

Avaliação da qualidade dos dados populacionais e estatísticas vitais do estado de São Paulo, Brasil, 1872-1969

Evaluation of population data quality and vital statistics: São Paulo, Brazil, 1872-1969

Matheus Alves Albino

ma@matheusalbino.com

ORCID: <https://orcid.org/0000-0003-3102-7663>

Pesquisador independente, Brazil

Resumo

O objetivo deste artigo é apresentar os arquivos e fontes para subsidiar o estudo da mortalidade da população do estado de São Paulo, Brasil, bem como definir os procedimentos técnicos e metodológicos empenhados para extrair suas potencialidades e superar as limitações. Recuperamos o histórico das fontes e levantamentos de população e óbitos com considerações críticas sobre a qualidade das informações populacionais dos primeiros censos brasileiros e do registro civil. Apresentamos dois principais conjuntos de dados que fundamentam a análise da mortalidade, estabelecendo os pressupostos para reconstrução das séries de distribuição por sexo e idade da população e dos óbitos do estado de São Paulo.

Abstract

The objective of this article is to present the sources and data to support the study of mortality in the population of the State of São Paulo, Brazil, as well as to define the technical and methodological procedures employed to extract their potential and overcome limitations. We review the history of sources and population data and deaths registration with critical considerations on the quality of population information from the first Brazilian censuses and the civil registry. We present two main datasets that underpin the analysis of

Palavras-chave

Censos
Estatísticas Vitais
Demografia
Histórica
História do Brasil

Keywords

Census
Vital Statistics
Historical
Demography
History of Brazil

mortality, establishing the assumptions for reconstructing the series of distribution by sex and age of the population and deaths in the State of São Paulo.

Enviado: 06/11/24

Aceptado: 22/01/25

Introdução

As evidências da mortalidade e da saúde das populações do passado podem ser extraídas de diversas fontes e, mesmo atualmente, a apuração rigorosa e completa é um ideal. Algumas possibilidades de aproximação envolvem os atestados de óbito de origem eclesiástica ou civil, os registros de sepultamento, documentos fiscais e militares, inventários, anuários estatísticos, testamentos, relatos e relatórios administrativos de hospitais e de delegacias, túmulos e achados arqueológicos. Apesar das inúmeras alternativas, os ingredientes fundamentais para um quadro da mortalidade são duas fontes principais de informação: registros de óbitos e contagens de população.

No Brasil, a partir do Recenseamento Geral do Império de 1872, inaugurou-se uma fase das fontes, denominada “era estatística”, a partir da qual as informações demográficas foram planejadas e sistematizadas periodicamente pelo Estado. Na passagem do Império para a República, particularmente a partir do Decreto nº 9.886, de 7 de março de 1888, os efeitos civis dos registros eclesiásticos foram cessados. Os documentos reunidos nesta discussão se situam nesse espaço, no qual os levantamentos censitários e os registros de óbito estarão sujeitos a instabilidades políticas e conflitos em um caminho difícil rumo à consolidação.

O objetivo deste artigo é apresentar o conjunto de informações reunidas para reconstituir, ao menos parcialmente, o quadro da mortalidade no estado de São Paulo nas primeiras décadas do século XX. Nosso propósito é caracterizar os arquivos e fontes coligidos para subsidiar o estudo da mortalidade da população do estado de São Paulo, bem como delinear os procedimentos técnicos e metodológicos empenhados para extrair suas potencialidades. Apresentamos uma avaliação dos dados demográficos que constam nos levantamentos populacionais e no registro civil de óbitos, concentrando as aplicações às informações disponíveis para o estado de São Paulo desde 1872, com especial ênfase à primeira metade do século XX.

Contribui-se para a organização de dois conjuntos de dados. O primeiro apresenta informações selecionadas dos municípios paulistas, tendo como referência os levantamentos censitários e as estatísticas vitais do registro civil em um período de análise mais amplo, do século XIX ao XXI. O segundo conjunto de dados engloba os óbitos por sexo, idade e causa de morte para o município de São Paulo (1900-1969) e demais municípios do interior do estado (1933-1969) com base nas Estatísticas Demógrafico-Sanitárias e tabelas de óbitos gerais do registro civil.

Recuperamos o histórico dos levantamentos populacionais de São Paulo para derivar as estatísticas de população e sua composição por sexo e idade. Esboçamos considerações críticas sobre a qualidade das informações dos primeiros recenseamentos produzidos e pressupostos assumidos nesse estudo. Entre os principais problemas, destacamos que esses levantamentos possuíam informações pouco confiáveis ou sem cruzamentos de variáveis fundamentais para o dimensionamento da população, de maneira que as análises estão condicionadas a pressupostos sobre a distribuição da população, em especial por sexo e idade. Adicionalmente, examinamos as limitações às quais estão sujeitas as informações populacionais, como os problemas de enumeração.

Discutimos os métodos, técnicas, arquivos e fontes para observação do comportamento da mortalidade a partir dos óbitos do registro civil. Com base nos subsídios fornecidos pela análise crítica das fontes, avaliamos a plausibilidade das estimativas. Pretende-se, com esse exercício, apresentar os ingredientes necessários para a construção de padrões completos e indicadores derivados das tábuas de mortalidade e sobrevivência.

Como resultado, apresentamos uma série de registros de óbitos e contagens de população que produzem um quadro razoavelmente coerente da mortalidade. Com todas as cautelas, a adoção desses diversos procedimentos preparatórios mostra grande potencial para os estudiosos de população, na medida em que permitem revelar novos elementos sobre a transição da mortalidade e epidemiológica em um momento de intensas transformações. Uma implicação importante desse último resultado é a compreensão de que, como reflexo de uma evolução do contexto socioeconômico e político-institucional, a consolidação das estatísticas é mais que simples descrição necessária para a análise da mortalidade.

Trata-se, portanto, de um componente de um processo abrangente de mudanças na organização do sistema de saúde. Como tradução de descobertas no campo tecnológico, ações de combate a epidemias e das reformas sanitárias, as informações demográfico-sanitárias disponíveis e a necessidade de melhor compreender os processos de adoecer e falecer acompanharam uma transição da atenção à saúde. O próprio ato contínuo de aperfeiçoar a mensuração e as ações diante dos indicadores do estado de saúde da população podem provocar um impacto nessas medidas.

Arquivos e fontes

Os levantamentos populacionais

Para a província de São Paulo, existiram diversas tentativas de levantamentos populacionais anteriores ao Recenseamento Geral do Império de 1872, que, embora instrutivos, são pouco seguros ou incompletos. Apesar das omissões e erros, reconhece-se na literatura que ao menos dois documentos possuem significativa importância, não podendo ser descartados. O primeiro é o *Quadro Estatístico*, organizado por Daniel Pedro Müller, em 1836. Trata-se do único levantamento populacional publicado de que se dispõe para a primeira metade do século XIX para São Paulo. Um segundo documento é o *Quadro Estatístico da População de 1854*, que apresenta a população de São Paulo por povoações, grupos etários, nacionalidade, estado conjugal, condição de liberdade e sexo.

O primeiro levantamento censitário nacional de fato foi o Recenseamento do Império do Brasil de 1872, único realizado sob a vigência do Império e da escravidão legal, no qual foram reunidas informações sobre a população escravizada, nacional e estrangeira de todo o território brasileiro (Recenseamento do Império do Brasil de 1872, 1874). Ele apresenta a distribuição da população livre e escravizada por sexo e idade ao nível paroquial, sendo a unidade de referência do levantamento. Contudo, é necessário ter em vista que, para a província de São Paulo, a operação censitária foi realizada em 30 de janeiro de 1874 (Bassanezi, 1998, p. 20).

Diversas críticas a este levantamento são amplamente conhecidas e debatidas, desde a ausência de informações, problemas no formulário, até erros de agregação e soma das tabulações originais, corrigidos por

diversos estudiosos. Uma das principais cautelas ao utilizar os dados deste recenseamento refere-se às relevantes incoerências apontadas por Moura (2022) quanto às idades infantis, com subenumeração de crianças e razões de sexo implausíveis, de forma que o próprio autor propõe ajustes aos totais publicados para tornar qualquer análise nestas faixas de idade minimamente precisa.

Em 1888, foi apresentado o *Censo da Província de 1886*, contendo informações gerais e um censo da população. O documento contém informações sobre a estrutura etária da população livre e escrava, por municípios, tendo como data de referência 30 de setembro de 1886. Também apresenta volumes destinados a estatísticas sobre os filhos livres de mulheres escravizadas, imigrantes e médias anuais de casamentos, nascimentos e óbitos da província com base nos anos de 1883 a 1886 (Bassanezi, 1998, p. 15).

Esse levantamento não reuniu dados de todos os municípios, mas, apesar de sua insuficiência, Camargo (1981, p. 21) observou que essas informações podem ser preferíveis às do Censo de 1890 para o estudo das tendências do crescimento da população. Esta também é a conclusão de Bassanezi (1998, p. 12), argumentando que, a partir das informações constantes desse levantamento, é possível preencher as lacunas ou minimizar as distorções existentes no Censo de 1890, com maior probabilidade de acerto.

O Censo de 1890 se valeu, em boa parte, dos relatórios legados pelo recenseamento do Império. A despeito de problemas com paróquias que não devolveram formulários, o ponto mais crítico deste levantamento foi a apuração e divulgação de seus resultados, considerada bastante ambiciosa pela quantidade de livros planejados para publicação. Essa divulgação fracassou, existindo para esse Censo apenas alguns volumes temáticos, dados para o Distrito Federal, uma síntese do Brasil e nada mais (Senra, 2006, p. 68). Apesar disso, é preciso lembrar que o Censo de 1890 teve uma cobertura bastante razoável no estado de São Paulo, com apenas quatro paróquias não recenseadas (Bassanezi, 1998, p. 12).

O recenseamento de 1900, realizado em todo o país, não foi apurado e publicado completamente; e o planejado para 1910 não ocorreu. Os escassos recursos para a repartição de estatística foram suficientes para

a montagem de uma oficina tipográfica capaz de contemplar a complexidade das tabulações do Censo e o volume de materiais a serem confeccionados (Senra, 2006, p. 341).

O Censo de 1920 é considerado o mais seguro de todos os anteriores e constitui presumivelmente a enumeração mais completa da população até 1940 (Camargo, 1981). Estudiosos do Censo destacam que nunca o país havia planejado tão minuciosamente a operação, desde o trabalho de campo até a apuração e divulgação – um dos grandes gargalos dos levantamentos anteriores (Senra, 2018). Entretanto, a cautela não eximiu o levantamento de omissões importantes como a ausência dos quesitos cor e religião, fundamentais para uma coerente caracterização da população brasileira.

Com o fim da Primeira República, não houve recenseamento em 1930. O governo provisório de Getúlio Vargas e o Estado Novo promoveram reorganizações administrativas no sistema de estatísticas. Senra (2006, 2018) destaca que a mudança político-institucional provocada pela alteração na correlação de forças entre o poder maior das unidades federativas para uma centralização do poder no governo federal traduziu-se em condições favoráveis para o aparecimento de um órgão nacional de estatística com capacidade sistematizadora e normalizadora. Em 1932, foram criados o Instituto Nacional de Estatística (INE) e o Conselho Nacional de Estatística.

Em 1934, foi realizado o Recenseamento Demográfico, Escolar e Agrícola-Zootécnico do Estado de São Paulo (Recenseamento demográfico, escolar e agrícola-zootécnico do estado de São Paulo: 20 de setembro de 1934, 1936). Publicado em 1936, uma das potencialidades deste levantamento é a apresentação de totais populacionais para o estado no agregado, bem como a primeira estimativa de população por situação de domicílio urbano ou rural, com enumeração considerada bastante completa (Camargo, 1981, p. 22).

Logo após a Constituição do Estado Novo, o Decreto-Lei nº 218, de 26 de janeiro de 1938, criou o Instituto Brasileiro de Geografia e Estatística (IBGE). Esse período marca uma consolidação da atuação do IBGE como instituição de pesquisa e as estatísticas passam a ser organizadas em uma coordenação institucional. Especialmente a partir de 1940, o

Brasil iniciou uma nova etapa na história das estatísticas populacionais, sobretudo no que diz respeito às questões voltadas à fecundidade e mortalidade (Instituto Brasileiro de Geografia e Estatística [IBGE], 1940).

Os censos do IBGE de 1940 e 1950 oferecem elementos com os quais Giorgio Mortara inicia a série de estudos demográficos para o país, contribuindo para o desenvolvimento da demografia latino-americana, com o desenvolvimento de métodos e técnicas para avaliação e aperfeiçoamento dos censos para reconstrução do movimento da população brasileira. Em 1960, mediante instabilidade política, como ocorreu em diversos países latino-americanos, os resultados censitários não foram divulgados e se questionou a validade dos próprios dados (IBGE, 1960).

Para a recomposição da evolução populacional e das mudanças na distribuição espacial de São Paulo província e estado, utilizamos com cautela as informações disponíveis desde os primeiros levantamentos, que, muito embora apresentem inúmeras deficiências em seus detalhes, são importantes para a recuperação do quadro geral da história econômica e demográfica paulista.

As estatísticas vitais

O estado civil é um conjunto de qualidades que distinguem o indivíduo na sociedade e na família, iniciando-se com o nascimento e sendo encerrado com a morte. Entre o atar dos laços do indivíduo com a sociedade e o rompimento definitivo, uma série de atos e fatos se constituem como resultados das modificações do estado civil, tal como o casamento, a adoção, a legitimação, o divórcio, a tutela. Os registros vitais, a partir dessas definições, são um processo de obtenção de informações cujo fundamento é anotar cada um dos fatos e acontecimentos que tornam conhecido o estado civil (Silveira e Laurenti, 1973).

As fontes de dados mais conhecidas e amplamente utilizadas pelos demógrafos e historiadores para período anterior ao registro civil são as declarações de batismo, casamentos e óbitos elaboradas por autoridades eclesiásticas, os registros paroquiais. Muito embora tenha havido tentativas de consolidação do registro civil no Brasil em período anterior, um dos momentos críticos foi a lei de criação da Diretoria Geral de Estatística (DGE). Trata-se da primeira legislação que definiu, por meio

das atribuições da diretoria, o funcionamento de duas fontes clássicas de dados demográficos: o censo e o registro civil (Altmann e Ferreira, 1979, p. 400).

A instabilidade política e as crises enfrentadas tiveram profundos reflexos no processo de evolução dos levantamentos populacionais, de forma que tanto o Império quanto a Primeira República falharam em prover recenseamentos em periodicidade decenal, como exigia a legislação. Com a criação do IBGE, a instituição passou a coordenar as atividades estatísticas no país e a conduzir estudos para padronizar leis estaduais, regularizar o levantamento dos eventos vitais e centralizar as estatísticas (Altmann e Ferreira, 1979).

Apesar disso, os Departamentos de Estatística deixaram de ser intermediários nesse fluxo de informações apenas a partir da centralização da coleta pelo IBGE em 1973 (Lei nº 6.015, 1973). Isso possibilitou a consolidação do atual Sistema Nacional de Estatísticas Vitais, a partir do Sistema de Estatísticas do Registro Civil coordenado pelo IBGE, além do Sistema de Informação sobre Mortalidade (SIM), criado em 1975, e do Sistema de Informações sobre Nascidos Vivos (SINASC) em 1990 – os dois últimos sob responsabilidade do Ministério da Saúde. As deficiências no registro civil de óbitos em algumas regiões do Brasil, contudo, são há muito tempo conhecidas.

Mortara (1957, p. 3) afirmou que “a falta de estatísticas fidedignas dos nascimentos e dos óbitos coloca o Brasil em condição de penosa inferioridade no domínio da estatística demográfica internacional”. Felizmente para nosso caso, estudos anteriores já evidenciavam que, mediante alguns ajustes e precauções, a mortalidade poderia ser estudada para populações de alguns municípios urbanos, capitais brasileiras e para o estado de São Paulo, onde o registro de óbitos possuía boa cobertura.

Desde o final do século XIX, os eventos vitais do estado de São Paulo são coletados sistematicamente a partir dos mapas do registro civil enviados pelos cartórios, constituindo-se em rica fonte de informação para o estudo da dinâmica da mortalidade (Grupo Especial de Análise Demográfica [GEADE], 1977). A produção dessa série se iniciou em 1892, sob a responsabilidade da Repartição de Estatística e Arquivo da Secretaria de Negócios do Interior.

Até 1900, essas informações de óbitos gerais e detalhados por *causas mortis* foram publicadas em relatórios anuais enviados à Secretaria do Interior. A partir de 1901, foram publicadas sob o título de *Anuário Demográfico da Seção de Estatística Demógrafo-Sanitária*, em volumes encadernados até 1929. A partir de 1930, as informações foram reunidas em tabelas e mapas do movimento do registro civil enviadas pelos cartórios. A organização dos dados foi de responsabilidade do Departamento Estadual de Estatística (DEE) até que, passando por diversas reorganizações e desdobramentos, teve suas atribuições incorporadas à Fundação Seade.

Anuários e Boletins das Estatísticas Demógrafo-Sanitárias (1896-1929)

Os *Anuários Estatísticos de São Paulo* foram publicações periódicas, de 1896 a 1929, sob responsabilidade da Repartição de Estatística e Arquivo do Estado de São Paulo. Compondo a lista de documentos que tinham como objetivo a divulgação dos resultados das estatísticas vitais, os boletins das Estatísticas Demógrafo-Sanitárias eram compilados pelo Serviço Sanitário, contendo informações sobre nascimentos, casamentos e óbitos para a capital e demais municípios do estado.

A periodicidade dos boletins variou ao longo dos anos de sua produção pela dificuldade de reunir informações mensais de todos os cartórios. O Serviço Sanitário também produziu *Anuários*, com a consolidação dos boletins mensais publicados. De 1906 a 1929, o *Anuário Demográfico da Seção de Estatística Demógrafo-Sanitária* foi publicado ininterruptamente.

No que se refere à confiabilidade e cobertura, as informações sobre a capital do estado são mais completas e bem elaboradas. Também, devido à existência das delegacias de higiene e repartições do Serviço Sanitário, nos principais municípios do interior as lacunas e imperfeições são igualmente poucas ou inexistentes. Em outros, no entanto, as omissões são graves e impedem um panorama completo, mesmo em nível regional.

Nos primeiros anos de publicação, apenas a capital, o município de Campinas – muito provavelmente devido à ação do Serviço Sanitário na epidemia de febre amarela enfrentada pela cidade – e poucos outros municípios divulgaram informações. Com o tempo, as estatísticas são apresentadas de forma mais detalhada e a cobertura se amplia. Contudo,

as estatísticas vitais apenas atingem a cobertura total de municípios levantados a partir de 1918. As Estatísticas Demógrafo-Sanitárias da Capital são publicadas de 1900 a 1969 com óbitos distribuídos por distritos, sexo, idade e causas de morte. Trata-se do conjunto mais completo de informações inicialmente constantes dos Anuários e Relatórios e depois publicadas em tabelas manuscritas enviadas pelos cartórios à Diretoria Estadual de Estatística (DEE).

Tabelas de óbitos gerais do Departamento de Estatística (1930-1969)

A sistematização dos eventos vitais pelo DEE era realizada por pesquisa mensal nos cartórios de registro civil, com base em dois instrumentos de coleta: mapas estatísticos com tabulações específicas dos registros; e as cópias das declarações de óbito. Dessa forma, o Estado sempre processou integradamente as declarações de óbitos e nascimentos e os registros e, ao longo do tempo, aprimorou uma rotina de verificação de consistência das variáveis baseadas na comparação das informações entre as duas fontes (GEADE, 1977).

A partir dessas estatísticas disponíveis na série de tabelas de óbitos gerais, consolidamos uma base de dados dos óbitos da capital detalhados por sexo, grupos de idade e causas de morte, no período de 1900 a 1969, e do interior do estado de São Paulo ao nível agregado para o período de 1933 a 1968. A partir do esforço para coligir todas essas informações, conseguimos os elementos básicos para o estudo da dinâmica da mortalidade no estado.

Os óbitos da capital para 1900, 1901 e 1902 não estão desagregados por grupos de idade, mas apenas por distritos. De 1903 até 1918, eles estão organizados por sexo e grandes grupos etários (0 a 2, 2 a 5, 5 a 9, 10 a 19, 20 a 49 e 50 anos ou mais). De 1918 até 1923, as informações são apresentadas em grupos de idade simples de 0 a 5 anos (correspondentes às faixas de mortalidade infantil e mortalidade na infância), de 10 a 19 anos em idades quinquenais (mortalidade juvenil) e de 20 a 100 anos ou mais, em faixas decenais (mortalidade adulta e em idades avançadas). Finalmente, de 1924 em diante, os óbitos estão desagregados em faixas de idade simples até os 5 anos e depois em idades quinquenais até os 100 anos e mais.

Os óbitos correspondentes ao interior do estado são menos detalhados e de menor cobertura e, portanto, devem ser tratados com cautela. De 1933 a 1968, os dados estão desagregados em idade simples até os 5 anos e depois em idades quinquenais até os 100 anos e mais. No entanto, no que concerne à mortalidade por causas, identificamos que a proporção de óbitos por causas desconhecidas é elevada, em muitos casos superior a 25 % do total registrado, indicando uma maior precariedade da informação. Além disso, por serem menos detalhadas, as informações do interior ocultam as desigualdades regionais, que devem ser ponderadas com o uso de outras informações ao nível municipal.

Métodos e técnicas

Ajustes aos dados brutos

Os dados da população presente recenseada, ao menos para a capital e o total do estado desagregados por sexo e grupos de idade, estão disponíveis apenas para os censos de 1872, 1920, 1940, 1950 e 1960. A partir da informação para os anos cuja distribuição por sexo e idade está disponível, é possível obter estimativas para anos intercensitários, visto que é desejável obter totais para os momentos em que não houve recenseamento. Particularmente, as análises derivadas de estimativas de estrutura da população para o período entre o início da Primeira República e o recenseamento de 1920 estão intrinsecamente condicionadas aos pressupostos estabelecidos sobre a distribuição da população e o comportamento das variáveis demográficas.

Obtemos uma série de totais populacionais considerando um crescimento exponencial entre dois levantamentos, a partir de uma taxa de crescimento intercensitária. Para reconstruir as estruturas etárias no período entre dois censos, foram utilizadas taxas de crescimento anuais médias específicas por grupos etários quinquenais. Além disso, com as informações da data de referência dos censos, reposicionamos mediante interpolação os totais populacionais para gerar estimativas com referência no meio do ano.

É razoável supor que as contagens populacionais aferidas nos recenseamentos do final do século XIX e primeira metade do século XX estejam sujeitas não apenas ao erro de declaração de idade (*age heaping*), mas

também a erros de exagero na idade declarada (*age exaggeration/age overstatement*) nos grupos etários mais avançados, fenômeno especialmente relevante em populações latino-americanas (Palloni e Pinto-Aguirre, 2004; Palloni et al., 2014). Em vista disso, a reconstrução das estruturas etárias exigiu a adoção de alguns procedimentos preliminares de validação e avaliação da consistência.

Para lidar com algumas prováveis fontes de erros de declaração de idade e com o fato de que os grupos etários dos levantamentos são, em alguns casos, apenas decenais ou irregulares, adotamos procedimentos de suavização dos dados por sexo e idade. Esse procedimento é tradicionalmente destinado a gerar estimativas mais plausíveis ou corrigidas de contagens populacionais a partir de informações censitárias. Diversas propostas para extrair estimativas por idade simples e grupos etários quinquenais a partir de grupos etários mais amplos são consagradas na literatura. Sobre tudo com a finalidade de calcular tábuas de mortalidade para países em desenvolvimento com dados deficientes, vale mencionar os métodos de United Nations (1955), Carrier e Farrag (1959) e Arriaga (1968) que ajustam proporcionalmente estruturas etárias em grupos decenais podendo ser posteriormente desagregados em grupos etários menores.

Além disso, para construir distribuições por idade simples, muitas vezes faz-se necessária a adoção de mais de um procedimento para minimizar as irregularidades, de forma que a soma dos valores interpolados seja consistente com os valores originais ou corrigidos anteriormente dos grupos etários. Esses procedimentos de interpolação, também chamados de graduação, são modificações dos dados originais para extrair séries mais suaves de distribuições completas por idade. As mais conhecidas e amplamente utilizadas são as fórmulas de interpolação osculatória de Sprague (1880), Karup-King (Karup, 1899; King, 1908), Beers (1945b) (Beers Ordinário) e Beers (1945a) (Beers Modificado). A seleção do método a ser usado depende, principalmente, do equilíbrio entre suavidade e proximidade de ajuste aos dados, de forma que o pesquisador deve fazer considerações sobre os pressupostos do uso de cada um deles.

Mais recentemente, Rizzi, Gampe e Eilers (2015) propuseram um método flexível de suavização de informações a partir de um modelo chamado *penalized composite link model* (PCLM). O método proposto é baseado no pressuposto de que as contagens de população ou óbitos verificadas

são observações indiretas de uma sequência (teórica) por idade mais suave, o que geralmente é atendido em contagens distribuídas por idade simples. Para que esse método seja eficiente para os anos iniciais e finais da distribuição por idade, os autores recomendam que seja definido um tamanho para o intervalo aberto, isto é, uma idade máxima para a qual não existe nenhuma observação, além de informações mais desagregadas no primeiro ano, que também não costumam atender ao pressuposto de uma distribuição suave.

Os resultados a partir de vários métodos não apresentaram grandes divergências e, portanto, optamos pelo PCLM para estimar estruturas etárias completas, incluindo menores de um ano até os 105 anos, reagrupando posteriormente a informação em faixas quinquenais para o cálculo de tábuas abreviadas de mortalidade. O intervalo aberto considerado nos cálculos foi de 85 anos ou mais, com o objetivo de aproveitar ao máximo a informação disponível para idades avançadas constantes nos óbitos, sem estarmos tão sujeitos aos problemas de exagero na declaração de idade, como a sobre-estimação de centenários nas fontes. A vantagem mais significativa desses critérios e métodos adotados foi a possibilidade de utilizar procedimentos com pressupostos relativamente mais simples, mais recentes e uniformes tanto para a população quanto para os óbitos.

Na construção das séries nesse artigo utilizamos o grupo etário de 85 anos e mais, derivando por meio do método PCLM de Rizzi, Gampe e Eilers (2015) os óbitos de menores de 1 ano até os 105 anos de idade, e reagrupando posteriormente em idade simples no primeiro ano, de 1 até 5 anos, e em grupos quinquenais nas idades seguintes.

Métodos de avaliação da qualidade da informação

Tanto as contagens populacionais aferidas nos recenseamentos quanto os óbitos registrados pelos cartórios de registro civil do final do século XIX e primeira metade do século XX estão sujeitos a diversos tipos de imprecisão: nas idades infantis, é comum o grave sub-registro dos casos; nas idades adultas, os erros de declaração de idade por excesso ou omissão (*age heaping*); e nas idades avançadas, os erros de exagero na idade declarada (*age exaggeration/age overstatement*), fenômeno especialmente relevante em populações latino-americanas (Palloni e Pinto-Aguirre, 2004; Palloni et al., 2014).

As formas de avaliação da qualidade da declaração por idade são variadas, bem como as propostas de correção e redistribuição de totais populacionais para estimar estruturas por idade coerentes. Para avaliar as omissões nas idades infantis e acessar um diagnóstico dos números apurados para São Paulo nos diversos levantamentos, realizamos um exercício de comparação dos resultados dos censos com dados calculados a partir de diversas hipóteses sobre o crescimento da população. Esse tipo de comparação é útil para os casos em que não estão disponíveis informações padronizadas por idade.

Para avaliar o fenômeno de preferência por dígitos na declaração de idade, o índice proposto por Whipple (1919) é um dos indicadores mais amplamente utilizados. Trata-se de uma medida que calcula a proporção relativa da soma das idades múltiplas de 5 e 10 em comparação à soma dos valores das idades de 23 a 62 anos. Dessa forma, o objetivo é medir o quanto as idades terminadas nos dígitos 0 e 5 desviam-se do valor esperado, caso a distribuição da população fosse uniforme entre as idades. Na ausência de atração por dígitos, o valor esperado do índice é 100. Em outro extremo, se todas as idades declaradas terminarem em 0 e 5, o índice atinge o valor máximo de 500, informando que as idades terminadas entre 0 e 5 representam 5 vezes o valor esperado.

A preferência por dígitos depende em grande medida dos aspectos culturais de cada país ou região. Embora a atração pelos dígitos 0 e 5 seja a mais comum, em menor medida pode haver uma repulsão por outros, como dígitos ímpares. Essas anomalias também podem estar presentes nas informações de idade extraídas de datas de nascimento. Para acessar uma medida que considere a preferência por dígitos de uma forma global, Spoorenberg (2007) propôs uma versão modificada do índice de Whipple: calcula-se o índice específico de Whipple (Wi) de atração para todos os dígitos, todos eles cobrindo um intervalo de 20 anos.

Diferentemente do índice de Whipple tradicional, este apresenta valor mínimo de 0 na ausência de qualquer atração por dígitos, até o valor máximo de 18, em que a gravidade da atração por dígitos é máxima. Os primeiros testes de valores de W (0 e 5) revelam que os primeiros censos do IBGE para o estado de São Paulo sofrem do fenômeno de preferência por dígitos de forma não desprezível, classificados de maneira geral como dados irregulares.

Para avaliar o fenômeno de envelhecimento, existem estratégias distintas para os dados de população e de óbitos. Como um exercício para avaliar a magnitude dos erros de exagero de declaração nas idades avançadas nas contagens populacionais, procedemos com cálculo que se aproxima dos procedimentos realizados por Mortara (1949), cujo objetivo foi analisar a consistência das informações apuradas nos censos de 1920 e 1940 mediante algumas hipóteses sobre o crescimento da população brasileira em largo período. Calculamos o número de sobreviventes à data do Censo de 1960, sob determinadas hipóteses sobre a mortalidade e a fecundidade para compará-lo com a população apurada neste recenseamento.

O efeito de um sistemático exagero na declaração de idade em populações idosas é amplamente conhecido pelos analistas, mas nem sempre considerado. Em uma população fechada, os indivíduos acima de 60 ou 70 anos diminuem muito rapidamente devido à alta mortalidade. Contudo, o padrão usual de exagero na idade declarada, que geralmente se inicia com uma modesta transferência para a faixa de 60 anos ou mais, aumenta rapidamente com a idade (Coale e Kisker, 1986, p. 397).

O exagero nas idades avançadas que afeta as estimativas de população também se reflete em imprecisões nas declarações de óbitos, de forma que os dados brutos devem ser compreendidos com cautela e estarem sujeitos a ajustes antes da aplicação de qualquer método para estimar níveis e padrões de mortalidade. Com as tabelas publicadas para o município de São Paulo e o interior do estado, os efeitos de atração por dígitos nos óbitos estão parcialmente ocultos pelo fato de que a informação sobre as mortes foi publicada com agregação em faixas de idade.

Um indicador é comumente utilizado para detectar esse tipo de fenômeno em uma distribuição de óbitos por idade, baseado nos trabalhos de Coale e Kisker (1986). Trata-se da razão de mortes em idades extremas (D_{90+}/D_{60+}) com relação às registradas na população idosa. Os valores de referência dessa medida devem ser os de uma população com dados de mortes considerados precisos, preferencialmente retirados de populações com esperança de vida mais elevada. Em nosso caso, utilizamos a distribuição de mortes da população do Japão no período de 1960 a 1969 extraída da *Human Mortality Database* (2000) como padrão, uma população que neste período já havia passado a transição da

mortalidade e apresentava uma esperança de vida ao nascer ao redor de 70 anos, indicador que o estado de São Paulo apenas atingiria na década de 1990.

Por fim, o problema de *invasão de óbitos* ocorre quando os eventos vitais de uma localidade são “inflados” em razão da ocorrência de óbitos de não residentes. Como o registro de óbitos do município de São Paulo era realizado até 1969 segundo o local de ocorrência do óbito sem a informação de local de residência, é provável que o acesso a serviços de saúde e assistência hospitalar em cidades diferentes daquela da residência provoque esse fenômeno (Waldvogel e Ferreira, 2003). Aqui, utilizamos o mesmo procedimento de Siviero (2009) para corrigir os óbitos relativos ao período de 1920 a 1968 com os dados de Berquó e Gonçalves (1974). Criamos uma versão da série do trabalho de Berquó e Gonçalves (1974) para todos os anos mediante interpolação linear.

Qualidade e ajuste da população recenseada

Omissões e deslocamentos nas idades infantis

Nos primeiros censos brasileiros era comum a declaração de idades infantis sofrer erros por excesso nas idades 0 e 5 anos e por omissão nas idades intermediárias. Uma abordagem intuitiva de avaliar esse problema é por meio da seguinte questão: qual deveria ser o número de nascimentos nos anos anteriores ao censo para que esses nascidos vivos, sujeitos a um nível de mortalidade razoável, fossem exatamente os dados apurados para a respectiva idade no censo? Queremos avaliar qual é a taxa de natalidade necessária no quarto ano anterior ao censo para a população sobrevivente ser rigorosamente a apurada. Se a taxa bruta de mortalidade for elevada sem que haja um motivo plausível para isso, é provável haver um erro por excesso na contagem de sobreviventes à data do censo.

Demonstramos os cálculos de avaliação para o Censo de 1872 na Tabela 1, aplicando a correção, em maior ou menor nível, aos demais até o Censo de 1970. Consideramos que um nível de mortalidade razoável para o período do Censo de 1872 é o da Tábua Brasileira de Mortalidade 1870-1890, calculada por Giorgio Mortara. As taxas brutas de natalidade para o

Brasil no período 1872-1890 foram calculadas em 46,5, mantendo-se em 44 nascidos por mil habitantes até o período 1950-1960.¹

Dessa forma, observa-se que as taxas implícitas estimadas a partir dos recenseados em 1872 e da Tábua de Mortara apresentam variabilidade não razoável e valores implausíveis. O Censo de 1886 para a província de São Paulo indicou uma média anual de 34.778 nascimentos, resultando em uma taxa bruta de natalidade de 33 nascidos por mil habitantes. No entanto, não há informação para 11 municípios, faltam dados de uma ou mais paróquias para três e outros seis não apresentam a informação desagregada por paróquias, indicando sub-registro de nascimentos.

Tabela 1. Estimativas do número de nascidos vivos nos anos anteriores ao censo. Província de São Paulo, 1872.

Idade / Anos anteriores ao Censo	População recenseada em 1/8/1872	Coeficiente de sobrevivência (Mortara)	Nascimentos estimados	População total estimada	Taxa bruta de natalidade (por mil habitantes)
0	30.297	0,783	38.704	835.333	46,3
1	11.726	0,702	16.696	811.467	20,6
2	12.422	0,663	18.726	787.667	23,8
3	16.023	0,643	24.915	763.866	32,6
4	17.098	0,63	27.158	740.066	36,7
5	22.940	0,62	36.997	716.200	51,7

Nota: População total para o período intercensitário estimada mediante interpolação linear. Coeficientes de sobrevivência extraídos da Tábua Brasileira de Mortalidade (1870-1890) de Giorgio Mortara. Nascimentos estimados a partir da aplicação dos coeficientes sobre a população recenseada.

Fonte: Recenseamento do Império do Brasil (1874).

Para o 5º ano de idade, uma taxa de natalidade de 51,7 parece sobre-estimada. Uma média de 35,1 nascidos vivos por mil habitantes para o período é pouco provável, tendo em vista que o nível é relativamente baixo. Aliás, como ressaltou Mortara (1940), os cálculos para os primeiros censos parecem indicar que, num longo período de 1840 a 1920, a natalidade no Brasil se manteve nos níveis em torno de 45-50 nascimentos por mil habitantes. Isso pode indicar uma grave anomalia decorrente da atração pelos dígitos 0 e 5 nas idades infantis, que também se propaga no intervalo quinquenal imediatamente subsequente, nas idades de 6 a 10 anos.

1 Ver, por exemplo, as estatísticas retrospectivas do IBGE: https://seculoxx.ibge.gov.br/images/seculoxx/arquivos_download/populacao/1970/populacao1970ser_04_1.pdf

Os recenseados até o primeiro ano apresentam valores convergentes com a tendência da fecundidade do período, considerando um coeficiente de sobrevivência bastante conservador e a ausência de migração significativa nessa faixa de idade. Os recenseados no primeiro, segundo e terceiro anos muito provavelmente estão errados fortemente por falta. Na faixa etária seguinte, as idades de 6 a 10 anos estão superestimadas, resultado que persiste quando efetuados os cálculos para os censos de 1890, 1900 e 1920. Tanto Mortara (1941, p. 462) quanto Chackiel e Macció (1978) propõem que uma agregação em grupos polianuais para compensar parcialmente os erros por excesso e por falta nas idades infantis.

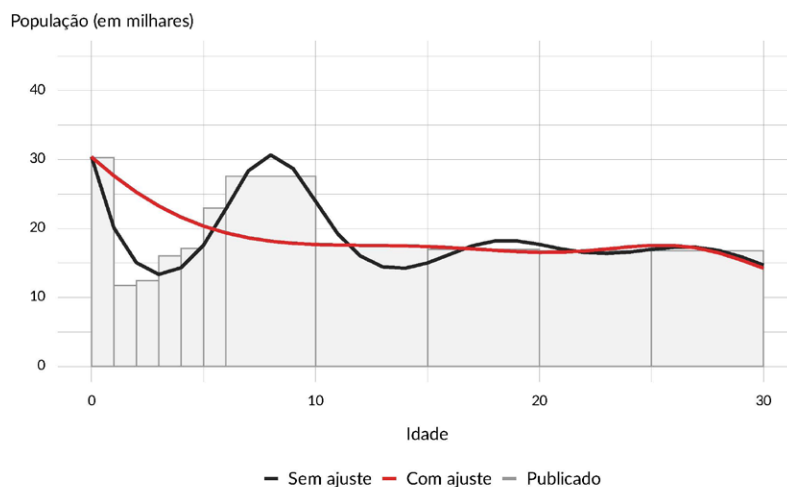
Diante desse problema, faz-se necessário estabelecer procedimento de ajuste das omissões e deslocamentos ocorridos na declaração das idades infantis. Algumas alternativas podem ser adotadas, como o uso de grupos quinquenais não convencionais, que evitem o término em idades múltiplas de 10 e 5, ou ainda o agrupamento em idades decenais. Como alerta Chackiel e Macció (1979), não há para diferentes populações, ou uma mesma população no tempo um padrão-ouro. Conforme Mortara (1940, p. 465), as análises da declaração de idade para diversos algoritmos censitários revelam fortes deficiências nas idades terminadas em 1, 3 e 4, cujo excedente nas idades 0 e 5 correspondem às deficiências nas idades seguintes. Além disso, a gravidade dos problemas de arredondamento cresce com a idade.

A título de ilustração, mostramos o efeito de diferentes procedimentos de agrupamento nas estimativas da população recenseada do Censo de 1872, análise que estendemos para os censos subsequentes. Primeiro, agregamos as informações disponíveis por idade em diferentes faixas para, em um segundo momento, desagregar em idades simples mediante interpolação, preservando os totais dos grupos calculados. Uma vantagem do método PCLM para cindir grupos etários é a robustez quanto a grupos de diferentes tamanhos e intervalos não convencionais.

No procedimento de suavização realizado sobre as idades em intervalos quinquenais (sem ajuste), observamos que permanecem os erros por excesso na faixa de 6 a 10 anos. Além disso, os erros por falta nas idades de 1 a 5 anos são graves. No segundo procedimento, mantivemos o número recenseado de menores de 1 ano e agrupamos o intervalo seguinte até os 10 anos (com ajuste). Esse tipo de agregação apenas redistribui os valores

entre 1 e 10 anos, o que se torna uma forma de supor confiáveis os grupos decenais do Censo, mas encarar com cautela os totais quinquenais, considerando que não absorvem todos os problemas de declaração e preferência por dígitos. Os resultados são ilustrados na Figura 1.

Figura 1. Diferentes estratégias de agrupamento e redistribuição de contagens para a compensação de erros de declaração de idade. Província de São Paulo, 1872.



Fonte: Dados por idade simples extraídos do Recenseamento do Império do Brasil, reagrupados sob diversas estratégias e cindidos mediante interpolação.

As irregularidades da estrutura etária permanecem nos grupos subsequentes, de forma que o número de vivos, que deveria diminuir gradualmente conforme o avançar das idades, continua mais numeroso nas idades terminais em 0 e 5. Em alguns testes, Mortara (1940, p. 464) observou que, mesmo para algumas capitais no Censo de 1920, não há grupamento quinquenal que forneça uma distribuição regularmente decrescente ao avançar da idade, “o que pode explicar porque os elaboradores dos censos brasileiros adotaram constantemente o grupamento decenal para idades seguintes ao trigésimo aniversário”.

Preferência por dígitos na declaração de idade da população adulta

Por não estarem disponíveis as distribuições por sexo e idade simples da população do estado de São Paulo até o Censo de 1940, nos limitamos a avaliar o fenômeno de atração por dígitos na declaração de idade apenas

para os censos do IBGE. O índice de Whipple permite identificar se existe algum tipo de distorção na declaração de idade, algo frequente em regiões com estatísticas de baixa qualidade. Geralmente, valores do índice de Whipple maiores que 125 sugerem informações irregulares ou muito irregulares.

A Tabela 2 demonstra que a gravidade da preferência por dígitos é de fato maior conforme o avançar da idade. Isso nos traz evidências adicionais de que, muito provavelmente, os censos brasileiros estejam afetados pelo problema de exagero na declaração das idades avançadas, tratado adiante. Além disso, o problema parece ser mais pronunciado na distribuição etária das mulheres.

Tabela 2. Grau de atração pelos dígitos 0 e 5 (Whipple) e todos os dígitos (Whipple modificado por Spoorenberg) em faixas de idade selecionadas. Estado de São Paulo, 1940-1960.

Anos	Whipple (23 a 62)		Whipple (58 a 77)		Whipple (78 a 97)		W de Spoorenberg	
	Homem	Mulher	Homem	Mulher	Homem	Mulher	Homem	Mulher
1940	125,7	130,7	156,3	174,9	178,5	203,8	0,95	1,12
1950	121	123,1	141,1	157,6	164	185,6	0,72	0,81
1960	127,3	128,7	155,9	165,4	157,6	171	1,13	1,19
Referência	100,0	100,0	100,0	100,0	100,0	100,0	0	0
Suíça (1960)	101,5	101,0					0,1	0,08

Nota: Índice de Whipple: dados geralmente classificados como muito precisos (menor que 105), relativamente precisos (entre 105 e 110), aproximados (entre 110 e 125), irregulares (entre 125 e 175) ou muito irregulares (175 ou mais). Índice de Spoorenberg: valores próximos de zero indicam dados mais acurados.

Fonte: Censos Demográficos do IBGE (1940, 1950, 1960). Cálculos para a Suíça realizados por Spoorenberg (2007).

O índice Whipple modificado avalia o nível geral de preferência pela declaração concentrada em determinadas idades. Dessa forma, para uma população com dados precisamente apurados, é esperado um valor próximo de zero nessa medida de avaliação. De fato, ao utilizar a população da Suíça como referência, tal qual no trabalho original de Spoorenberg (2007), identificam-se inconsistências na declaração por idade da população paulista apurada nos recenseamentos do IBGE na primeira metade do século XX.

Em perspectiva comparativa entre regiões e temporalidades distintas, os valores calculados para a população escravizada da paróquia da Madre

de Deus de Porto Alegre, a partir do Rol de Confessados de 1779, são 325,9 homens e 333,3 mulheres, isto é, classificados como muito irregulares (Scott, 2020, p. 120). Paes e Albuquerque (1999, p. 36) apontam que, para 1980, as informações de idade dos estados de Roraima, Acre, Maranhão e Alagoas possuem o índice acima de 125, sendo classificados como irregulares, fato que só foi alterado para todas as regiões do país no Censo de 1990.

***Exagero de declaração nas idades avançadas
(fenômeno do “envelhecimento”)***

Segundo Mortara (1941, p. 469), nos países que procederam a uma apuração mais detalhada da idade declarada, grande parte dos pretendidos centenários tornam-se nonagenários, octogenários, septuagenários e até sexagenários. Os censos brasileiros confirmam a experiência de que o fenômeno de envelhecimento, isto é, de exagero na idade declaração, em especial nas idades avançadas, tem uma gravidade relativamente maior do que o fenômeno contrário de rejuvenescimento em casos específicos, como nas mulheres adultas. Embora a população tenha crescido substancialmente entre 1872 e 1920, os números dos quatro primeiros censos mostram uma diminuição contínua de indivíduos em idades avançadas, evidenciando particularmente esses erros na declaração.

Utilizamos as informações da população total do estado de São Paulo disponíveis em 1854 e 1872 para estimar a população nos 106 anos anteriores ao censo a partir da aplicação de uma taxa média de crescimento anual linear. A taxa bruta de natalidade no ano de 1854 foi apurada em 45,8 nascidos vivos por mil habitantes (Quadro estatístico do movimento da população da Província de São Paulo durante o ano de 1854, 1855). Embora mereça cautela, esse nível de nascimentos é bastante próximo da média estimada por Mortara (1949) para o Brasil. Supondo que essa taxa tenha sido constante entre os 1854 e 1860, podemos calcular uma aproximação do número de nascidos vivos até 106 anos anteriores ao Censo de 1960.

A taxa de mortalidade infantil estimada para 1960 foi calculada em 78 óbitos por mil nascidos vivos. Dessa forma, supomos as probabilidades de morte iguais às da Tábua de Mortalidade West de Coale e Demeny (1966), revisada e estendida por United Nations (2011), ao nível correspondente à

mortalidade calculada. Efeitos da migração foram considerados desprezíveis nestas faixas de idade.

A Tabela 3 traz evidências de que, sob essas hipóteses, a população com 100 anos ou mais para o estado de São Paulo em 1º de setembro de 1960 foi estimada em 25 indivíduos caso o padrão de mortalidade se aproxime das tábuas Coale-Demeny. O valor apurado pelo censo é de 1.494 habitantes, sugerindo graves erros de exagero na declaração de idade. A natalidade nos 106 anos anteriores ao censo já encontra níveis bastante elevados, de maneira que as elevar ainda mais, a níveis nunca observados, não altera significativamente os resultados. A mortalidade experimentada nesse período deveria ser muito superior à utilizada, o que reduziria ainda mais o número de sobreviventes estimados.

Tabela 3. Estimativas de sobreviventes a partir dos nascidos vivos nos 106 anos anteriores ao censo. Estado de São Paulo, 1960.

Ano de referência	População média estimada	Nascimentos estimados (TBN = 45,8)	Idade / anos anteriores ao Censo	Sobreviventes estimados em 01/09/1960	
				CD West	UN General
1860	558.139	25.563	100	12	33
1859	534.871	24.497	101	6	20
1858	511.604	23.431	102	4	12
1857	488.336	22.366	103	2	7
1856	465.068	21.300	104	1	4
1855	441.800	20.234	105	0	2
1854	418.532	19.169	106	0	1
P100 + Estimada				25	79
P100 + Censo				1.494	1.494

Nota: Hipótese de Taxa Bruta de Natalidade constante em 45,8 nascimentos por mil habitantes, calculada pelo método direto com dados de São Paulo (1855). Coeficientes de Sobrevivência extraídos das Tábuas de Mortalidade Coale-Demeny West (CD West) e Geral das Nações Unidas (UN General), $e_0 = 60$.

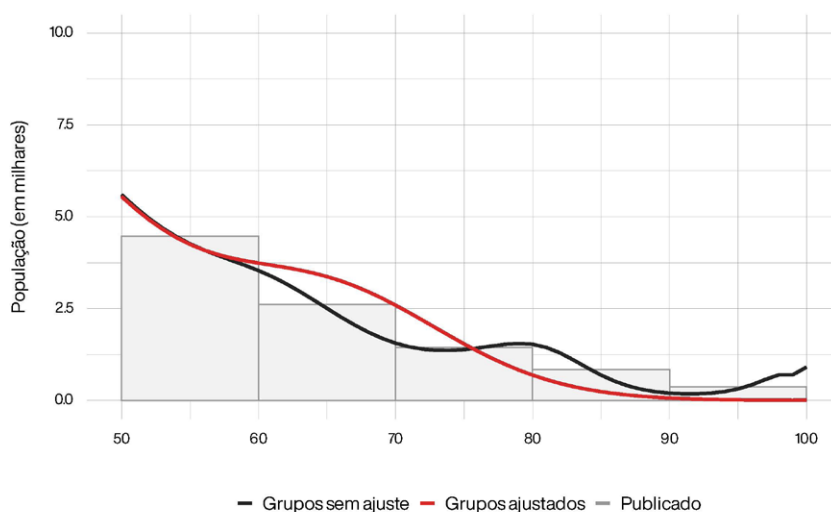
Fonte: Censo Demográfico do IBGE (1960). Quadro estatístico do movimento da população da Província de São Paulo durante o ano de 1854. (1855). Quadro estatístico da população da Província de São Paulo recenseada no ano de 1854. (1856).

As estimativas indiretas da população centenária brasileira de Nepomuceno e Turra (2020) sugerem uma grave sobre-enumeração dos censos nas idades avançadas, de tal forma que o número informado de centenários nos primeiros levantamentos não foi alcançado nem mesmo um

século depois, conforme as estimativas indiretas. Os autores sustentam que antes do século XX possivelmente não existiu nenhum indivíduo recenseado com 100 anos ou mais. As estimativas para o Brasil em 1900 apontam para um máximo de 2 indivíduos, enquanto nos anos 1970 seria de, no máximo, 200 habitantes. Esses valores são consistentes com nossas estimativas.

Tendo em vista esses problemas, executamos correções adicionais aos dados brutos antes de utilizá-los. Para as informações populacionais do Censo de 1872, cujas incoerências em idades infantis são hoje conhecidas e documentadas, reunimos a população apurada em grupos etários decenais. Os recenseamentos de 1872, 1890 e 1920 apresentam a informação de idade até o grupo de 100 anos e mais. No Censo paulista de 1934, o grupo etário aberto era o de 65 anos e mais. Em 1940, 1950 e 1960, o intervalo é de 80 anos e mais. Em 1970, a informação está disponível até 70 anos e mais. Essa heterogeneidade na apresentação das informações apuradas nos censos impõe um dilema entre aproveitar ao máximo as informações para as idades avançadas e ser criticamente prudente com a presença de centenários em tais levantamentos. A Figura 2 apresenta uma saída para esse problema.

Figura 2. Estratégias de agrupamento e redistribuição de contagens para compensação de exagero na idade declarada. Província de São Paulo, 1872.



Fonte: Dados por idade simples extraídos do Recenseamento do Império do Brasil, reagrupados sob diversas estratégias e cindidos mediante interpolação.

Em seu estudo sobre a demografia histórica do Rio de Janeiro escravista, Moura (2022, p. 476) apresentou um conjunto de ajustes críticos aos dados do Censo de 1872. O autor apontou para a necessidade de três ajustes da população nas idades avançadas: eliminar a ocorrência de idades superiores a 100 anos; reduzir a idade média das ocorrências acima de 80 anos; e eliminar a concentração de registros em idades múltiplas de 10. Em nosso entendimento, para os censos de 1872 a 1920, é ainda mais prudente reduzir o intervalo aberto das informações por idade até a faixa de 60 anos e mais, considerando que ela representa o número de fato apurado pelo recenseamento, e a partir da qual os erros de declaração de idade tornam a informação muito imprecisa.

Adicionalmente, determinamos uma idade máxima a partir da qual não existirá nenhuma ocorrência, tal como o primeiro ajuste realizado por Moura (2022). Em outras palavras, definimos um tamanho para o intervalo aberto, em que a idade é de 105 anos. Um procedimento adicional é adotado: os valores gerados por esse ajuste matemático são novamente redistribuídos de acordo a função de mortalidade do Brasil para 1960 estimada por United Nations (2019). Para os anos em que não houve recenseamento, estimamos a população a partir de uma taxa de crescimento dos grupos etários.

Para os anos nos quais existe uma estimativa da população total, mas a estrutura da população por sexo e idade não está disponível, pressupomos uma tendência linear da mudança nas razões de sexo de cada grupo etário e redistribuímos o total da população conhecido entre as faixas de idade e sexo. Para facilitar o cálculo de estimativas com base no período intercensitário, todas as informações populacionais foram trazidas da data de referência original dos levantamentos para o dia 1º de janeiro mediante interpolação linear. A população sem idade declarada foi redistribuída proporcionalmente conforme as faixas de idade presumida, produzindo estimativas que reproduzem os totais publicados dos censos, corrigidos os erros de soma e preservando as estruturas etárias estimadas.

Qualidade e ajuste dos óbitos registrados

As informações de óbitos estão sujeitas a fontes semelhantes de imprecisões. Os óbitos da capital paulista para 1900, 1901 e 1902 não estão

desagregados por grupos de idade, mas apenas por distritos. De 1903 até 1918, eles estão organizados por sexo e grandes grupos etários (0 a 2, 2 a 5, 5 a 9, 10 a 19, 20 a 49, 50 anos ou mais). De 1918 até 1923, as informações são apresentadas em grupos de idade simples de 0 a 5 anos (correspondentes às faixas de mortalidade infantil e mortalidade na infância), de 6 a 9 anos e de 10 a 19 anos em idades quinquenais (mortalidade juvenil) e de 20 a 100 anos ou mais, em faixas decenais (mortalidade adulta e em idades avançadas). Finalmente, de 1924 em diante, os óbitos estão desagregados em faixas de idade simples até os 5 anos e depois em idades quinquenais até os 100 anos e mais.

Os óbitos correspondentes ao interior de São Paulo são menos detalhados e de menor cobertura, devendo, portanto, ser tratados com maior cautela. De 1933 a 1968, os dados estão desagregados em idade simples até os 5 anos e depois em idades quinquenais até os 100 anos e mais. No entanto, no que concerne à mortalidade por causas, identificamos que a proporção de óbitos por causas desconhecidas é elevada, em muitos casos superior a 25 % do total registrado, indicando uma maior precariedade da informação. Além do mais, por serem menos detalhadas, as informações do interior ocultam as desigualdades regionais, que devem ser ponderadas com o uso de outras informações em nível municipal.

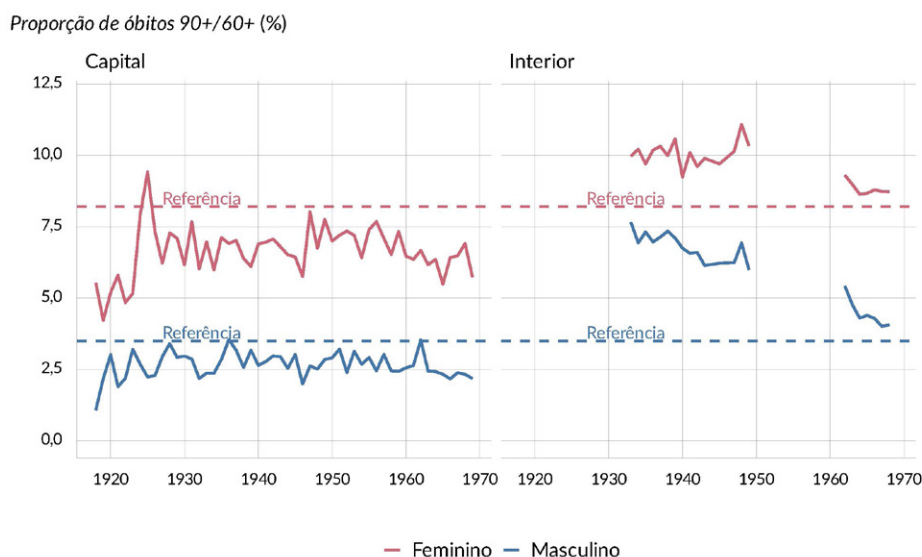
O envelhecimento na declaração da idade ao óbito

As medidas de Coale-Kisker são adaptações da proposta desses autores, que avaliaram se as taxas de mortalidade da população negra em idades avançadas nos Estados Unidos eram de fato menores do que as da população branca, ou se isso decorria de um problema dos dados. Por esse motivo, são úteis na avaliação de exagero na declaração de idade ao óbito (Coale e Kisker, 1986).

A Figura 3 traz evidências de que os problemas de exagero na idade ao óbito são significativos para o período 1920-1970, sendo mais graves nas estatísticas do interior do estado do que na capital. Em todo o período em que a informação para o interior é disponível, a razão entre óbitos de pessoas em idades extremas sobre a população idosa é mais elevada do que os valores de referência para ambos os sexos.

Ao observar o mesmo fenômeno para as idades avançadas na faixa dos 70 anos e mais, com relação ao total da população idosa, esse efeito de exagero é atenuado, sugerindo que uma possível agregação da informação em intervalos abertos menores corrija a informação de óbitos por idade.

Figura 3. Óbitos de pessoas de 90 anos ou mais sobre aqueles registrados na população idosa (D90+/D60+). Capital e interior do estado de São Paulo, 1900-1969.



Fonte: *Annuario demographico: Secção de estatística demographo-sanitaria. Anno X XVI-1929 (Vol. 1). (1932). Estatísticas demógrafo-sanitárias da capital e do interior de São Paulo (1900-1969). Valor de referência: Japão (1960-1969) extraído de Human Mortality Database (2000).*

No estudo comparativo de Rizzi et al. (2016), um foco particular foi dado à comparação do desempenho de diversos métodos de desagregação de distribuições agrupadas quando há um amplo intervalo aberto no final da distribuição. Interpolações usando *splines* são amplamente utilizadas em *Human Mortality Database* para desagregar contagens de mortes agrupadas em faixas quinquenais em mortes por idade simples (Wilmoth et al., 2021, Apêndice B), exigindo tratamento especial para idades avançadas (Wilmoth et al., 2021, Apêndice C). O método PCLM foi o menos afetado, distribuindo suavemente os valores nas idades extremas. Adicionar a informação de idade máxima para a qual nenhuma observação é plausível melhora a estimação do método e é essencial para a redistribuição eficiente das observações nessas idades extremas (Rizzi et al., 2016, p. 8).

Em virtude dessas evidências, as distribuições de óbitos foram corrigidas do original a partir dos mesmos procedimentos que aplicamos para a população: os óbitos foram agrupados e depois desagregados mediante interpolação com o método PCLM até a idade máxima de 105. Essa distribuição foi, após a aplicação do método, reagrupada até os 85 anos e mais, para coincidir com a distribuição da população estimada. Toda a série de óbitos da capital e do interior de São Paulo foi corrigida por esses procedimentos, bem como todos os censos demográficos.

Considerações finais

A partir da condução de uma análise documental dos censos e estatísticas demógrafo-sanitárias, identificamos que os primeiros levantamentos censitários brasileiros estão sujeitos a omissão nas idades infantis, exagero no número de centenários e erros de declaração de idade. Por sua vez, o registro de óbitos cobriu consistentemente todos os municípios apenas a partir de 1918, e qualquer análise com base nessas fontes deve levar isso em consideração.

Estatísticas populacionais dos municípios do estado de São Paulo (1836-2022) consolidam diversas informações regionais e desdobramentos territoriais de São Paulo, juntando fontes demográficas, cartográficas e econômicas e de saúde dos municípios paulistas durante o período 1836 a 2022, provenientes de diversas fontes. Estatísticas Demógrafo-Sanitárias da Capital e do Interior de São Paulo (1900-1969) é um conjunto construído a partir da leitura das tabelas de óbitos fornecidas pela Fundação Seade e das apurações dos censos demográficos. Além disso, consolidam a informação dos óbitos por sexo, idade e causas de morte para o Município de São Paulo (1900-1969) e do Interior do Estado (1933-1969). A descrição completa das bases de dados pode ser consultada no Anexo.

Com as informações necessárias para a análise da dinâmica da mortalidade do estado de São Paulo em um longo período, temos consolidadas as principais séries históricas que compõem dois conjuntos principais que subsidiam o quadro da mortalidade, com implicações de avaliação e correção dos dados que podem ser utilizadas em outros contextos de países da América Latina.

Referências

- Albino, M. A. (2024a). *Estatísticas demógrafo-sanitárias da capital e do interior de São Paulo (1900-1969)*. Harvard Dataverse, V1. UNF:6:mlt/D37vSWx3mBi32+Jbqw. <https://doi.org/10.7910/DVN/37MCA2>
- Albino, M. A. (2024b). *Estatísticas populacionais dos municípios do Estado de São Paulo (1836-2022)*. Harvard Dataverse, V1. <https://doi.org/10.7910/DVN/FOUZS3>
- Altmann, A. M. G. e Ferreira, C. E. C. (1979). Evolução do censo demográfico e registro civil como fontes de dados para análise da fecundidade e mortalidade no Brasil. *Revista Brasileira de Estatística*, 40(160), 399-453.
- Annuario demographico: Secção de estatística demographo-sanitaria. Anno X XVI-1929* (Vol. 1). (1932). Imprensa Oficial.
- Arriaga, E. E. (1968). *New life tables for Latin American populations in the nineteenth and twentieth centuries*. University of California.
- Barbieri, M., Wilmoth, J. R., Shkolnikov, V. M., Glej, D. Jasilionis, Dmitri Jdanov, D., Boe, C., Riffe, T., Grigoriev, P. e Winant, C. (2015). Data resource profile: The Human Mortality Database (HMD). *International Journal of Epidemiology*, 44(5), 1549-1556. <https://doi.org/10.1093/ije/dyv105>
- Bassanezi, M. S. C. B. (Org.). (1998). *São Paulo do passado: Dados demográficos*. Nepo/Unicamp.
- Beers, H. S. (1945a). Modified-interpolation formulas that minimize fourth differences. *The Record of the American Institute of Actuaries*, 34(69), 14-20.
- Beers, H. S. (1945b). Six-term formulas for routine actuarial interpolation. *The Record of the American Institute of Actuaries*, 34(69), 59-60.
- Berquó, E. S. e Gonçalves, M. A (1974). *A invasão de óbitos no município de São Paulo*. CEBRAP.
- Camargo, J. F. (1981). *Crescimento da população do estado de São Paulo e seus aspectos econômicos*. FIPE.
- Carrier, N. H. e Farrag, A. M. (1959). The reduction of errors in census populations for statistically underdeveloped countries. *Population Studies*, 12(3), 240-285.
- Chackiel, J. e Macciò, G. (1978). *Evaluación y corrección de datos demográficos: VI. Análisis de la población por edades*. CELADE.
- Chackiel, J. e Macciò, G. (1979). *Evaluación y corrección de datos demográficos: VII. Técnicas de corrección y ajuste de la mala declaración de la edad*. CELADE.

- Coale, A. e Demeny, P. (1966). *Regional model life tables and stable populations*. Princeton University Press.
- Coale, A. e Kisker, E. E. (1986). Mortality crossovers: Reality or bad data? *Population Studies*, 40(3), 389-401.
- Diretoria Geral de Estatística. (1890). *Recenseamento da população da República dos Estados Unidos do Brasil em 31 de dezembro de 1890*. Diretoria Geral de Estatística.
- GEADE. Grupo Especial de Análise Demográfica. (1977). *Diagnóstico da situação das estatísticas vitais no Estado de São Paulo, 1930 a 1976*. (1977). Secretaria de Economia e Planejamento.
- Human Mortality Database. (2000). *Human mortality database*. University of California. <https://mortality.org/>
- IBGE. Instituto Brasileiro de Geografia e Estatística. (1950). *Censo Demográfico de 1950*.
- IBGE. Instituto Brasileiro de Geografia e Estatística. (1960). *Censo Demográfico de 1960*.
- IBGE. Instituto Brasileiro de Geografia e Estatística. (1940). *Recenseamento Geral do Brasil de 1940*.
- Karup, J. (1899). On a new mechanical method of graduation. In *International Actuarial Congress, 1898, 2. Transactions of the Second International Actuarial Congress* (pp. 78-109). C. & E. Layton.
- King, G. (1908). On the construction of mortality tables from census returns and records of deaths. *Journal of the Institute of Actuaries*, 42(3), 238-246.
- Lei nº 6.015, de 31 de dezembro de 1973. (1973). Dispõe sobre os registros públicos, e dá outras providências. *Diário Oficial da República Federativa do Brasil*, 111(249), Seção 1.
- Moura, H. P., Filho. (2022). As falhas no registro de crianças no Censo de 1872: O caso fluminense. *Revista Brasileira de Estudos de População*, 39, 1-20. <https://doi.org/10.20947/S0102-3098a0203>
- Mortara, G. (1940). Estudos sobre a utilização do censo demográfico para a reconstrução das estatísticas do movimento da população do Brasil. *Revista Brasileira de Estatística*, 1(3), 443-472.
- Mortara, G. (1941). Estudos sobre a utilização do Censo Demográfico para a reconstrução das estatísticas do movimento da população do Brasil. *Revista Brasileira de Estatística*, 2(7), 493-538.
- Mortara, G. (1945). Estudos sobre a mortalidade no Distrito Federal e no município de São Paulo. *Revista Brasileira de Estatística*, 6(24), 583-776.

- Mortara, G. (1949). *Methods of using census statistics for the calculation of life tables and other demographic measures (with applications to the population of Brazil)*. United Nations.
- Mortara, G. (1957). Incógnitas da mortalidade no Brasil. *Revista Brasileira de Estatística*, 18(69), 1-18.
- Nepomuceno, M. e Turra, C. (2020). The population of centenarians in Brazil: historical estimates from 1900 to 2000. *Population and Development Review*, 46(4), 813-833.
- Paes, N. A. e Albuquerque, M. E. (1999). Avaliação da qualidade dos dados populacionais e cobertura dos registros de óbitos para as regiões brasileiras. *Revista de Saúde Pública*, 33(1), 33-43.
- Palloni, A. e Pinto-Aguirre, G. (2004). One hundred years of mortality in Latin America and the Caribbean: The fragile path from hunger to longevity. In *Population Association of America Meeting*, 2004 (pp. 1-37). PAA.
- Palloni, A., Pinto-Aguirre, G. e Beltrán-Sanchez, H. (2014). *Latin American Mortality Database (LAMBDa)*. University of Wisconsin. <https://www.ssc.wisc.edu/cdha/latinmortality2/>
- Pascariu, M. D., Daňko, M. J., Schöley, J. e Rizzi, S. (2018). ungroup: An R package for efficient estimation of smooth distributions from coarsely binned data. *The Journal of Open Source Software*, 3(29), 937. <https://doi.org/10.21105/joss.00937>
- Quadro estatístico do movimento da população da Província de São Paulo durante o anno de 1854.* (1855). Typographia 2 de Dezembro.
- Quadro estatístico da população da Província de São Paulo recenseada no anno de 1854.* (1856). Typographia 2 de Dezembro.
- Recenseamento demographico, escolar e agrícola-zootécnico do Estado de São Paulo: 20 de setembro de 1934.* (1936). Imprensa Oficial.
- Recenseamento do Império do Brazil de 1872.* (1874). Typographia de Georges Leuzinger.
- Rizzi, S., Gampe, J. e Eilers, P. H. C. (2015). Efficient estimation of smooth distributions from coarsely grouped data. *American Journal of Epidemiology*, 182(2), 138-147.
- Rizzi, S., Thinggaard, M., Engholm, G., Christensen, N., Johannesen, T. B., Vaupel, J. W. e Lindahl-Jacobsen, R. (2016). Comparison of non-parametric methods for ungrouping coarsely aggregated data. *BMC Medical Research Methodology*, 16(59), 1-12. <https://doi.org/10.1186/s12874-016-0157-8>

- Scott, D. (2020). *Livres e escravos: população e mortalidade na Madre de Deus de Porto Alegre (1772-1872)* (Tese de Doutorado, Universidade Estadual de Campinas). Repositório da Produção Científica e Intelectual da Unicamp. <https://repositorio.unicamp.br/acervo/detalhe/1129229>
- Senra, N. de C. (2006). *História das estatísticas brasileiras* (Vol. 4). IBGE.
- Senra, N. de C. (2018). Da DGE até hoje, com o IBGE, uma sucessão em linha reta. *Estatística y Sociedad*, (5), 56-81.
- Silva, F. S. (2009). *Tendências e projeção da mortalidade do município de São Paulo – 1920 a 2100* (Dissertação de Mestrado, Universidade Federal de Minas Gerais). Repositório Institucional da UFMG. https://repositorio.ufmg.br/handle/1843/AMSA-7WZETS?locale=pt_BR
- Silveira, M. H. e Laurenti, R. (1973). Os eventos vitais: Aspectos de seus registros e inter-relação da legislação vigente com as estatísticas de saúde. *Revista de Saúde Pública*, 7(1), 37-50. <https://doi.org/10.1590/S0034-89101973000100004>
- Siviero, P. C. L. (2009). *Níveis e padrões do diferencial de mortalidade por sexo no município de São Paulo, 1920-2005* (Dissertação de Mestrado, Universidade Federal de Minas Gerais). Repositório Institucional da UFMG. <https://repositorio.ufmg.br/handle/1843/AMSA-7UTLAM>
- Spoorenberg, T. (2007). Quality of age reporting: Extension and application of the modified Whipple's index. *Population*, 62(4), 729-741.
- Sprague, T. B. (1880). Explanation of a new formula for interpolation. *Journal of the Institute of Actuaries*, 22(4), 270-285.
- United Nations. (1955). *Manual II: Methods of appraisal of quality of basic data for population estimates*.
- Waldvogel, B. e Ferreira, C. E. de C. (2003). Estatísticas da vida. *São Paulo em Perspectiva*, 17(3-4), 55-66.
- Whipple, G. C. (1919). *Vital statistics: An introduction to the science of demography*. Chapman & Hall.
- Wilmoth, J. R., Andreev, K., Jdanov, D. e Gleijer, D. A. (2021). *Methods protocol for the Human Mortality Database*. University of California. <https://mortality.org/File/GetDocument/Public/Docs/MethodsProtocolV6.pdf>

Anexo

Tabela 1. Estatísticas demógrafo-sanitárias da capital e do interior de São Paulo, 1900-1969.

Conjunto	Descrição	Abrangência	Variáveis
ESP_pop_5x1	Série anual da população do estado de São Paulo por sexo e idade, em grupos	1872-1970	Código, sexo, idade, data de referência, população
MSP_pop_5x1	Série anual da população do município da Capital por sexo e idade, em grupos quinquenais	1872-1970	Código, sexo, idade, data de referência, população
INT_pop_5x1	Série anual da população dos municípios do interior de São Paulo por sexo e idade, em grupos quinquenais	1872-1970	Código, sexo, idade, data de referência, população
ESP_deaths_5x10	Série intercensitária dos óbitos do estado de São Paulo por sexo e idade, em grupos quinquenais	1900/19	Código, sexo, idade, data de referência, óbitos
MSP_deaths_5x10	Série intercensitária dos óbitos do interior de São Paulo por sexo e idade, em grupos quinquenais	1900/19	Código, sexo, idade, data de referência, óbitos
MSP_nrdeaths_5x10	Série intercensitária dos óbitos de não residentes da capital por sexo e idade, em grupos quinquenais	1920/33; 1960/69	Código, sexo, idade, data de referência, óbitos
INT_deaths_5x10	Série intercensitária dos óbitos do Interior de São Paulo por sexo e idade, em grupos quinquenais	1900/19	Código, sexo, idade, data de referência, óbitos
XXX_cause_deleted_5x10	Série intercensitária dos óbitos por sexo e idade, em grupos quinquenais, excluído um grupo de causa de morte	1900-1969 (capital) e 1933-1969 (interior)	Código, sexo, idade, data de referência, óbitos

Fonte: Albino (2024a).

Tabela 2. Estatísticas populacionais dos municípios do estado de São Paulo, 1836-2022.

Variáveis	Descrição	Fonte
codigo	Código composto de 7 dígitos, sendo os dois primeiros referentes ao código da Unidade da Federação, de acordo com a Tabela de Códigos de Municípios do IBGE	IBGE (2011)
mun	Nome atualizado do município	IBGE (2011)
codigo_perodo	Código de identificação na data do Censo correspondente	IBGE (2011)
nome_perodo	Nome do município na data do Censo correspondente	IBGE (2011)
ano_criacao	Ano de criação do município	IBGE (2011)
ano_misericordia	Ano de fundação da Santa Casa de Misericórdia	Khoury (2004)
zona	Região Demográfico-Econômica do Estado de São Paulo	Holloway (1984) e Colistete (2015)
ref	Data de referência da estimativa de população	-
ano	Ano de referência	-
pop	População total estimada, de acordo com os levantamentos disponíveis. Informações intercensitárias estimadas mediante interpolação linear.	Bassanezi (1998), São Paulo (1936) e Censos Demográficos (1920 a 2022)
pop_livre	População livre recenseada em 1872	Recenseamento do Império do Brasil de 1872 (1874)
pop_escrava	População escravizada recenseada em 1872	Recenseamento do Império do Brasil de 1872 (1874)
pop_rural	População rural estimada a partir de diversos levantamentos	Recenseamento demographico, escolar e agrícola-zootécnico do Estado de São Paulo: 20 de setembro de 1934 (1936) e Censos Demográficos do IBGE (1940 a 2010)
pop_urbana	População urbana estimada a partir de diversos levantamentos	Recenseamento demographico, escolar e agrícola-zootécnico do Estado de São Paulo: 20 de setembro de 1934 (1936) e Censos Demográficos do IBGE (1940 a 2010)
pib	Produto Interno Bruto municipal, em milhares de reais, a preços de 2010	Ipeadata (2023)
nasc_vivos	Nascidos vivos no ano de referência	Fundação Seade (2000)
nasc_mortos	Natimortos no ano de referência	Fundação Seade (2000)
casamentos	Total de casamentos no ano de referência	Fundação Seade (2000)
obitos_inf	Óbitos de crianças menores de um ano de idade no ano de referência	Fundação Seade (2000)
obitos_inf_nr	Óbitos de crianças menores de um ano de idade não residentes em São Paulo atribuídos ao Interior	Berquó e Gonçalves (1974)
obitos	Óbitos gerais, excluídos os natimortos	Fundação Seade (2000)
obitos_nr	Óbitos de não residentes no município de São Paulo atribuídos ao interior	Berquó e Gonçalves (1974)
hospitais_1929	Número de hospitais no ano de 1929	São Paulo (1932)
agua_esgoto	Existência de sistema de água e esgotos no ano de 1929	São Paulo (1932)

Fonte: Albino (2024b).