

Revista Vectores de Investigación
Journal of Comparative Studies Latin America
ISSN 1870-0128
ISSN online 2255-3371

Heitor Pinto de Moura Filho
**MODELAGEM DEMOGRÁFICA: VERIFICAÇÃO E
DESCOBERTA O TRATAMENTO HISTORIOGRÁFICO
DE REGISTROS DE ÓBITOS
DEMOGRAPHIC METHODOLOGY**

Vol. 6 No. 6, 55-68 pp.

MONOGRAFÍA METODOLOGÍA Y ANÁLISIS

**Heitor Pinto
de Moura Filho**

*Associação
Biblioteca de
Cultura, Brasil*

*Palabras claves:
modelo
demográfico,
tratamiento
historiográfico,
natalidad-
mortalidad*

Modelagem demográfica: verificação e descoberta o tratamento historiográfico de registros de óbitos

DEMOGRAPHIC METHODOLOGY

**ENVIADO 17-7-2006 REVISADO 3-12-2012
ACEPTADO 17-1-2013**

RESUMEN Este texto discute problemas encontrados na análise de estatísticas de óbitos, em especial no cálculo de taxas de mortalidade geral ou específica, com base em registros demográficos incompletos. Propõe conceitos e procedimentos de tratamento dos dados destinados a explicitar os limites de validade desses dados e a esclarecer a abrangência das hipóteses empregadas para corrigir erros ou estimar informações

desconhecidas, dentro de uma perspectiva de buscar a aproximação possível entre as práticas metodológicas de historiadores e de demógrafos.

ABSTRACT This paper discusses problems found when analysing death statistics, especially when calculating general or specific death rates based on incomplete demographic sources. It proposes concepts and methodological procedures destined to expose the limits of the validity of the data, and to clarify the scope of hypothesis employed in correcting errors or in estimating unknown information, within a perspective of searching for the possible approximation between the methodological practices of historians and those of demographers.

1 Tratamento historiográfico de registros de óbitos¹

Este texto discute problemas encontrados na análise de estatísticas de óbitos, em especial no cálculo de taxas de mortalidade geral ou específica, com base em registros demográficos incompletos. A maior parte dessas questões

¹ Uma versão resumida deste trabalho foi apresentada no XV Congresso da AHILA-Associação de Historiadores Latinoamericanistas Europeus, em Leiden, 26 a 29 de agosto de 2008, no Simpósio N05 "Metodologia y análisis histórico y económico sobre América Latina".

também se aplica a registros e taxas de natalidade e nupcialidade. O tema vem adquirindo interesse crescente na medida em que os pesquisadores dispõem atualmente de poderosas ferramentas eletrônicas de guarda, classificação e tratamento de dados, associado ao fato de se multiplicarem, no Brasil e no mundo, pesquisas primárias em arquivos paroquiais e cartoriais. Neste dinâmico contexto de aceleração da produção de conhecimento, entretanto, há elementos que restringem o melhor aproveitamento do esforço coletivo.

Em primeiro lugar, presenciamos a manutenção de um fosso metodológico entre historiadores e demógrafos. Historiadores se valem, por um lado, de estatísticas gerais de baixa confiabilidade e, por outro, de seus detalhados, porém limitados e potencialmente pouco representativos, levantamentos nas fontes primárias. Demógrafos, ao contrário, dispõem de modernas estatísticas censitárias e de pesquisas pontuais abrangentes, com ampla cobertura da população total e com a possibilidade de operar sobre registros individuais, a partir de sofisticadas técnicas matemáticas e de tratamento de informação. Além das óbvias diferenças na qualidade dos dados manuseados, também há distintas preferências por abordagens metodológicas mais matematizadas, pelos demógrafos, e mais voltadas para a história social e da população, pelos historiadores, conforme apontado por Bacelar, Scott e Basanezzi (2005).

Em segundo, há importantes barreiras de informação entre as pesquisas originais e o público especializado que podem eventualmente comprometer o grau de confiança que a comunidade acadêmica se sente confortável em atribuir às conclusões publicadas. Estas barreiras decorrem da rara publicação dos dados originais e, mais relevante, da infreqüente discussão quanto à validade historiográfica dos dados. Nem sempre os detalhes do processo de construção dos dados, de sua combinação por tratamentos numéricos e das hipóteses empregadas para substituir dados desconhecidos são apresentados aos leitores. Na busca da aproximação possível entre as práticas metodológicas de historiadores e de demógrafos, cremos que, mesmo diante de fontes falhas e incompletas, cabe ao historiador procurar alcançar o maior rigor lógico compatível com as possibilidades de seus dados. Ao se impor padrões metodológicos potencialmente inatingíveis, porém explicitados, o historiador estará fortificando seu próprio trabalho, estará esclarecendo aos leitores suas hipóteses e métodos, e ainda estará iluminando caminhos para pesquisas futuras. No caso dos registros de óbitos em contexto de dados incompletos, propomos alguns conceitos e procedimentos de tratamento dos dados destinados a explicitar os limites de sua validade e a esclarecer a abrangência das hipóteses empregadas para corrigir erros ou estimar informações desconhecidas.

2 O que informa uma série de registros de óbitos?

É extremamente variável a qualidade das fontes demográficas quantitativas (ou das que podem gerar informações quantitativas) e, conseqüentemente, também é variável a qualidade dos dados que nelas podem ser recolhidos. Ao examinarmos um clássico da demografia histórica francesa –como *Tourouvre-au-Perche aux XVII^e et XVIII^e siècles*, do *quebequois* Hubert Charbonneau

(1970)– vemos dados minuciosamente reunidos sobre uma pequena paróquia, porém cobrindo mais de dois séculos e com cobertura quase total dos eventos demográficos: nascimentos(batismos), casamentos e óbitos(enterras). Além da qualidade intrínseca dos dados (considerando-se aí preservação dos originais, poucas omissões, precisão dos registros, anotação sobre existência de assinaturas etc.), essa regularidade de informações autoriza técnicas de análise praticamente impossíveis em contextos de períodos curtos ou com deficiência de cobertura dos eventos demográficos. E mais, o fato de haver informações simultâneas sobre todos os tipos de eventos demográficos, principalmente de nascimento e óbitos, é particularmente útil, pois permite reconstituir a população com alto grau de confiabilidade, obtendo assim séries paralelas não só para os eventos individuais, como para as diversas taxas demográficas que relacionam duas séries de eventos ou uma série de evento ao contingente de certo grupo populacional². Esta característica de completude das fontes, disponível em alguns países, oferece aos seus estudos demográficos uma precisão que talvez só seja possível no Brasil para paróquias específicas, durante curtos períodos³.

É interessante lembrarmos aqui algumas diferenças entre o uso de registros de eventos demográficos para a reconstituição de famílias ou de redes de relacionamentos, em oposição a seu uso para a reconstituição agregada de populações. Para a reconstituição de grupos familiares ou de interrelacionamentos, a ausência de registros relativos a certos períodos reduz a abrangência e a representatividade da pesquisa, mas não a invalida. Como exemplo, poderíamos supor uma lacuna de registro de vários anos, mas que não atingisse os eventos vitais dos membros de certa família. Em situação menos propícia, no caso em que tais lacuna escondam o nascimento ou óbito de membros de famílias já conhecidas, a ficha dessa família poderá ser deixada de lado ou poderão ser empregadas técnicas alternativas (como o cômputo a partir do último eventos vital registrado), sem afetar a análise das fichas restantes. No caso da reconstituição agregada, no entanto, qualquer lacuna tem efeito direto sobre a precisão do resultado, possivelmente invalidando, a menos de estimativas substitutivas, o uso dos dados daquela paróquia naquele período.

O que podemos buscar nas incertas e irregulares fontes brasileiras? O caso dos registros de óbitos, crescentemente estudados, apresenta interesse especial na medida em que são mais precisos e talvez apresentem abrangência mais geral do que os registros de nascimento. Por reduzir-se (até a consolidação do registro civil) ao registro de batismo, o registro de nascimento foi afetado pela altíssima mortalidade neonatal, pela dispersão geográfica e

² Maria Luiza Marcílio explicita esse objetivo: “A Demografia Histórica começou (...) testando o método da Reconstituição de famílias de uma paróquia ou de conjunto de paróquias, e com a finalidade expressa de calcular taxas gerais e especiais de fecundidade, de nupcialidade e de mortalidade, para populações pré-estatísticas (...). (Marcílio, 2004)”.

³ Alguns exemplos: Canadá (Henripin, 1954), Noruega (Dyrvik, Mykland, *et al.* 1976), Bélgica (Alter, Gutmann, 2005), Islândia (Guttormsson, Gaoarsdóttir, 2002), Islândia, Dinamarca e Suécia (Hansen, 1975).

talvez até pelo caráter “menos público” do evento a ser anotado – o parto⁴.

O problema básico para a utilização proveitosa dos registros de óbito decorre da fragilidade argumentativa de uma série isolada de registros. Em si, ela só informa isso mesmo: a série de óbitos, classificados por data e sexo do falecido, possivelmente complementada por sua idade aproximada, sua condição social, *causa mortis* e outras informações esparsas. É preciso insistir no fato de que terem morrido mais homens do que mulheres, mais crianças do que adultos, mais escravos do que livres etc. não autoriza *a priori* qualquer conclusão sobre a mortalidade diferencial entre essas categorias⁵. Só sabemos que houve tantas mortes de um grupo e tantas de outro. Mesmo dispondo de uma série mais longa de óbitos, a evolução desses números tampouco nos permite qualquer conclusão sobre a evolução da mortalidade nas respectivas categorias! Essa impossibilidade é uma consequência lógica da ausência de referencial. Ou seja, saber que, em certo período, o número de óbitos de escravos masculinos foi maior do que o de homens livres só nos permite concluir que os escravos masculinos sofriam de maior mortalidade caso também saibamos, por exemplo, que a população de um e de outro grupo tinham tamanhos próximos.

O indicador demográfico fundamental, a taxa de mortalidade geral, requer que se conheça, além da série de óbitos, a série paralela com o tamanho da população na qual ocorreram esses óbitos, a “população em risco”. Até no caso de uma série temporal de óbitos, não podemos concluir grande coisa sobre a evolução da mortalidade sem sabermos qual a evolução paralela do contingente populacional corespondente, pois, por exemplo, numa população com grande movimentação de imigrantes ou de emigrantes, o fato de num ano haver maior número de óbitos não significa que a mortalidade tenha aumentado, podendo, até, ter tido uma queda importante, caso tenha havido grande incremento da imigração de um período para outro.

E há ainda outras dificuldades em torno da questão, além da tradicional preocupação com emigração e imigração permanentes. Supondo que de fato conheçamos qual a população da paróquia (ou mesmo da cidade) para qual dispomos da série de registros de óbitos, a comparação direta entre óbitos registrados e população conhecida pode ser bastante falha, seja por omissão de registros, seja pela movimentação da população. Um ótimo exemplo dessa dificuldade diz respeito à situação ocorrida em meados do século XIX, quando a proibição a enterros em igrejas e a obrigação de realizá-los nos cemitérios públicos levou ao transporte de defuntos para serem enterrados nas capitais (até de Petrópolis para o Rio de Janeiro). Mesmo antes dessa época, o trânsito de doentes de outras cidades para hospitais nas capitais, onde viriam a falecer e, portanto, serem enterrado, já criava distorções aos números de óbitos registrados relativamente à população conhecida.

Outro exemplo de distorção entre número de óbitos e população residente

⁴ Maria Luiza Marcílio (2004) e Sérgio Nadalin (2004: 39-61) discutem a abrangência dos registros paroquiais no Brasil.

⁵ Uma pequena exceção poderia ser aberta para *causa mortis*, pois, a menos de problemas classificatórios, de identificação e das omissões, a mortalidade devida a certa doença, relativamente às demais, seria proporcional ao número de óbitos registrados com aquela causa, mantida a distribuição etária.

aconteceu durante as grandes secas do Nordeste, na segunda metade do século XIX, quando a movimentação de retirantes, que permaneciam fora de suas cidades por muitos meses e lá adoeciam e morriam, inflou as estatísticas de óbitos urbanos em desproporção a suas populações. As épocas de crises de mortalidade em decorrência de epidemias oferecem mais uma situação em que o cálculo de taxas de mortalidade fica dificultado. A alta morbidade e mortalidade entre marinheiros – estrangeiros e brasileiros de outras regiões – nos portos brasileiros afetados, possivelmente em proporções bem maiores do que entre a população nativa da cidade, precisaria ser descontada das estatísticas de óbitos para calcularmos taxas gerais correspondentes à população registrada. Para a cidade do Rio de Janeiro em meados do século XIX, há indicações que não só os marinheiros, mas toda a população de imigrantes tinha mortalidade bem maior do que a dos brasileiros livres, porém menor do que a dos escravos⁶. Como um último exemplo, as situações de guerra ou outras formas de violência coletiva surgem como uma combinação de distorções causadas por mobilidade e mortalidade, seja no local de origem das tropas, que perdem população, seja no local dos conflitos, que ganham em óbitos.

Assim, dado um conjunto de registros de óbitos, impõe-se o problema de conciliar estes dados com a população de onde se teriam originado. Para dificultar mais um pouco a estimação dos dados, não se pode, *a priori*, supor que os registros de certa freguesia, município ou província correspondam necessariamente à população conhecida dessas mesmas unidades geográficas, pois existe grande probabilidade de que este conjunto de registros não inclua todos os óbitos ocorridos nesta população e, além dessas omissões, sabemos que os levantamentos populacionais disponíveis provavelmente não teriam contado com precisão a efetiva população da região.

Conhecer as dificuldades não impede, contudo, a realização de pesquisas ou de analisar dados reconhecidamente imperfeitos. No caso dos registros de óbitos, temos uma relação fundamental para nos auxiliar na avaliação da correção e na estimação dos dados. Esta relação define o número de óbitos, por faixa etária, em função da população existente naquela faixa, multiplicada pela taxa de mortalidade específica da faixa etária. Na medida em que podemos supor que a distribuição de taxas de mortalidade específicas por idade da população em estudo e da própria amostra possam ser aproximadas por alguma distribuição conhecida – a distribuição de uma tábua de mortalidade calculada sobre os dados censitários, por exemplo – temos um importante instrumento para estimar diversos dados desconhecidos. Antes de propor um método de estimação, contudo, é essencial dispormos de uma terminologia para descrever os vários conjuntos populacionais com que estamos lidando.

3 Relações teóricas e relações entre dados coletados

Estimar com precisão grupos populacionais com equações que relacionam população, por faixa etária, a óbitos, também por faixas etárias, através de

⁶ Sobre uma população de 78.676 imigrantes no Rio de Janeiro, entre 1865 e 1870, Maria Bárbara Levy (1973: 650-1) calculou a mortalidade dos estrangeiros em 53/1000, dos brasileiros livres em 43/1000 e a dos escravos em 67/1000.

uma distribuição de mortalidade específica, pressupõe a disponibilidade de uma quantidade de dados dificilmente encontrada na prática, isto é, recenseamentos anuais da população residente e dos movimentos migratórios, e completo registro civil dos óbitos. No caso mais comum de pesquisa, possivelmente conheceremos somente uma estimativa para a população da região, num momento qualquer dentro do período de estudo ou simplesmente próximo a ele, combinado com dados referentes a registros de óbitos em paróquias da região. Estes dados coletados apresentam diferenças importantes com relação às variáveis teóricas requeridas.

Devemos lembrar que um registro de óbito para ter utilidade num estudo demográfico deve conter, minimamente, além da identificação da fonte, indicações de data do óbito, sexo e idade do falecido. Embora muitos registros indiquem outros dados, como *causa mortis*, condição civil, nacionalidade ou naturalidade, proprietário (no caso de escravos) e endereço (preciso ou aproximado) do falecido, ou, inversamente, não contenham indicação de sexo ou, principalmente, de idade, as quatro informações acima são aquelas necessárias para o uso demográfico do registro.

Todo o trabalho descrito aqui requer que se possa associar o conjunto de registros de óbitos disponível a uma população cuja características demográficas mínimas também sejam conhecidas. Só assim teremos a possibilidade de analisar adequadamente a mortalidade desta população (seja num nível geral ou num nível mais detalhado, incluindo, por exemplo, *causa mortis*). Existem diversas maneiras de definirmos essa população na qual os óbitos ocorreram e cada opção tem certas vantagens e desvantagens com relação ao objetivo prático de analisar a amostra e estimar as taxas de mortalidade. O pesquisador deve examinar com cuidado as opções à mão, pois a solução mais imediata, de adotar a população nominalmente associada à base geográfica dos registros, pode não ser necessariamente a melhor. Dentre as possíveis populações de referência, é interessante situarmos duas populações limites. Entre esses limites, temos uma extensão da população falecida, que chamaremos de população mínima, e a população que poderíamos considerar como a mais provável, definida por critérios administrativo-geográficos.

- 1 POPULAÇÃO FALECIDA. Seria a população composta unicamente pelas pessoas que, ao falecer, passaram a compor o conjunto de óbitos conhecidos. É, logicamente, a menor população que poderia originar os óbitos conhecidos, o que constitui uma definição com a vantagem da absoluta simplicidade, pois conhecemos cada indivíduo da população. Não corresponde, no entanto, a uma população “demograficamente real”, isto é, uma população com nascimentos, casamentos e mortes, em que os óbitos atingissem anualmente somente uma parcela pequena da população e na qual faça sentido se estudar uma dinâmica demográfica, determinada por sua composição etária e por sexo e pelas taxas específicas de fecundidade e mortalidade⁷.

⁷ A análise de uma população exclusivamente a partir da própria amostra de falecidos é típica da paleo-demografia, onde os pesquisadores se esforçam por extrapolar os poucos dados contidos em pequenas amostras, nas quais se procura estimar a idade ao morrer de cada esqueleto, osso ou, até, fragmento de osso. Tais investigações, combinadas com hipóteses de

- 2 POPULAÇÃO MÍNIMA OU RECONSTITUÍDA. Supondo que tivéssemos extraído todos os registros de óbitos à mão de uma reconstituição de famílias, poderíamos considerar o conjunto mínimo de famílias que incluisse todos os óbitos da amostra como a população mínima que também apresentasse características demográficas “reais”. Esta definição é a que autoriza a maior precisão no cálculo de indicadores demográficos, embora se deva avaliar cuidadosamente sua representatividade frente à população total da região. As taxas demográficas apresentadas nos estudos que cobrem pequenas regiões durante longos períodos podem ser consideradas, portanto, as mais fidedignas no sentido de serem derivadas de um quociente com numerador e denominador com alto grau de associação⁸.
- 3 POPULAÇÃO MÁXIMA. Sem considerarmos o máximo absoluto, definido pela população global, haveria diversos outros recortes máximos interessantes. Dentre estes, podemos situar a população do país ou de uma grande macroregião, como a “máxima prática”. Esta população deveria nos garantir maior controle sobre as flutuações migratórias intra-regionais. Neste caso, a amostra de óbitos disponível possivelmente representaria uma quantidade pequena, ou mesmo muito pequena, relativamente ao total de óbitos da população. O uso desta população máxima poderá ser necessária, no entanto, em pesquisas nas quais se estudam amostras com ampla cobertura geográfica.
- 4 POPULAÇÃO PROVÁVEL. Seria a população mais fortemente associada à fonte dos registros: para os livros de registro de óbitos de uma paróquia, a população daquela paróquia. Eventualmente talvez seja necessário fazermos ajustes nas estatísticas disponíveis para melhorar a estimativa desta população, de forma a levar em conta movimentos migratórios ou a corrigir discrepâncias na estrutura etária ou por sexo representada nas estatísticas. Nesse caso, sendo o conjunto de registros de óbitos suficientemente representativo, podemos até supor que os óbitos não registrados sejam pouco relevantes. A principal vantagem desta escolha é trabalhar com uma população demograficamente real, porém na qual o número total de óbitos seria da mesma ordem de grandeza do número de nossa amostra. Cada contexto, no entanto, precisará ser examinado individualmente para avaliarmos a confiabilidade da cobertura do registros de óbitos relativamente ao total de óbitos da população provável.

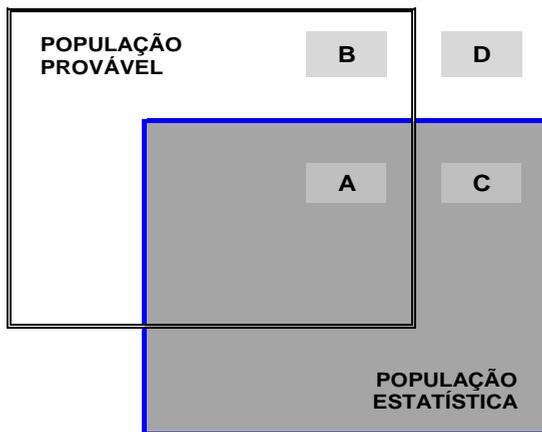
É essencial considerarmos outra dificuldade inerente à estimação demográfica em situação de estatísticas incompletas. Por mais que possamos escolher a população-base próxima da população efetivamente associada aos óbitos conhecidos, esta população-base não será exatamente a população para a qual dispomos de estatísticas censitárias.

população estacionária, produziram até estimativas de “tábuas de mortalidade”. Wood, *et al.*, discutem criticamente a validade desses métodos (Wood, Holman, *et al.*, 2001).

⁸ Alter e Gutmann mostram como, a partir de 1846, todos os eventos demográficos da população belga foram anotados, individualizados por família, num registro geral de população, o que permite a detalhada reconstituição de praticamente toda a população do país, bem como o cálculo de taxas gerais e específicas de alta precisão (Alter, Gutmann, 2005: 528-9).

Para analisar com maior precisão as diferenças entre essas populações e, portanto, entre as amostras de óbitos à mão e o conjunto de óbitos associados a estas populações, necessitamos introduzir a população que chamaremos de “população-estatística”, a população sobre a qual temos informações estatísticas (e que é freqüentemente confundida com a população-base ou -provável). Vemos, na figura abaixo, que a interseção dessas duas populações gera quatro conjuntos populacionais, A, B, C e D, que serão objeto de nossa análise.

Figura 1. Possíveis combinações de população-provável e população-estatística



Fuente: Elaboración propia

Estão representados na figura os dois conjuntos de população, a provável e a estatística, mostrando seus subconjuntos possíveis: (A) a população provável e incluída na estatística; (B) a população provável não incluída na estatística; (C) a população estatística mas que não corresponde à provável; e (D) o conjunto de pessoas que nem pertencem à população provável nem à estatística. Lembramos alguns exemplos: pertencem ao subconjunto B, da população provável, porém não computada estatisticamente, entre outros, imigrantes entrados na região desde o momento da estimativa de população e que, portanto, podem ter morrido ali; ou crianças, escravos ou jovens em idade militar omitidos nos levantamentos provinciais. Pertencem ao subconjunto C, da população-estatística, porém não incluída na população-provável, os emigrantes saídos da região entre o momento da estimativa e o do registro de óbitos, bem como qualquer população fictícia indevidamente incluída nos levantamentos.

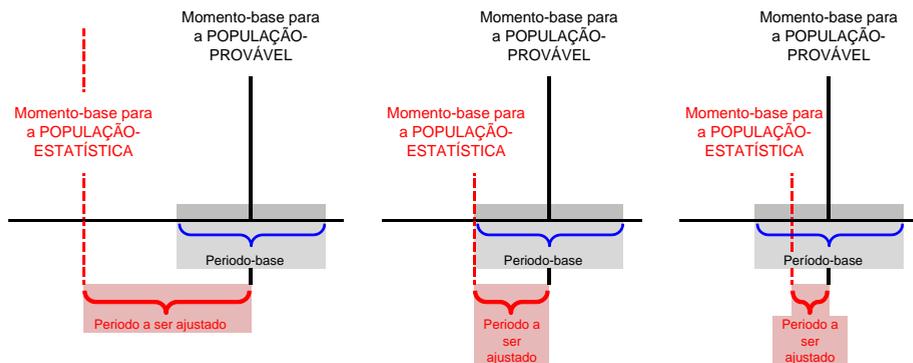
O principal objetivo desta classificação é termos claro, seja conceitualmente, seja na prática, relativamente aos dados que estamos manipulando, qual a representatividade dos registros de óbitos coletados e também das estatísticas sobre a população-provável, o que nos permitirá avaliar a validade das operações matemáticas e estatísticas que podemos vir a fazer sobre esses dados. Para atingir esse objetivo, precisamos introduzir alguns conceitos importantes para o desenvolvimento de estimativas sobre dados incompletos. O primeiro conceito é o de “período-base” dos óbitos conhecidos, ou seja, o período (anual, pluri-anual ou mesmo infra-anual) durante o qual o-

correram os óbitos anotados. Outro conceito é o de “população-base”, significando a população “demograficamente real”, (conforme comentado acima) da qual se supõe serem originados os óbitos de certo conjunto de registros. Numa interpretação demográfica mais matematizada, isso significará que, ao multiplicarmos a população de cada faixa etária desta população-base pelas taxas de mortalidade específica da população (padronizadas para a duração do período-base), chegaremos ao total de óbitos ocorridos durante este período-base. Para manter essa relação dentro da convenção usual de considerar a população média para o cálculo das taxas referentes a um período como aquela existente no meio deste período, consideraremos que a população-base seja aquela existente no meio do período-base.

Para estudar as diferenças que, *a priori*, podemos estabelecer entre a população-provável (aqui associada à população-base) e a população-estatística, é necessário levarmos em conta a linha do tempo, distinguindo as possíveis posições da população-estatística relativamente à população-provável e ao período-base.

A figura acima ilustra as três possibilidades de população-estatística anterior à população-provável. Considerando os três casos análogos com a população-estatística referenciada a momento posterior ao da população-provável, mais o caso de coincidência dos momentos-base das duas populações, teremos sete situações potenciais. O período entre o momento de referência da população-estatística e o da população-provável será sempre o período a ser ajustado, isto é, aquele cujos eventos demográficos precisaremos analisar para entender essas diferenças.

Figura 2. Possíveis situações do período de ajuste sendo a população-estatística anterior à população-provável



PE fora do período-base

PE no limite do período-base

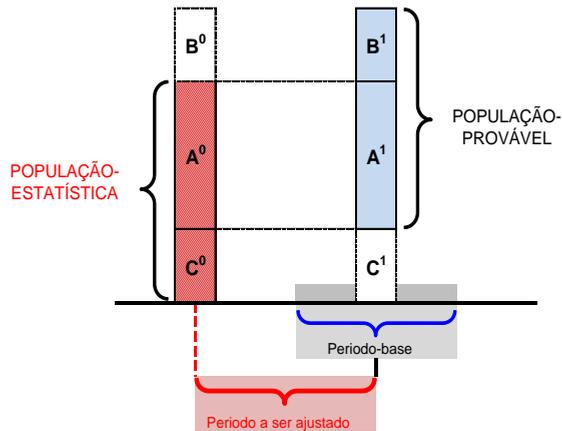
PE dentro do período-base

Fuente: Elaboración propia

Precisamos, agora, localizar nesse esquema temporal os conjuntos A, B, C e D apresentados na Figura 1. O problema metodológico será conciliar esses subconjuntos, considerando as modificações demográficas ocorridas em cada um durante o período de ajuste. Ao empregar essa tipologia de subconjuntos populacionais, estaremos mais bem preparados para conciliar numerador

(óbitos durante o período-base) e denominador (população-base) ao calcular taxas de mortalidade.

64 **Figura 3. Subconjuntos da população-estatística e da população-provável, dispostos no tempo (caso da PE anterior a PP e fora do período-base)**



Fuente: Elaboración propia

4 Estimación da população-provável a partir da população-estatística

Desejamos exprimir a população-provável (PP) em termos da população-estatística (PE) e empregar o máximo de informação contida na amostra de óbitos conhecida, de modo a chegar a uma estimación da taxa de mortalidade geral da população. Este problema consiste: 1) no ajuste temporal da população do conjunto A, das pessoas que pertencem tanto à PE quanto à PP; 2) na estimación da população do conjunto B, das pessoas que devem ser somadas à PE; e 3) na estimación da população do conjunto C, dos que devem ser retirados da PE. Em termos dos conjuntos da figura acima, queremos estimar A^1 mais B^1 , conhecendo somente A^0 , C^0 e amostras dos eventos demográficos ocorridos durante o período-base.

O conjunto A^1 se compõe das pessoas pertencentes à PP que já pertenciam a PE no momento 0 e sobreviveram até o momento 1, tendo, além disso, permanecido na região. Também devemos considerar em A^1 os acréscimos a este grupo, ou seja, os nascidos durante o período de ajuste a mães do grupo A^0 e que sobreviveram na região até o momento 1.

Compõem o conjunto B^1 , das pessoas que, no momento 1, pertencem à população-provável (PP), mas não constaram da população-estatística (PE):

- A Pessoas de fato pertencentes à PP, que foram omitidas da PE no momento 0 e, além disso, sobreviveram até o momento 1 (certas faixas etárias ou grupos ou ainda residentes de áreas indevidamente excluídas);
- B Os nascidos, durante o período de ajuste, a mães do grupo anterior e que sobreviveram na região até o momento 1;

- C Pessoas que entraram na PP durante o período de ajuste e sobreviveram, na região, até o momento 1; e
- D Nascidos a mães da categoria anterior, desde o momento em que estas entraram para a PP, e sobreviveram na região até o momento 1.

Compõem o conjunto C^0 , das pessoas incluídas na PE, porém que não pertenciam à PP no momento 0:

- A A população fictícia incluída na PE;
- B Populações transitórias incluídas em PE;
- C Inclusões geográficas indevidas.

Com relação aos óbitos, partimos de um conjunto de registros conhecidos, definindo nossa população-provável como sua correspondente população-base, aquela na qual esses óbitos teriam ocorrido. Como é certo que, além dos óbitos conhecidos pela amostra, tenha havido outros óbitos nesta população-base, chegamos ao conceito de “óbitos complementares” a nossa amostra, isto é, os óbitos ocorridos na população-base, porém ausentes de nossa amostra. Complicando nossas estimativas, devemos reconhecermos que tanto os óbitos de nossa amostra, como seus óbitos complementares, podem pertencer a cada uma das populações A, B ou C.

Descrevemos, a seguir, uma série de procedimentos preparatórios que deverão nos aproximar de um conhecimento mais preciso sobre os óbitos e sua respectiva população-base. Tal lista é indicativa, não exaure as possibilidades e sempre deverá ser adaptada ao caso em estudo.

- 1 Reunir as estatísticas disponíveis sobre possíveis populações-estatísticas e sobre os eventos demográficos, referentes à região pesquisada e em torno do período amostral.

Um cuidado preliminar a este levantamento é entender o recorte geográfico-administrativo de cada estatística e os procedimentos utilizados em sua construção. Principalmente em estudos cobrindo períodos mais longos isto se torna essencial, embora, mesmo para curtos períodos, modificações administrativas podem ter sido relevantes à determinação dos contornos geográficos dos dados.

Em segundo lugar, é fundamental preparar os dados de forma a eliminar incoerências e erros de transcrição. Como exemplo, mencionamos os resultados do Recenseamento Geral de 1872 que, apesar de disponíveis no site da Biblioteca do IBGE, requerem algum trabalho para serem padronizados. É preciso, inicialmente, digitar os valores numa planilha e verificar que esses dados mantêm sua coerência interna, ou seja, que as somas de linhas e colunas correspondam aos totais publicados. Por este procedimento devemos eliminar (ou reduzir bastante) os erros de digitação devidos à dificuldade de leitura dos dados. Estes problemas dizem respeito à distinção entre certas duplas de dígitos: 3 e 8, 6 e 0, 9 e 0, 1 e 4, 1 e 7, 3 e 5.

Mesmo sendo feita uma digitação cuidadosa, é surpreendente a quantidade de erros que aparecem na primeira passada. Esses erros precisarão ser resol-

vidos, num “sudoku paleográfico”, pela revisão e alteração sucessiva dos dígitos mais propensos a erros, até termos todas as linhas e colunas com somas iguais aos totais impressos.

- 66
- 2 Ajustar a população-estatística. Há diversos ajustes possíveis à PE, seja pela análise das razões de masculinidade, seja pela composição etária ou por condição social. Os ajustes propostos por G. Mortara aos dados censitários, por exemplo, se baseiam principalmente sobre descompasso nas distribuições etárias, que sugerem erros de classificação ou omissões (Mortara 1940c, a, b, d, 1941). Tais procedimentos se tornaram técnicas reconhecidas de ajustes censitários⁹. O resultado desses ajustes nos fornecerão estimativas para certos grupos omitidos da PE (população B) ou para grupos indevidamente incluídos nela (população C).
 - 3 Completar as estimativas de B e C. Sobre as primeiras estimativas dos conjuntos B e C obtidas na etapa anterior, podemos agregar informações adicionais, de outras origens, que auxiliem na sua estimação.
 - 4 Preparação da amostra de óbitos. Há diversos procedimentos preliminares importantes a serem realizados sobre a amostra de óbitos. Tais procedimentos, mesmo trabalhosos, devem ser realizados sobre amostras de qualquer tamanho, sob pena de introduzirmos “lixo” em nossos dados. São eles:
 - A Padronizar todos os registros quanto a sua numeração, data do óbito, indicação de idade, sexo, condição social etc. – É comum aparecerem “idades” indicadas como “meses”, “ao nascer”, “1 ano” etc., o que impossibilita a correta classificação dos registros nas planilhas. Semelhantemente, podemos encontrar na coluna “sexo”, indicações como “crioula”, “ingênuo”, “desconhecido”, “inocente” etc.
 - B Criar uma gama de códigos possíveis que sejam adequados aos dados – Usar 4 classificações para “sexo”, por exemplo, pode ser necessário para dar conta das opções de [masculino], [feminino], [óbitos de recém-nascidos e crianças sem especificação de sexo] e [óbitos em geral sem sexo não especificados]. Um caso especial desse problema é a escolha de classes etárias: anuais, quinquenais ou decenais.
 - C Procurar eliminar duplicidades – Este trabalho, embora subjetivo e dependendo do bom-senso do pesquisador, inevitavelmente consegue “limpar” a amostra.
 - D Rever todas as informações disponíveis que auxiliem no preenchimento dos registros indispensáveis (data do óbito, idade, sexo, condição social) – Muitas vezes não há informação sobre condição social, por exemplo, porém existem indicações assessórias de “liberto africano” ou do “proprietário”, ou falta menção ao sexo, porém existe indicação do nome do falecido, o que nos permite a correta classificação do óbito.
 - E Ordenar todos os registros por sexo e condição social – Esta classifica-

⁹ Ver, por exemplo, George Barclay, *Techniques of Population Analysis* (1958).

ção preliminar permite a análise dos registros ainda sem indicação em cada categoria, bem como reexaminar os conjuntos supostamente homogêneos à procura de discrepâncias de codificação e de classificação. Após esta “reescapagem”, a amostra pode ser reclassificada e estará pronta para a pesquisa, agora como um banco de dados organizado.

Concluído este tratamento preliminar, precisamos buscar estimativas para os três conjuntos A, B e C, conforme discutido acima. Para esta empreitada, propomos utilizar tábuas de mortalidade conhecidas, preferencialmente calculadas a partir da população-estatística. Tais distribuições de mortalidade específica poderão ser usadas em dois sentidos: a) aplicadas (por multiplicação) a uma população ou seu subconjunto, fornecerão estimativas para o total de sobreviventes no momento 1 e para o total de óbitos ocorridos durante o período de ajuste; ou b) aplicadas (como divisor) a um total de óbitos conhecidos, fornecerão sua população-base. O fato de realizarmos estes cálculos por faixa etária, dentro de cada grupo por sexo e condição social, e não por médias gerais de toda a população transforma a informação existente nas distribuições etárias em aproximações melhores para o total de cada subconjunto. Diante das inúmeras dificuldades e incoerências nas estatísticas, um cuidado adicional é aplicarmos parâmetros em grupos sociais tão semelhantes quanto possível àqueles de onde foram derivados.

BIBLIOGRAFÍA

- ALTER, George, GUTMANN, Myron (2005) "Belgian Historical Demography as Viewed from North America: Protoindustrialization, Fertility Decline, and the Use of Population Registers", *Revue Belge d'Histoire Contemporaine*, vol. XXXV, N° 4, 523-546, <http://www.flwi.ugent.be/btng-rbhc/pdf/BTNG-RBHC,%2035,%202005,%204,%20pp%20523546.pdf>
- BACELLAR, Carlos de Almeida Prado, SCOTT, Ana Sílvia Volpi, et al. (2005) "Quarenta anos de demografia histórica", *Revista Brasileira de Estudos de População*, vol. 22, N° 2, 339-350, <http://www.scielo.br/pdf/rbe-pop/v22n2/v22n2a09.pdf>
- BARCLAY, George (1958) *Techniques of Population Analysis*, New York, John Wiley & Sons.
- CHARBONNEAU, Hubert (1970) *Tourouvre-au-Perche aux XVII^e et XVIII^e siècles*, Paris, Presses Universitaires de France.
- DYRVIK, Ståle, MYKLAND, Knut, et al. (1976) *The demographic crises in Norway in the 17th and 18th centuries: Some data and interpretations*, Bergen, Department of History, University of Bergen.
- GUTTORMSSON, Loftur, GAOARSDOTTIR, Ólöf (2002) "The Development of Infant Mortality in Iceland, 1800-1920", *Hygiea Internationalis*, vol. 3, N° 1, 151-176.
- HANSEN, Hans-Oluf (1975) "Some age-structural consequences of mortality variation in pre-transitional Iceland and Sweden", *The great mortalities: methodological studies of demographic crises in the past*, Hubert Charbonneau, Larose, André, Liège, International Union for the Scientific Study of Population, 113-132.
- HENRIPIN, Jacques (1954) *La population canadienne au début du XVIII^e siècle*, Paris, Institut national d'études démographiques, Presses Universi-

taires de France.

LEVY, Maria Bárbara (1973) *Participação da população livre e escrava numa codificação sócio-profissional do Rio de Janeiro (1850-1870): Alguns aspectos*, Anais do VI Simpósio Nacional dos Professores Universitários de História, São Paulo, 639-658.

MARCILIO, Maria Luíza (2004) "Os registros paroquiais e a história do Brasil", *Varia Historia*, N° 31, 13-20.

MORTARA, Giorgio (1940a), "Estudos sobre a utilização do censo demográfico para a reconstrução das estatísticas do movimento da população do Brasil. I. Estimativa do número dos nascimentos", *Revista Brasileira de Estatística*, vol. I, N° 1, 7-35, [http://biblioteca.ibge.gov.br/\(Coleçãoodigital\)](http://biblioteca.ibge.gov.br/(Coleçãoodigital)).

- (1940b) "Estudos sobre a utilização do censo demográfico para a reconstrução das estatísticas do movimento da população do Brasil. II. Conjecturas sobre os níveis da natalidade e da mortalidade no Brasil no período 1870-1920", *Revista Brasileira de Estatística*, vol. I, N° 2, 229-242, [http://biblioteca.ibge.gov.br/\(Coleçãoodigital\)](http://biblioteca.ibge.gov.br/(Coleçãoodigital)).

- (1940c) "Estudos sobre a utilização do censo demográfico para a reconstrução das estatísticas do movimento da população do Brasil. III. Análise dos erros existentes nas distribuições por idade da população do Brasil, baseadas nos censos", *Revista Brasileira de Estatística*, vol. I, N° 3, 433-472, [http://biblioteca.ibge.gov.br/\(Coleçãoodigital\)](http://biblioteca.ibge.gov.br/(Coleçãoodigital)).

- (1940d) "Estudos sobre a utilização do censo demográfico para a reconstrução das estatísticas do movimento da população do Brasil. IV. Ensaio de ajustamento das tábuas de mortalidade brasileiras calculadas por Bulhões de Carvalho", *Revista Brasileira de Estatística*, vol. I, N° 4, 673-693, [http://biblioteca.ibge.gov.br/\(Coleçãoodigital\)](http://biblioteca.ibge.gov.br/(Coleçãoodigital)).

- (1941) "Estudos sobre a utilização do censo demográfico para a reconstrução das estatísticas do movimento da população do Brasil. V. Retificação da distribuição por idade da população natural do Brasil, contante dos censos, e cálculo dos óbitos, dos nascimentos e das variações dessa população no período 1870-1920", *Revista Brasileira de Estatística*, vol. II, N° 5, 39-89, [http://biblioteca.ibge.gov.br/\(Coleçãoodigital\)](http://biblioteca.ibge.gov.br/(Coleçãoodigital)).

NADALIN, Sérgio Odilon (2004) *História e demografia*, Campinas, ABEP.

WOOD, James, HOLMAN, Darryl, *et al.* (2001) "Mortality Models for Paleodemography", *University of Washington CSDE Working Papers*, vol. 1, N° 1, <http://csde.washington.edu/downloads/01-02.pdf>.