma avaliação prática do impacto das perdas em covariáveis. **Cad. Saúde Pública**, Rio de Janeiro, 27(12):2299-2313, dez, 2011.
- Case A., Deaton A. Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. **Proceedings of the National Academy of Sciences (PNAS)**, November 2, 2015.

Clogg C. C., Donald R. B., Schenker N., Schultz B., Weidman L. Multiple Imputation of Industry and Occupation Codes in Census Public-Use Samples Using Bayesian Logistic Regression. **Journal of the American Statistical Association**. 86(413): 68-78, 1991.

Coleman, J. S. 1964. **Introduction to Mathematical Sociology**. New York: Free Press.

Collins, L. M., Schafer, J. L., & Kam, C. M. A comparison of inclusive and restrictive strategies in modern missing-data procedures. **Psychological Methods**, 6, 330–351, 2001.

Crimmins E. Demography: the 30 past years, the present, and the future. **Demography**. 30(4): 579-591, 1993.

Cutler, D. M.; Lleras-Muney A. Education and health: evaluating theories and evidence". In House, J. S., Schoeni R. F., Kaplan, G. A., Pollack H. (eds). **The Health effects of social and economic policy**. New York: Russel Sage Foundation, 2007.

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

Denisova I. Mortality in Russia: Microanalysis. Center for Economic and Financial Research and New Economic School. Working Paper 128. March, 2009.

DHSS. Department of Health and Social Security. Inequalities in Health: Report of a Research Working Group (The Black Report). London: Department of Health and Social Security, 1980.

Duarte, E. C., Schneider, M. C., Paes-Sousa, R., Silva, J. B., Castillo-Salgado, C. Expectativa de vida ao nascer e mortalidade no Brasil em 1999: análise exploratória dos diferenciais regionais. **Revista Panamericana de Salud Publica**, Washinton, v. 12, n. 6, p. 436-444, 2002.

Duleep, H. O. Measuring mortality differentials over time. **Demography**, 26(2): 345-351, 1989.

Elo, I.T. and S.H. Preston. Educational Differentials in Mortality: United States, 1979–1985. **Social Science and Medicine**. 42:47–57, 1996.

Elo I T. Social Class Differentials in Health and Mortality: Patterns and Explanations in Comparative Perspective. **Annual Rev. Sociol.**, 35, 2009, pp 553-572.

Elo I. T., Martikainen P., Smith E. K. P. Socioeconomic differentials in mortality in Finland and the United States: the role of education and income. **Eur J Population**, 22:179–203, 2006.

Enders C. K. **Applied Missing Data Analysis**. The Guilford Press, New York, 2010.

Fuchs, V. Who we shall live? Helath, economics, and social choice. New York. 1974.

Garcia, L. P.; Santana L. R. Evolução das desigualdades socioeconômicas na mortalidade infantil no Brasil, 1993-2008. **Ciênc. saúde coletiva**[online]. 16(9): 3717-3728, 2011.

Goldman D.; Smith J. P. The Increasing Value of Education to Health. **Soc Sci Med.** May 72(10): 1728–1737, 2011.

Gomes M. M. F., Turra C., Fígoli M. G. B., Duarte Y. A. O., Lebrão M. L. Associação entre mortalidade e estado marital: uma análise para idosos residentes no Município de São Paulo, Brasil, Estudo SABE, 2000 e 2006. **Cad. Saúde Pública**, Rio de Janeiro, 29(3):566-578, mar, 2013.

Gomes M. M. F., Turra C., Fígoli M. G. B., Duarte Y. A. O., Lebrão M. L. Passado e presente: condições de vida na infância e mortalidade de idosos. **Rev Saúde Pública** 2015. pp 49-93.

Guedes G. R.; Siviero P. C. L.; Queiroz B. L.; Machado C. J. Approximating the educational differences in mortality: demographic indirect techniques. **Cad. Saúde Colet.** Rio de Janeiro, 19(2) 240-3, 2011.

Graham J. W., Hofer S. M., Donaldson S. I. Mackinnson D. P., Schafer J. L. Analysis with Missing Data in Prevention Research. In Bryant K. J., Windle M., West S. G. **The science of prevention: methodological advances from alcohol and substance abuse research.** Washington DC: American Association, 1997. p 325-366.

Grossman, M. On the concept of health capital and the demand for health. **The Journal of Political Economy**, Chicago, 80(2): 223-55, 1972.

Grossman, M. The relationship between health and schooling. **Eastern Economic Journal.** 34(3): 281–292, 2008.

Hayward, M. D., Hummer R. A., Sasson I. Trends and Group Differences in the Association between Educational Attainment and U.S. Adult Mortality: Implications for Understanding Education's Causal Influence. **Social Science & Medicine.** 127: 8-18, 2015.

Heitjan D. F., Rubin D. B. Inference from Coarse Data Via Multiple Imputation with Application to Age Heaping. **Journal of the American Statistical Association.** 85(410): 304-314, 1990.

Heitjan D. F., Rubin, D. B. Ignorability and coarse data. **Annals of Statistics.** 19(4): 2244–2253, 1991.

House J. S., Landis K. R., Umberson D. Social relationship and health. **Science**, 241: 540-545, 1988.

House J. S., Kessler R. C., Herzog R. Age, socioeconomic status and health. **The Milbank Quartely.** 68(3): 383-411, 1990.

House et al 1994.. (ver lauderdale)

Huisman et al Socioeconomic inequalities in mortality among elderly people in 11 European populations. **J Epidemiol Community Health**. 58:468–475, 2004.

Hummer R. A., Rogers R. G., Eberstein I. W. Sociodemographic differentials in adult mortality: a review of analytic approaches. **Population and Development Review**. 24(3): 553-578, 1998.

Hurt L. S., Ronsmans C., Saha S. Effects of education and other socioeconomic factors on middle age mortality in rural Bangladesh. **J. Epidemiol. Community Health** 58:315–20, 2004.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Observações sobre a evolução da mortalidade no Brasil: o passado, o presente e perspectivas**. Rio de Janeiro 2010a.

IBGE. Instituto Brasileiro De Geografia E Estatística. **Tábuas Abreviadas de Mortalidade por Sexo e Idade: Brasil, Grandes Regiões e Unidades da Federação - 2010**. Rio de Janeiro: 2010b.

Illeys R. Occupational class, selection and the production of inequalities in health. **Quartely Journal of Social Affairs**, 2(2) 1986.

Kitagawa E. M., Hauser P. M. **Differential Mortality in the United States: A Study in Socioeconomic Epidemiology**. Harvard University Press, Cambridge, MA, 1973.

Koch et al. Desigualdad educacional y socioeconómica como determinante de mortalidad en Chile: análisis de sobrevivencia en la cohorte del proyecto San Francisco. **Rev Méd Chile** 2007; 135: 1370-1379, 2007.

Krieger N., Williams D. R, Moss N. E. Measuring social class in us public health research: concepts, methodologies, and guidelines. **Annu. Rev. Public Health**. 18:341–78, 1997.

Kunst A. E, Mackenbach J. P. The size of mortality differences associated with educational level in nine industrialized countries. **Am. J. Public Health**. 84:932–37, 1994.

Lauderdale D. S. Education and survival: birth cohort, period, and age effects. **Demography**. 38(4): 551-561, 2001.

Leinsalu M, Vagero D, Kunst AE. Estonia 1989–2000: enormous increase in mortality differences by education. **Int. J. Epidemiol**. 32(6):1081–87, 2003.

Liang et al. Socioeconomic Gradient in Old Age Mortality in Wuhan, China. **Journal of Gerontology: Social Sciences**, 55B(4): S222–S233, 2000.

Liberatos P., Link B. G., Kelsey J. L. The measurement of social class in epidemiology. **Epidemiologic Review**. 10(1): 87-121, 1988.

Lillard L. A.; Waite L. J. 'Til death do us part': marital disruption and mortality. **American Journal of Sociology**. 100: 1131-56, 1995.

Lima L. C. **A natureza da relação entre escolaridade materna e mortalidade infantil e na infância no Brasil**. 2013, ??f. Tese (Doutorado em Demografia) – Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais, Belo Horizonte, 2013.

Lima E. E. C., Queiroz B. L. Método de estimação de grau de cobertura em pequenas áreas: uma aplicação nas microrregiões mineiras. **Cad. Saúde Colet**. Rio de Janeiro, 22 (4): 409-18, 2014a.

Little R. J. A., Rubin D. B. **Statistical Analysis with Missing Data**. 2 ed. Hoboken, NJ: Wiley, 2002.

Little T. D., Rhemtulla M. Planned missing data designs for developmental researchers. *Child Development Perspectives*. 7(4): 199-204, 2013.

Lynch J. Income inequality and health: expanding the debate. **Social Science & Medicine**. 51: 1001-1005, 2000.

Lynch J. Smith G. D., Kaplan G. A., House J. S. Income inequality and mortality: importance to health of individual income, psychosocial environment, or material conditions. **BMJ**. 320: 1200-04, 2000.

Mackenbach JP, Stirbu I, Roskam A-JR, Schaap MM, Menvielle G, *et al.* 2008. Socioeconomic inequalities in health in 22 European countries. **N. Engl. J. Med**. 358:2468–81, 2008.

Manzelli H. Educational attainment and adult mortality differentials in Argentina. **Revista Latinoamericana de Población**. Vol 14, 2014.

Meara E., Richards S., Cutler D. The Gap Gets Bigger: Changes in Mortality and Life expectancy by Education, 1981–2000. **Health Aff (Millwood)**. 27(2): 350–360, 2008.

Mistler, S. A., & Enders, C. K. Planned missing data designs for developmental research. In B. Laursen, T. D. Little, & N. A. Card (Eds.), **Handbook of developmental research methods** (pp. 742–754). New York, NY: Guilford Press. 2012.

Montez J. K., Hayward M. D, Brown D. C, Hummer R. A. Why Is the Educational Gradient of Mortality Steeper for Men? *Journal of Gerontology: Social Sciences*. 64B(5): 625-634, 2009.

Montez, J. K., Hummer, R. A., Hayward, M. D., Woo, H., and Rogers, R. G. Trends in the Educational Gradient of U.S. Adult Mortality from 1986 through 2006 by Race, Gender, and Age Group. **Research on Aging**. 33(2):145-71, 2011.

Montez, J. K.; Hummer, R. A.; Hayward M. D. Educational Attainment and adult mortality in United States: a systematic analysis of functional form. **Demography**. 49: 315-336, 2012.

MS - Ministério da Saúde. Secretaria de Vigilância em Saúde. Portaria nº 116, de 11 de fevereiro de 2009.

MS - Ministério da Saúde. Coordenação Geral de Informações e Análise Epidemiológica - CGIAE Sistema de Informações sobre Mortalidade – SIM: Consolidação da base de dados de 2011. Available from: http://tabnet.datasus.gov.br/cgi/sim/Consolida_Sim_2011.pdf.

NCHS. National Center for Health Statistics. Vital Statistics - Special Reports, vol 53(1), Washington DC., 1961.

NCHS. National Center for Health Statistics. Vital and Health Statistics, series 4(29), 1968.

NCHS. National Center for Health Statistics. Vital and Health Statistics, series 2(30), 1969.

NCHS. National Center for Health Statistics. **Highlights from Wave 1 of the National Survey of Personal Health Practices and Consequences: United States, 1979**. Vital and Health Statistics, Series 15(1), 1981.

Olshansky S. J. *et al.* A Potential Decline in Life Expectancy in the United States in the 21st Century. **New England Journal of Medicine**. 352(11): 1138-1145, 2005.

Pampel F. C, Krueger P. M., Denney J. T. Socioeconomic disparities in health behaviors. **Annu. Rev. Sociol.** 36:349–70, 2010.

Pamuk E. R. Social class inequality in mortality from 1921 to 1972 in England and Wales. **Popul. Stud.** 39(1):17–31, 1985.

Pappas G., Queen S., Hadden W. and Fisher G. The increasing disparity on mortality between socioeconomic groups in the United States, 1960 and 1986. **N. Engl. J. Med.** 329, 103, 1993.

Pereira, F. N. A. Diferenciais de mortalidade jovem no Brasil - a importância dos fatores socioeconômicos dos domicílios e das condições de vida nos municípios e UFs. 2014, 122f. Dissertação (Mestrado em Demografia) – Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais, Belo Horizonte, 2014.

Pereira F. N. A., Queiroz B. L. Diferenciais de mortalidade jovem no Brasil: a importância dos fatores socioeconômicos dos domicílios e das condições de vida nas localidades. 2016. Cadernos de Saúde Pública. No prelo.

Perez E. R., Turra C. M. Desigualdade social na mortalidade no Brasil: diferenciais por escolaridade entre mulheres adultas. **Anais do XVI Encontro da ABEP**. Caxambu, Brasil, 2008.

Perez E. R. **Estimativas de mortalidade adulta feminina por nível de escolaridade no Brasil**. 2010, 146f. Tese (Doutorado em Demografia) – Centro de Desenvolvimento e

Planejamento Regional, Universidade Federal de Minas Gerais, Belo Horizonte, 2010.

Preston S. H., Taubman P. Socioeconomic differences in adult mortality and health status. In: Martin L., Preston S. H. **The Demography of Aging**. pp. 279-318. National Academy Press, Washington, DC, 1994.

Preston S. H., Elo I. T. Are educational differentials in adult mortality increasing in the United States? **Journal of Aging Health**. 7: 476-496, 1995.

Preston S. H., Wang H. Sex mortality differences in the United States: the role of cohort smoking patterns. **Demography**. 43(4): 631-646, 2006.

Queiroz. Estimativas do Grau de Cobertura e da Esperança de Vida para as Unidades da Federação no Brasil entre 2000 e 2010. Anais XVIII Encontro Nacional de Estudos Populacionais: transformações na população brasileira: complexidades, incertezas e perspectivas, 19 a 23 de novembro de 2012 / Águas de Lindóia, SP: ABEP, 2012.

Queiroz, B. L.; Sawyer, D. O. T. O que os dados de mortalidade do Censo de 2010 podem nos dizer? **Revista Brasileira de Estudos de População**, vol.29, n.2, pp. 225-238, 2012.

Rios-Neto, E., Guimarães, R. R. M, The demography of education in Brazil: inequality of educational opportunities based on Grade Progression Probability (1986-2008). **Vienna Yearbook of Population Research**, vol 8, 2010, pp. 283-312.

Rofman R. Diferenciales de mortalidad adulta en Argentina. Centro de Estudios de Población, CENEP, Buenos Aires, 2004.

Roger R. G. Marriage, sex and mortality. **Journal of Marriage and the family**. 57(2): 515-526, 1995.

Roger R. G. The effects of family composition, health, and social support linkages on mortality. **Journal of Health and Social Behavior**. 37(4): 326-338, 1996.

Rogers, R. G., Hummer R. A., Nam C. B. **Living and Dying in the USA: Behavioral, Health, and Social Differences of Adult Mortality**. Academic Press, 1999b. Cap 7. The effects of basic socioeconomic factors on mortality, p. 115-139.

Rogers R. G, *et al.* Social, behavioral, and biological factors, and sex differences in mortality. **Demography**, 47(3): 555-578, 2010.

Rogot E., P. D., Johnson N. J. Life Expectancy by Employment Status, Income, and Education in the National Longitudinal Mortality Study. Public Health Reports. 107(4): 457-61, 1992.

Rosenzweig M. R., Schultz T. P., Estimating a household production function: heterogeneity, the demand for health inputs, and their effects on birthweight. **Journal of Political Economy**, 91(5):723-46, 1983.

Rosero-Bixby L. Adult Mortality Decline in Costa Rica. In: Chackiel, J., Ruzicka, L.; Timaeus, I. (Ed.). **Adult mortality in Latin America**. Oxford: Clarendon Press, 1996. p. 166-196.

Rosero-Bixby, L.; Dow, W. H. Surprising SES Gradients in mortality, health, and biomarkers in a Latin American population of adults. **Journal of Gerontology B: Psychological and Social Science**, Washington, v. 64, n. 1, p. 105-117, Jan. 2009.

Ross C. E., Wu C. L. Education, age and the cumulative advantage in health. **Journal of Health and Social Behavior**. 37(1): 104-120, 1996.

Ross C. E.; Masters R. K.; Hummer R. R. A. Educational and the gender gaps in health and mortality. **Demography**, 49: 1157-1183, 2012.

Rubin D. B., Inference and missing data. **Biometrika**, 63(3): 581-592, 1976.

Rubin D. B., Multiple imputations in sample surveys. **Proceedings Survey Research Methods Section, American Statistical Association**: 20-34, 1978.

Rubin D. B.; Schenker N. Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. **Journal of the American Statistical Association**, 81: 366-374, 1986.

Rubin D. B., Schenker N. Multiple imputation in health care databases: an overview and some applications. **Statistics in Medicine**. 10: 585-598, 1991.

Rubin D. B. Multiple imputation after 18+ years. **Journal of the American Statistical Association**. 91(434): 473-489, 1996.

Rubin D. B. **Multiple imputation for nonresponse in surveys**. 2nd ed. New York: Wiley, 2004.

Saikia N., Bhat P. N. M. Factors Affecting Adult Mortality in India: An Analysis of National Family Health Surveys of 1992-93 and 1998-99 (NFHS I and II) **Demography India**, 37(2), pp. 291-302, 2008.

Sandoval M. H., Turra C. M. El gradiente educativo en la mortalidad adulta en Chile. **Revista Latinoamericana de Población**. 17(9): 7-35, 2015.

Sasson I. Trends in Life Expectancy and Lifespan Variation by Educational Attainment: United States, 1990–2010. **Demography**, 53:269–293, 2016.

Sastry, N. What explains rural-urban differentials in child mortality in Brazil? **Social Science and Medicine**, Oxford, v. 44, n. 7, p. 989–1002, Oct. 1997.

Sawyer, D. O., Soares. E. S. Child mortality in different contexts in Brazil: variations in the effects of the socio-economic variables. In: **Infant and child mortality in the third world**. Paris: WHO, 1982. p. 145-160.

Schafer J. L. **Analysis of Incomplete Multivariate Data**. Chapman & Hall: London, 1997.

Schafer J. L., Multiple imputation: a primer. **Statistical Methods in Medical Research**. 8: 3-15, 1999.

Schafer J. L., Graham J. W. Missing data: our view of the state of the art. **Psychological Methods**, 7: 147-177, 2002.

Shkolnikov et al 1998. Educational level and adult mortality in Russia: an analysis of routine data 1979 to 1994. **Soc. Sci. Med.**, 47(3): 357-369, 1998.

Shkolnikov V. M., Andreev E. M., Jasilionis D., Leinsalu M., Antonova O. I., McKee M. The changing relation between education and life expectancy in Central and Eastern Europe in the 1990s. **J. Epidemiol. Community Health**. 60:875–81, 2006.

Sickles. R. C., Taubman P. Mortality and morbidity among adults and t he elderly. In: Rosenzweig M., Stark O. **Handbook of Population and Family Economics**. Elsevier: 1997.

Silva L. E. Diferenciais de mortalidade adulta por nível de escolaridade no Brasil e Regiões. 2014, 122f. Dissertação (Mestrado em Demografia). Universidade Federal do Rio Grande do Norte, Natal, 2014.

Silva L. E., Freire F. H. M. A, Pereira R. H. M. Diferenciais de mortalidade por escolaridade da população adulta brasileira. **Cad. Saúde Pública, Rio de Janeiro**, 32(4): 1-12, 2016.

Simões, C. C. S. **Brasil: estimativas da mortalidade infantil por micro-regiões e municípios**. Brasília: Ministério da Saúde, 1999.

Simões, C. C. S. **Perfis de saúde e de mortalidade no Brasil: uma análise de seus condicionantes em grupos populacionais específicos**. Brasília: Organização Pan-Americana da Saúde, 2002, 141p.

Singh-Manoux A, Marmot M. Role of socialization in explaining social inequalities in health. **Social Science & Medicine**. 60: 2129–2133, 2005.

StataCorp. Sata 14.0. College Station, Texas, 2015a.

StataCorp. Stata Multiple Imputation. Reference Manual. Release 14. Stata Press, 2015b.

Subramanian, S. V. *et al*. The mortality divide in India: the differential contributions of gender, caste, and standard of living across the life course. **American Journal of Public Health**. 96(5), p.818-825, 2006.

Townsend, P.; Davidson, N. (Ed.). **Inequalities in health: the black report & the health divide**. London: Penguin Books, 1988.

US Public Health Services. **US Surgeon General's Advisory Committee on Smoking and Health, Smoking and Health**. Washington DC, 1964.

Valkonen T. Adult mortality and level of education: a comparison of six countries. In: Fox J. **Health Inequalities in European Countries**. pp. 142-162. Gower, Aldershot, 1989.

Van Buren S., Boshuizen H. C., Knook D. L. Multiple imputation of missing blood pressure covariates in survival analysis. **Statistics in Medicine**. 18: 681-694, 1999.

Walque D., Filmer D. Trends and socioeconomic gradients in adult mortality around the developing world. **Population and Development Review**, 39(1): 1-29, 2013.

West P. Rethink the health selection explanation for health inequalities. **Social Science & Medicine**, 32, 1991.

Williams, D. F. Socioeconomic differentials in health: a review and redirection. **Social Psychology Quarterly**, 53(2): 81-99, 1990.

Wood, C. H.; CARVALHO, J. A. M. **The demography of inequality in Brazil**. London: Cambridge University Press, 1988.

Yunes J., Ronchezel V. S. C. Evolução da mortalidade geral, infantil e proporcional no Brasil. **Rev. Saúde Públ.**, São Paulo, 8(supl): 3-48, 1974.

APÊNDICE A

TAB A 1: Taxas de mortalidade específicas de mortalidade por idade estimadas pelos coeficientes da regressão de Poisson com base nas imputações, população feminina, estado de São Paulo, 2010

Escolaridade	Idade	Taxa de mortalidade (por 1000)	Erro padrão	Intervalo de confiança 95%	
≤ 7	25 a 29	0.84	0.10	0.65	1.03
	30 a 34	1.59	0.13	1.33	1.85
	35 a 39	2.22	0.16	1.91	2.54
	40 a 45	2.72	0.18	2.37	3.07
	45 a 49	3.93	0.22	3.50	4.36
	50 a 54	5.55	0.27	5.03	6.08
	55 a 59	8.41	0.35	7.72	9.10
≥ 8 e ≤ 11	25 a 29	0.43	0.05	0.34	0.53
	30 a 34	0.82	0.07	0.68	0.95
	35 a 39	1.14	0.09	0.97	1.31
	40 a 45	1.40	0.10	1.20	1.59
	45 a 49	2.02	0.13	1.76	2.27
	50 a 54	2.85	0.17	2.51	3.19
	55 a 59	4.32	0.25	3.83	4.81
≥ 12	25 a 29	0.45	0.05	0.34	0.56
	30 a 34	0.85	0.08	0.69	1.01
	35 a 39	1.19	0.10	0.99	1.40
	40 a 45	1.46	0.12	1.22	1.70
	45 a 49	2.10	0.16	1.79	2.42
	50 a 54	2.98	0.22	2.55	3.40
	55 a 59	4.51	0.32	3.88	5.13

Fonte: Com base nos dados do Censo 2010, amostra São Paulo, e correções com base no SIM/MS

TAB A 2: Taxas de mortalidade específicas de mortalidade por idade estimadas pelos coeficientes da regressão de Poisson com base nas imputações, população feminina, município de São Paulo, 2010

Escolaridade	Idade	Taxa de mortalidade (por 1000)	Erro padrão	Intervalo de confiança 95%	
≤7	25 a 29	0.91	0.26	0.40	1.42
	30 a 34	1.30	0.31	0.69	1.90
	35 a 39	1.44	0.34	0.79	2.10
	40 a 45	2.47	0.46	1.58	3.36
	45 a 49	3.76	0.58	2.62	4.90
	50 a 54	5.60	0.74	4.15	7.06
	55 a 59	7.51	0.91	5.72	9.29
≥8 e ≤ 11	25 a 29	0.48	0.14	0.22	0.75
	30 a 34	0.69	0.17	0.37	1.02
	35 a 39	0.77	0.18	0.41	1.13
	40 a 45	1.32	0.26	0.82	1.82
	45 a 49	2.01	0.34	1.34	2.67
	50 a 54	2.99	0.46	2.09	3.89
	55 a 59	4.01	0.60	2.84	5.18
≥12	25 a 29	0.34	0.10	0.14	0.54
	30 a 34	0.48	0.13	0.23	0.73
	35 a 39	0.53	0.14	0.26	0.81
	40 a 45	0.91	0.21	0.51	1.32
	45 a 49	1.39	0.28	0.83	1.95
	50 a 54	2.07	0.40	1.30	2.85
	55 a 59	2.78	0.53	1.75	3.81

Fonte: Com base nos dados do Censo 2010, amostra São Paulo, e correções com base no SIM/MS

TAB A 3: Taxas de mortalidade específicas de mortalidade por idade estimadas pelos coeficientes da regressão de Poisson com base nas imputações, população masculina, estado de São Paulo, 2010

Escolaridade	Idade	Taxa de mortalidade (por 1000)	Erro padrão	Intervalo de confiança 95%	
≤7	25 a 29	3.34	0.20	2.95	3.73
	30 a 34	4.02	0.22	3.59	4.45
	35 a 39	4.55	0.23	4.09	5.01
	40 a 45	5.24	0.25	4.74	5.73
	45 a 49	8.78	0.34	8.11	9.45
	50 a 54	10.91	0.40	10.13	11.69
	55 a 59	14.93	0.50	13.96	15.90
≥8 e ≤ 11	25 a 29	1.64	0.10	1.45	1.84
	30 a 34	1.98	0.11	1.76	2.19
	35 a 39	2.24	0.12	2.00	2.48
	40 a 45	2.58	0.14	2.31	2.84
	45 a 49	4.32	0.20	3.94	4.70
	50 a 54	5.37	0.24	4.90	5.84
	55 a 59	7.35	0.32	6.72	7.97
≥12	25 a 29	1.38	0.10	1.18	1.57
	30 a 34	1.66	0.11	1.43	1.88
	35 a 39	1.88	0.13	1.63	2.13
	40 a 45	2.16	0.14	1.88	2.45
	45 a 49	3.63	0.22	3.19	4.06
	50 a 54	4.50	0.27	3.98	5.03
	55 a 59	6.17	0.36	5.46	6.87

Fonte: Com base nos dados do Censo 2010, amostra São Paulo, e correções com base no SIM/MS

TAB A 4: Taxas de mortalidade específicas de mortalidade por idade estimadas pelos coeficientes da regressão de Poisson com base nas imputações, população masculina, município de São Paulo, 2010

Escolaridade	Idade	Taxa de mortalidade (por 1000)	Erro padrão	Intervalo de confiança 95%	
≤7	25 a 29	2.11	0.37	1.38	2.84
	30 a 34	2.31	0.39	1.54	3.08
	35 a 39	2.90	0.46	2.01	3.80
	40 a 45	4.09	0.57	2.97	5.20
	45 a 49	5.33	0.70	3.95	6.70
	50 a 54	8.31	0.96	6.42	10.20
	55 a 59	12.71	1.36	10.05	15.36
≥8 e ≤11	25 a 29	1.43	0.24	0.96	1.89
	30 a 34	1.56	0.26	1.06	2.07
	35 a 39	1.96	0.30	1.37	2.56
	40 a 45	2.77	0.38	2.02	3.52
	45 a 49	3.61	0.47	2.68	4.53
	50 a 54	5.63	0.66	4.34	6.92
	55 a 59	8.60	0.95	6.73	10.47
≥12	25 a 29	0.77	0.15	0.47	1.06
	30 a 34	0.84	0.16	0.52	1.16
	35 a 39	1.06	0.20	0.67	1.44
	40 a 45	1.49	0.26	0.98	1.99
	45 a 49	1.94	0.32	1.31	2.57
	50 a 54	3.03	0.47	2.10	3.95
	55 a 59	4.63	0.70	3.26	5.99

Fonte: Com base nos dados do Censo 2010, amostra São Paulo, e correções com base no SIM/MS

TAB A 5: Taxas específicas de mortalidade por idade e sexo estimadas pelos coeficientes da regressão de Poisson com base nos dados do SIM, município de São Paulo, 2012-13

Escolaridade	Idade	Taxas de mortalidade (por 1000)	
		Homens	Mulheres
≤ 7	25 a 29	3.42	0.95
	30 a 34	3.83	1.32
	35 a 39	4.48	1.83
	40 a 45	6.23	2.52
	45 a 49	8.94	3.95
	50 a 54	13.36	5.84
	55 a 59	20.80	8.68
$\geq 8 \text{ e } \leq 11$	25 a 29	1.55	0.47
	30 a 34	1.74	0.66
	35 a 39	2.03	0.91
	40 a 45	2.83	1.26
	45 a 49	4.05	1.98
	50 a 54	6.06	2.92
	55 a 59	9.44	4.34
≥ 12	25 a 29	0.79	0.35
	30 a 34	0.88	0.49
	35 a 39	1.03	0.67
	40 a 45	1.43	0.93
	45 a 49	2.05	1.46
	50 a 54	3.07	2.16
	55 a 59	4.78	3.21

Fonte: Com base no SIM/MS