MAT02262 - Estatística Demográfica I

Características, eventos, proporções, taxas e probabilidades: taxas e probabilidades

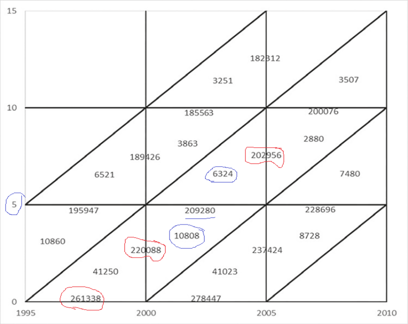
Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

Universidade Federal do Rio Grande do Sul
Instituto de Matemática e Estatística
Departamento de Estatística

Porto Alegre, 2023

Introdução

Relembrando



Relembrando

- Como o gráfico anterior, o diagrama mostra que **261.338 pessoas nasceram entre 1995 e 1999** e que destes **220.088 ainda estavam presentes no 1º de janeiro de 2000** e **202.956 no 1º de janeiro de 2005**.
- Mas além disso, este último gráfico mostra que **10.808 pessoas desta coorte morreram ou migraram antes do seu quinto aniversário**, de modo que se celebraram 209.280 quintos aniversários.
  - Os demais **6.324 morreram ou migraram antes do 1º de janeiro de 2005**, mas já tendo mais de 5 anos.

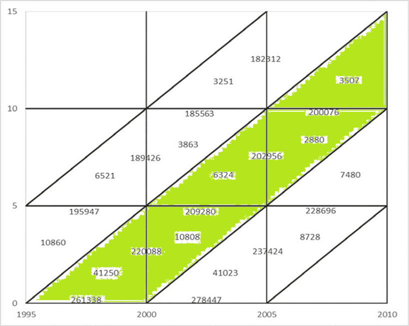
Taxas e probabilidades

↬ A contagem do número de eventos relevantes para quantificar a intensidade de uma variável de fluxo.

- Para construir um indicador de intensidade, esta quantidade de eventos precisa ser relacionada com **um denominador** que de alguma forma representa o **número potencial de eventos** que poderiam ter acontecido.
- O conceito mais correto para expressar este número potencial é o **número de anos-pessoa**[1] vividos pela população dentro da área relevante do diagrama de Lexis, ou seja, **o comprimento conjunto**, dentro da área, **de todas as linhas vitais** que cruzam a área de forma completa ou parcial.
  - Ou seja, é **soma dos tempos vividos** (em anos) por cada componente da população.

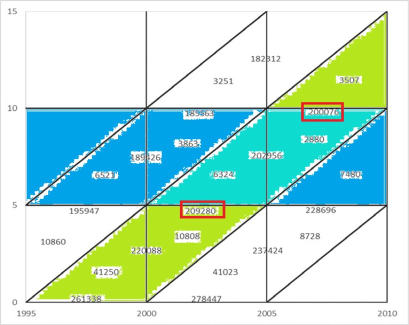
Taxas e probabilidades

⇝ Esta ideia é mais fácil de entender no caso de uma coorte como a **coorte de nascimentos de 1995-1999** no gráfico a seguir.



Taxas e probabilidades

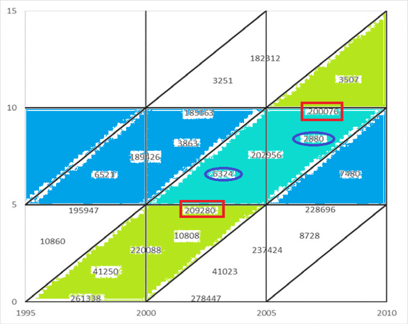
- No losango que define a vivência desta coorte para as idades de 5-9 anos, inicialmente houve 209.280 pessoas, das quais no final sobraram 200.076.



↬ Cada uma destas pessoas contribuiu **5 anos** ao conjunto de **anos-pessoa** vividos pela coorte, ou seja, um total de $5 \times 200.076 = 1.000.380$ anos.

Taxas e probabilidades

- Mas também houve $6.324 + 2.880 = 9.204$ pessoas que **não completaram o intervalo** e cujas linhas vitais pararam em algum lugar intermediário, seja porque morreram ou porque emigraram.



Taxas e probabilidades

- Muitas vezes se supõe que estas pessoas contribuíram a metade do período, ou seja $2,5$ anos cada uma, de modo que o número total de anos-pessoa acaba sendo a média da população inicial e final vezes o tamanho do intervalo.
  - Mas isso nem sempre é correto.
- Sabe-se, por exemplo, que as crianças que morrem no primeiro ano de vida tendem a morrer muito mais no início (o primeiro mês) do que mais tarde.
  - Portanto, o comprimento médio das linhas vitais destas crianças seria bem menos do que a metade do período.

Taxas e probabilidades

- Em teoria, seria possível calcular todas as contribuições feitas por todas as pessoas que saíram da população durante o período, mas na prática raramente dispõe-se de dados suficientemente detalhados para fazer isso.
- Quando não se pode supor que cada pessoa que não completa o intervalo contribui a metade do mesmo, a solução geralmente adotada na prática é a aplicação de **fatores de separação**, baseados na experiência, para quantificar a contribuição das linhas vitais incompletas.

Taxas e probabilidades

- Neste caso o fator de separação para o intervalo de 5 a 9 anos provavelmente seria bem próximo a 0,5, talvez 0,48. Sendo assim, o número de anos-pessoa seria

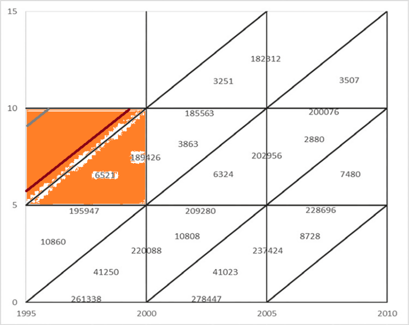
$$\text{Anos-pessoa} = 5 \times 200.076 + 0,48 \times 5 \times 9.204 = 1.022.469,6 \text{ anos.}$$

Taxas e probabilidades

- O critério de anos-pessoa pode ser o mais correto do ponto de vista conceitual, mas na prática o seu **cálculo exige certas aproximações**.
- Isso já é o caso numa análise de coorte, mas ainda mais numa análise de período.

Taxas e probabilidades

- A **área** relevante neste caso **é um quadrado**, por exemplo a **faixa etária de 5-9 anos em 1995-1999**, de modo que o resultado depende de onde exatamente as linhas vitais cruzam este quadrado, **mais perto da diagonal (onde o seu comprimento seria maior)** ou **mais perto dos cantos (onde seria menor)**.



Taxas e probabilidades

Na prática este tipo de detalhes normalmente não é levado em conta e o número de anos-pessoa é estimado de uma das três seguintes formas:

1. Como a média da população no início e no fim do período, ou seja, $5 \times (189.426 + 202.956)/2 = 980.955$ anos.
2. Eventualmente, se existem razões para supor uma distribuição desequilibrada, pode-se usar uma **ponderação** usando **fatores de separação**, como $5 \times (0,48 \times 189.426 + 0,52 \times 202.956)$.
3. Se existe uma estimativa para a população na metade do período, esta pode ser usada também, multiplicada pelo tamanho do intervalo.

$\rightsquigarrow$Os resultados de cada um destes procedimentos serão ligeiramente diferentes, mas normalmente as diferenças não devem ser muito significativas. **Na prática a primeira e a terceira solução são as preferidas.**

[1]Também chamado de **número pessoas-ano**.

Taxas centrais e iniciais

- Como foi mencionado, **taxas são razões**, mas nem toda razão é uma taxa.
- A **particularidade de uma taxa** é que o seu numerador mede o número de eventos que ocorrem num determinado período enquanto o denominador se refere à população que pode ser o objeto deste evento.

Taxas centrais e iniciais

Quando a população **se limita a uma determinada faixa etária**, distinguem-se dois tipos de taxas:

1. A **taxa central** (ou do tipo "m") combina eventos segundo o **critério de período no numerador** com uma população exposta segundo o número de anos-pessoa ou, na prática, segundo o critério 1 ou 3 do slide anterior.
2. A **taxa inicial** (ou do tipo "q") combina eventos segundo o **critério de coorte no numerador** com a população presente no início da faixa etária.

Taxas centrais e iniciais

↬ Ambas as taxas têm *vantagens* e *desvantagens*.

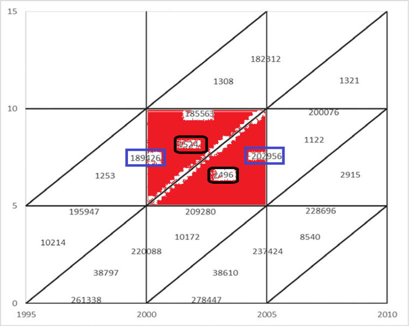
- A **vantagem da taxa central** é que ela costuma ser **fácil de calcular**, principalmente se forem usadas as aproximações 1 ou 3 e que ela se refere a um período único, enquanto a segunda taxa necessariamente combina informação de dois períodos.
- A **desvantagem da taxa central** é que ela **mistura a experiência de coortes distintas**.
  - Por esta razão, ela também é chamada **transversal**.
- Ela não pode ser interpretada como uma probabilidade e nem como proporção.

Taxas centrais e iniciais

- A **taxa inicial** é mais **difícil de calcular**, mas ela retrata a experiência de uma coorte real, ou seja, é uma medida **longitudinal**.
- **No caso de eventos não renováveis**, ela é uma proporção porque as pessoas afetadas pelos eventos no denominador fazem parte da população inicial no denominador.
    - Esta proporção também pode ser interpretada como uma **probabilidade**.

Taxas centrais e iniciais

- Considere o **grupo etário de 5-9 anos** e o **período de 2000-2004** no gráfico a seguir.



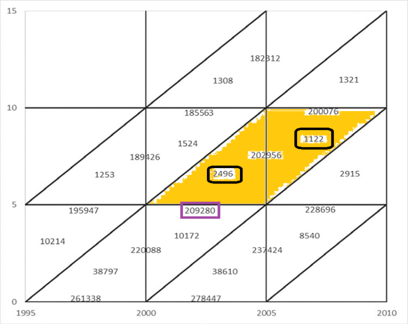
Taxas centrais e iniciais

- O **numerador da taxa central de mortalidade** para este grupo é $1.524 + 2.496 = 4.020$ óbitos.
- O **denominador** é mais facilmente calculado como $5 \times (189.426 + 202.956)/2 = 980.955$. Portanto, a taxa **(multiplicada por 1.000)** é

$$_5m_5 = 1.000 \times 4.020/980.955 = 4,10 \text{ por } 1.000.$$

Taxas centrais e iniciais

- Por outro lado, (**grupo etário de 5-9 anos**)a taxa inicial para os **períodos 2000-2004 e 2004-2009** tem um **numerador** de $2.496 + 1.122 = 3.618$ e um **denominador** de 209.280



Taxas centrais e iniciais

De modo que

$$ {}_5q_5 = 1.000 \times 3.618/209.280 = 17,3 \text{ por } 1.000. $$

Taxas centrais e iniciais

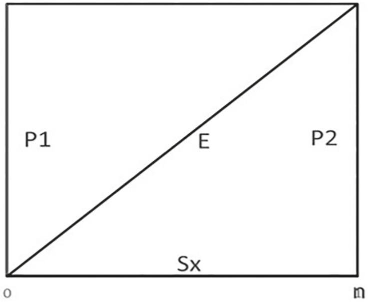
- Como a **mortalidade é um processo não renovável**, este último número pode ser **interpretado como a probabilidade de sobrevivência da idade exata de 5 anos até a idade exata de 10 anos no período entre 2000 e 2009**.
- No caso de eventos renováveis, como nascimentos, o cálculo destas taxas é o mesmo, mas raramente se calculam taxas do tipo "q" para estes eventos e quando são calculadas elas não podem ser interpretadas como probabilidades.
  - A sua interpretação correta é o **número esperado de vezes que o evento ocorre a cada pessoa dentro da faixa etária**.
- É de notar também que o denominador da taxa central **depende da amplitude do período** e o denominador da taxa inicial não.

Conversão de taxas centrais em taxas iniciais

- Uma das diferenças fundamentais entre taxas centrais e iniciais é que as taxas centrais se aplicam a um único período enquanto as taxas iniciais necessariamente se estendem por dois períodos.
- Entretanto, é comum encontrar situações (mortalidade, tábuas de vida) onde só se dispõe de taxas centrais referentes a um período, mas onde **é preciso estimar probabilidades**.
- Isso exige algum tipo de **mecanismo aproximado** para converter taxas do tipo m em taxas do tipo q.

Conversão de taxas centrais em taxas iniciais

- O seguinte procedimento é uma maneira para desmembrar um quadrado do diagrama de Lexis num losango para estimar uma taxa inicial.



Conversão de taxas centrais em taxas iniciais

- Supõe-se que não há informação sobre a divisão dos eventos E entre os dois triângulos que compõem o quadrado com tamanho de n por n anos do gráfico anterior, ou seja, que não se dispõe de uma dupla classificação.
- O primeiro passo portanto consiste em **dividir os eventos** E.
- O recurso geralmente usado para este propósito são os **fatores de separação**.
  - Supondo que o fator a simboliza a proporção dos eventos E que pertencem ao triângulo superior da esquerda, logicamente o número de eventos nesse triângulo seria $a \times E$ e no outro triângulo $(1 - a) \times E$.

Conversão de taxas centrais em taxas iniciais

- O próximo passo consiste em **deslocar o triângulo da esquerda para a direita**, para **formar um losango**.
- Mas isso não pode ser feito diretamente porque a população P_1 pode ser diferente de P_2.
- Portanto, todas as quantidades do triângulo superior da esquerda são multiplicadas por P_2/P_1.
  - Assim, P_1 se transforma em P_2 e $a \times E$ em $a \times E \times P2/P1$.
- Portanto o número ajustado de eventos é $E \times (a \times P2/P1 + 1 - a)$.
- Este número de eventos deve ser dividido pelo número inicial de pessoas entrando na faixa etária, que é S_x. Portanto,

$$ {}_nq_x = \frac{E(aP_2/P_1 + 1 - a)}{S_X}. $$

Conversão de taxas centrais em taxas iniciais

- Substituindo a definição da taxa central ${}_nm_x$ ($E = {}_nm_x \times n \times (P_1 + P_2)/2$), a seguinte fórmula aparece:

$$ {}_nq_x = n\,{}_nm_x \frac{(P_1 + P_2)(aP_2 + (1-a)P_1)}{2S_X P_1}. $$

Conversão de taxas centrais em taxas iniciais

- Esta fórmula ainda depende de várias incógnitas.
- Além da incerteza sobre o fator de separação a também há o fator S_x, cujo valor geralmente não se conhece exatamente.
    - No contexto da **tábua de vida** existem métodos padronizados para preencher essas lacunas.
- Aqui basta assinalar que em situações onde o evento E é responsável por (quase) toda a diferença entre S_x e P_2, é razoável supor que $S_x = P2 + 0,5 \times E$.

Conversão de taxas centrais em taxas iniciais

- Se além disso não houver muita diferença entre P_1 e P_2, a última fórmula pode ser aproximada da seguinte maneira:

$$ {}_nq_x \approx \frac{n\,{}_nm_x}{1 + 0,5 \times n\,{}_nm_x}. $$

- Esta fórmula constitui uma aproximação razoável em casos onde a modificação dos efetivos de população se deve basicamente a um único processo não renovável, como a mortalidade.

Próxima aula

- Medidas Básicas de Mortalidade.

Para casa

- Pequeno Trabalho 02: será postado no Moodle.
- Ler o capítulo 7 do livro "Métodos Demográficos Uma Visão Desde os Países de Língua Portuguesa"[2].

[2] FOZ, Grupo de. *Métodos Demográficos Uma Visão Desde os Países de Língua Portuguesa*. São Paulo: Blucher, 2021. https://www.blucher.com.br/metodos-demograficos-uma-visao-desde-os-paises-de-lingua-portuguesa_9786555500837

Por hoje é só!

Bons estudos!