

**Estimativas de mortalidade adulta por níveis de escolaridade no estado São Paulo  
em 2010<sup>1</sup>**

**Mirian Martins Ribeiro<sup>2</sup>**  
**Cássio Maldonado Turra<sup>3</sup>**  
**Cristine Campos de Xavier Pinto<sup>4</sup>**

**Palavras-chave: mortalidade adulta, imputação, diferenciais de mortalidade**

---

<sup>1</sup> Trabalho apresentado no VII Congreso de la Asociación Latinoamericana de Población e XX Encontro Nacional de Estudos Populacionais, realizado em Foz do Iguaçu/PR – Brasil, de 17 a 22 de outubro de 2016.

<sup>2</sup> Universidade Federal dos Vales do Jequitinhonha e Mucuri; Centro de Desenvolvimento e Planejamento Regional/Universidade Federal de Minas Gerais

<sup>3</sup> Centro de Desenvolvimento e Planejamento Regional/Universidade Federal de Minas Gerais

<sup>4</sup> Escola de Economia de São Paulo, Fundação Getúlio Vargas/SP

## **1. Introdução**

O objetivo do trabalho foi 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 ou diferenciais regionais (e.g Wood & Carvalho, 1988; Duarte *et al* 2002; Simões, 1999; Simões, 2002). Estudos que considerem a educação têm foco no papel da escolaridade materna sobre a mortalidade infantil ou na infância (e.g Barros & Sawyer, 1993). A partir dos anos 2000, surgiram investigações sobre o papel da educação do indivíduo, usando técnicas indiretas de estimação da mortalidade ou imputando 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, 2016) e o diferencial de mortalidade por escolaridade do responsável pelo domicílio (Silva et al 2016), a partir de informações do Censo.

Contudo, investigações sobre o efeito das características individuais na mortalidade são dificultadas pela falta deste tipo de informação no Censo. O quesito sobre óbito, que é aplicado a todos os domicílios particulares permanentes, só abrange o sexo e a idade do falecido. Deste modo, outra contribuição do trabalho é propor metodologias adequadas para imputar os dados de escolaridade dos indivíduos que faleceram durante o período de referência do Censo.

A amostra do estado de São Paulo foi selecionada para desenvolver e validar a metodologia de imputação, para que futuramente ela possa ser aplicada para estudos de diferenciais educacionais de mortalidade no Brasil.

## **2. Diferenciais de mortalidade por escolaridade**

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 por ser considerada a variável com menor grau de causalção reversa com mortalidade, a menos afetada pela proximidade com a morte e 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, trata-se de informação reportada com maior precisão em pesquisas de inquérito (Liberatos *et al*, 1988; Montez *et al*, 2012).

Nos países desenvolvidos, observa-se um gradiente educacional da mortalidade expressivo: quanto maior a escolaridade menor a mortalidade (Kitagawa & Houser, 1973; Rogers *et al*, 1999). 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; Subramaninan *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, após 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. Ela guarda uma relação com a renda, de modo que pessoas com melhores níveis de educação exibem médias salariais maiores. Portanto, suas funções orçamentárias são diferenciadas daquelas que representam os menos escolarizados. A dotação de recursos permite aos mais escolarizados consumirem 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).

### **3. Metodologia e fonte de dados**

As estimativas foram realizadas com base nos microdados da amostra do Censo Demográfico de 2010 do IBGE, visto que é 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.

O uso do censo demográfico para o estudo de mortalidade é uma vantagem por 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 (Queiroz e Sawyer, 2012).

Os dados do Sistema de Informações sobre Mortalidade (SIM) do Ministério da Saúde foram utilizados para análise de consistência das imputações e das estimativas de mortalidade. O SIM tem sido pouco utilizado para investigações sobre diferenciais de escolaridade por apresentar elevado grau de perda desta variável. Mas a partir de 2009 foram desenvolvidas 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 mudados foi a escolaridade do falecido, que passou a ser coletada em ciclos de ensino seguida da série concluída. Com isso, há uma expectativa de melhoria na cobertura e qualidade da informação, o que torna a base atrativa para avaliar consistência das imputações. 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.

Com relação à imputações, deve-se destacar que no Censo, os respondentes não foram inquiridos sobre outras características dos indivíduos que faleceram no domicílio, além de sexo e idade. A perda da informação sobre a escolaridade é intencional, não se tratando um problema comum de dado faltante, mas que pode ser tratado com metodologias de imputação.

Quanto ao mecanismo de perda, o pressuposto é de que os dados faltantes de escolaridade são ignoráveis ou aleatórios<sup>5</sup>. 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). 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.

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

---

<sup>5</sup> Esses conceitos são discutidos em Little & Rubin (2002) e foram revisados em Ribeiro (2016).

consistência das imputações foram aplicados métodos de imputação diferenciados.

A primeira imputação foi realizada usando imputação múltipla (IM). Ela 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. Serão estimadas regressões logísticas ordenadas.

Após a estimação dos modelos, são obtidos os parâmetros  $\theta$ , que chamamos de  $\theta^*$ . A etapa subsequente consiste estimação de um valor para  $Y_{mis}$  a partir de sua distribuição condicional e do valor estimado para  $\theta^*$ . Em seguida, o processo é repetido  $m$  vezes criando  $m$  extrações da distribuição conjunta *a posteriori* de  $(Y_{mis}, \theta)$ . 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, é possível alcançar estimativas válidas com  $m$  pequeno, tendo sido adotado no trabalho  $m=5$ .

A segunda imputação foi realizada com base no algoritmo *expectation-maximization* (EM) usando um modelo de regressão multivariada normal (MVN). O método MVN utiliza *data augmentation* (DA) com iterações por método de Monte Carlo via cadeias de *Markov* (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.

A seleção das variáveis utilizadas nos modelos das imputações 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 aleatoriedade do *missing* (Rubin, 1978).

Como a amostra do Censo é suficientemente grande para modelos mais inclusivos, a estratégia foi 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. As variáveis selecionadas são apresentadas no Quadro 1 do Apêndice.

Para estimativas da mortalidade, 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 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, 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, é 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.

No caso das taxas de mortalidade estimadas com dados do SIM, não serão 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 será 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, é necessário estimar a contribuição dos vivos com base em dados do Censo. O primeiro passo é 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, toma-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, aplica-se 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. Resultados**

### **4.1. Imputações por IM e algoritmo EM**

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.

As imputações por MVN produziram distribuições para o município de São Paulo estatisticamente iguais às observadas no SIM. Embora os intervalos de confiança das estimativas do censo sejam grandes, eles são menores se comparados aos produzidos por IM (Tabela 1 e Tabela 2). 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.

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 3, 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 IM e EM, 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 EM não aponta nenhum problema de tendência nas iterações, tendo a convergência sido alcançado com 500 iterações.

Tabela 1: Distribuição percentual dos óbitos segundo escolaridade imputada IM, estado e município de São Paulo, 2010

Anos de estudo	SIM	estado de São Paulo				município de São Paulo				
	2012/13	Proporção	Erro padrão	Intervalo de confiança		Proporção	Erro padrão	Intervalo de confiança		Extensão do intervalo
Proporção	95%			95%	95%			95%		
<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
<b>n (100%)</b>	<b>6437</b>	<b>1923</b>				<b>238</b>				
<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
<b>n (100%)</b>	<b>12351</b>	<b>3495</b>				<b>453</b>				

\*Proporção do SIM dentro do intervalo de confiança

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

Tabela 2: Distribuição percentual dos óbitos segundo escolaridade imputada por EM, 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
<b>n (100%)</b>	<b>6437</b>	<b>1898</b>				<b>239</b>				
<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
<b>n (100%)</b>	<b>12351</b>	<b>3477</b>				<b>468</b>				

\*Proporção do SIM dentro do intervalo de confiança

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

Tabela 3: Distribuição dos óbitos de 25-59 anos por sexo, 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	n	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	n	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	n	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

#### 4.2. Mortalidade: diferenciais por nível de escolaridade por idade e sexo

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 4 e Tabela 5).

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 1). Os diferenciais por sexo decrescem com a idade e com o nível de escolaridade (

Figura 2). Para o município de São Paulo, os diferenciais por idade são menores e o gradiente da educação mais marcado.

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 3).

Tabela 4: 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	<b>IRR</b>	Erro Padrão (coef)	Intervalo de confiança 95% (coef)		Coef	<b>IRR</b>	Erro Padrão (coef)	Intervalo de confiança 95% (coef)	
idgrup2										
30 a 34	0.6366	<b>1.8901</b>	0.1358	0.3706	0.9027 *	0.3549	<b>1.4260</b>	0.3564	-0.3436	1.0533
35 a 39	0.9721	<b>2.6434</b>	0.1298	0.7176	1.2265 *	0.4630	<b>1.5889</b>	0.3535	-0.2299	1.1559
40 a 45	1.1734	<b>3.2331</b>	0.1270	0.9246	1.4223 *	1.0005	<b>2.7197</b>	0.3261	0.3614	1.6396 *
45 a 49	1.5409	<b>4.6688</b>	0.1228	1.3003	1.7815 *	1.4213	<b>4.1423</b>	0.3110	0.8118	2.0308 *
50 a 54	1.8871	<b>6.6004</b>	0.1204	1.6511	2.1232 *	1.8195	<b>6.1690</b>	0.3019	1.2277	2.4113 *
55 a 59	2.3021	<b>9.9950</b>	0.1188	2.0692	2.5349 *	2.1120	<b>8.2645</b>	0.2994	1.5252	2.6987 *
educimp										
>=8 e <=11	-0.6669	<b>0.5133</b>	0.0516	-0.7680	-0.5659 *	-0.6277	<b>0.5338</b>	0.1360	-0.8943	-0.3611 *
>=12	-0.6242	<b>0.5357</b>	0.0665	-0.7546	-0.4938 *	-0.9947	<b>0.3698</b>	0.1781	-1.3438	-0.6457 *
_cons	-7.0802	<b>0.0008</b>	0.1146	-7.3048	-6.8556 *	-7.0039	<b>0.0009</b>	0.2873	-7.5671	-6.4408 *
ln(anopes)	1.0000	<b>1.0000</b>				1.0000	<b>1.0000</b>			

\* 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 5: 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

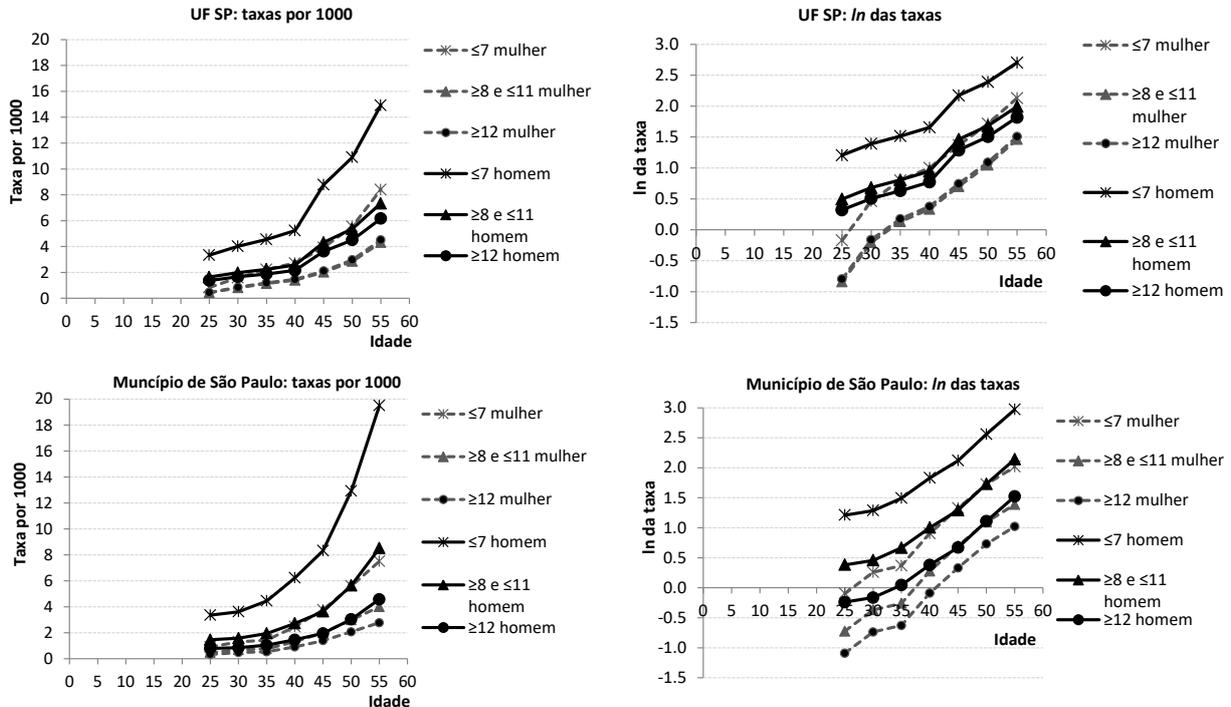
Variáveis	Escolaridade imputada MNV UF SP				Escolaridade imputada MNV município de SP			
	Coef	<b>IRR</b>	Erro Padrão (coef)	Intervalo de confiança 95% (coef)	Coef	<b>IRR</b>	Erro Padrão (coef)	Intervalo de confiança 95% (coef)
<b>idgrup2</b>								
30 a 34	0.1851	<b>1.2034</b>	0.0765	0.0352 0.3350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

\* P-valor ≤ 0.01

P-valor ≥ 0.17

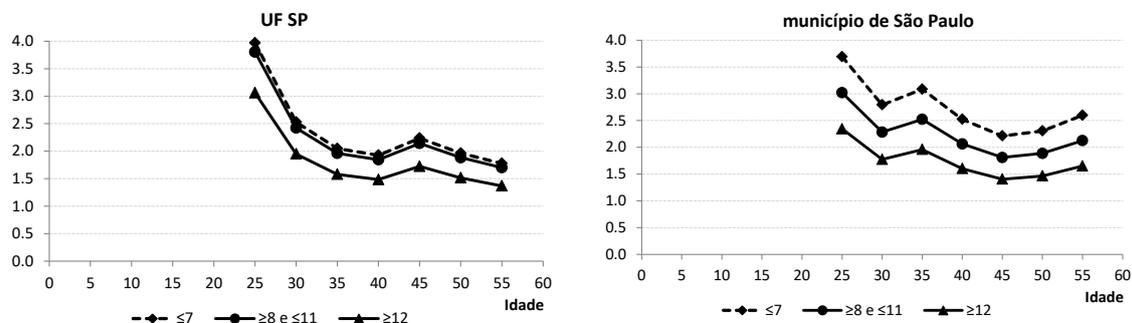
Fonte: IBGE - Censo Demográfico 2010, amostra estado de São Paulo; MS/SIM microdados de 2012 e 2013

Figura 1: Taxas de mortalidade por idade e sexo segundo níveis de escolaridade, estado e município de São Paulo, 2010



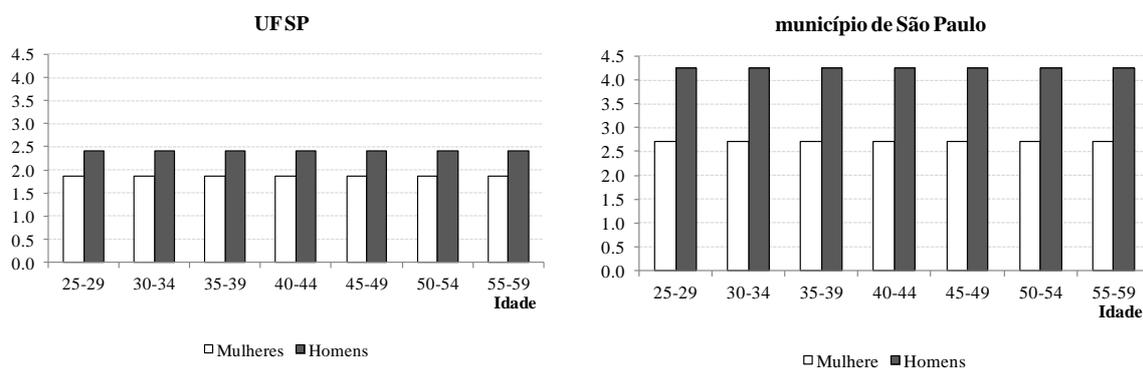
Fonte: IBGE - Censo Demográfico 2010, amostra estado e município de São Paulo

Figura 2: 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 3: Razão entre as taxas de mortalidade do grupo de menor e maior escolaridade por idade e sexo, estado e município de São Paulo, 2010



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

Um dos aspectos de interesse do trabalho foi discutir os ganhos por imputar escolaridade do indivíduo ao invés de usar a escolaridade dos responsáveis pelos domicílios, que se encontra disponível no Censo. 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).

Um teste realizado por Ribeiro (2016) mostrou 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 entre 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:  $\geq 8$ ,  $\geq 9$  e  $\leq 12$ ,  $\geq 13$  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 por idade. A partir da literatura apresentada no capítulo 2, era esperado um diferencial 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.

## **5. Considerações Finais**

O trabalho apresentou 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.

## 6. Referências Bibliográficas

- 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.
- 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.
- Coleman, J. S. 1964. *Introduction to Mathematical Sociology*. New York: Free Press.
- Denisova I. Mortality in Russia: Microanalysis. Center for Economic and Financial Research and New Economic School. Working Paper 128. March, 2009.
- 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.
- 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. and S.H. Preston. Educational Differentials in Mortality: United States, 1979–1985. *Social Science and Medicine*. 42:47–57, 1996.
- 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.
- 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.
- 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.
- House J. S., Landis K. R., Umberson D. Social relationship and health. *Science*, 241: 540-545, 1988.

- 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.
- 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 sobrevivida en la cohorte del proyecto San Francisco. *Rev Méd Chile* 2007; 135: 1370-1379, 2007.
- 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.
- Little R. J. A., Rubin D. B. **Statistical Analysis with Missing Data**. 2 ed. Hoboken, NJ: Wiley, 2002.
- Mackenbach J. P., Stirbu I, Roskam A. J. R., 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.
- 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](http://tabnet.datasus.gov.br/cgi/sim/Consolida_Sim_2011.pdf).
- 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.
- 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., 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. 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.

- 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.
- 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.
- 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.
- Ribeiro, M. M. 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. 2016. Tese (Doutorado em Demografia) – Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais, Belo Horizonte, 2016.
- 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.
- 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, 1999. Cap 7. The effects of basic socioeconomic factors on mortality, p. 115-139.
- 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.
- 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., Multiple imputations in sample surveys. Proceedings Survey Research Methods Section, American Statistical Association: 20-34, 1978.
- Sandoval M. H., Turra C. M. El gradiente educativo en la mortalidad adulta en Chile. Revista Latinoamericana de Población. 17(9): 7-35, 2015.
- Schafer J. L. Analysis of Incomplete Multivariate Data. Chapman & Hall: London, 1997.
- 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.
- Sickles. R. C., Taubman P. Mortality and morbidity among adults and the elderly. In: Rosenzweig M., Stark O. Handbook of Population and Family Economics. Elsevier: 1997.
- 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.

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.

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.

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

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.

## **7. Apêndice**

**Quadro 1: Variáveis selecionadas para as regressões das imputações<sup>6</sup>**

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	
	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	chsexo	sexo do responsável pelo domicílio	1 se homem, 0 se mulher	
	cheducsim2	escolaridade do responsável pelo domicílio	1: <=7 anos 2: >=8 e <=11 3: >=12	
	cheduc	escolaridade do responsável pelo domicílio	anos de estudo	
	chidgrup5	idade do responsável pelo domicílio	1: <30 anos 2: >=30 e <39 3: >=40 e <49 4: >=50 e <=59 5: >=60	
	chconjuge	estado civil do responsável pelo domicílio	1 se tem cônjuge, 0 se não	
	chfisico	responsável pelo domicílio tem pelo menos uma deficiência total ou grande?	1 se sim, 0 se não	
	chcor	raça/cor do responsável pelo domicílio	1: branca 2: preta 3: amarela 4: parda 5: indígena	
	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	indicador de condição socioeconômica do domicílio composto pelas variáveis: * 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	
brasinatod		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

<sup>6</sup> No modelo de imputação por MNV, o ideal é trabalhar com anos de escolaridade, dada a suposição de normalidade da variável dependente. Assim, as imputações foram realizadas usando a escolaridade quantitativa discreta, medida em anos de estudo e variando de 0 a 21. 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. Para o modelo de imputação múltipla (IM), a escolaridade foi categorizada em  $\leq 7$  anos, 8 a 11 anos e  $\geq 12$  anos de escolaridade.

O indicador de nível socioeconômico *ses* foi criado usando análise de componentes principais. Como as variáveis utilizadas são binárias, as componentes foram estimadas com base em matriz de correlação tetracórica, 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.