

Qualidade e ajuste de dados demográficos

Prof. José Marcos Pinto da Cunha

Qualidade e ajuste de dados demográficos

- Sub-enumeração e cobertura;
- Formas de detecção de problemas nos dados demográficos;
- A “Conciliação Censitária”;
- Índices de qualidade da declaração de idade;
- Interpolação e suavização;
- “Abrindo” grupos etários quinquenais e decenais

Qualidade e ajuste de dados demográficos

- **Erros de enumeração** (*coverage* em inglês: notar que o sentido é distinto do que se usa neste curso): ligado a omissão ou inclusão equivocada de algum evento ou indivíduo na contagem geral do levantamento;
- **Omissão;**
- **Sobreenumeração;**
- **“Subenumeração líquida”** (diferença entre omissão e sobre-enumeração);
- **Erros de falta de declaração e má declaração de idade;**

Qualidade e ajuste de dados demográficos

- Dependendo do tipo de erro existem distintas formas de identificação (detecção) e/ou correção dos dados.
 1. **Erros de enumeração**- método de identificação/correção;
 2. **Má declaração de idade:**

Qualidade e ajuste de dados demográficos

1. Erros de enumeração- método de identificação/correção:

- Pesquisa de avaliação pós-censitária (não demográfica)
- Métodos demográficos
 - Análise intercensitária por coortes
 - Uso de estatísticas vitais (estimar coortes)
 - Uso de razões de sexo/idade esperados (população, nascimentos, mortalidade). Pode haver variações segundo contextos ou países.
 - **Combinação de 2, 3 e 4 = conciliação censitária (além de combinar esses elementos também usa observações e pressupostos sobre a qualidade da declaração de certos grupos etários)**

Qualidade e ajuste de dados demográficos

2. Má declaração de idade:

- Medidas de preferência digital (Myers, Whipple, etc.)
- Agrupamento não convencionais/desagregação de grupos para eliminar preferência (Métodos de Sprague, Karup King etc. Essa metodologia também ajuda desagregar quando os dados estão agregados)

Qualidade e ajuste de dados demográficos

Contudo, há que se ter cuidado para não interpretar incorretamente variações nos comportamento dos indicadores, índices de qualidade ou nas análise de coorte ou estatísticas vitais, uma vez que se deve ter certeza que isso não seja real, por exemplo, decorrente de migração, descontinuidades demográficas etc...

Qualidade e ajuste de dados demográficos

Erros de enumeração:

métodos para identificação de
erro na declaração de idade

Qualidade e ajuste de dados demográficos

Razão de idade:

$$\frac{{}_5P_a}{[{}_5P_{a-5} + {}_5P_a + {}_5P_{a+5}] * (1/3)} * 100$$

Valores desviados de 100 indicam problemas (sem flutuações de nascimentos, mortes e migração) os três grupos deveriam formar uma série linear. Esse é um exemplo de que às vezes pode não existir erro

Qualidade e ajuste de dados demográficos

Índice de Whipple

(preferência para dígitos 0 e 5)

- Índice único

$$\frac{\sum (P_{25}, P_{30} \dots P_{60})}{1/5 * \sum_{i=23}^{62} P_i} * 100$$

Usa-se 23 a 62 anos para eliminar idade onde os problemas de declaração sejam mais complexo que a simples preferência (ex. esquecimento dos mais velhos):

Suposto: comportamento retangular nos dígitos.

Variação: o limite inferior do índice é 100 (sem preferência) e 500 (somente 0 e 5)

Qualidade e ajuste de dados demográficos

Índice de Myers

(preferência para todos os dígitos):

- índice para cada dígito;
- Índice geral:

Forma de Cálculo:

- Somar todas as populações por dígito de 10 a 89 anos;
- Somar todas as populações por dígito de 20 a 99 anos;
- Ponderar (a) por 1, 2, 3... e (b) pelo complemento de 10, ou seja, 9, 8, 7... (portanto, o dígito 9 será ponderado primeiro por 10 e depois por zero);
- Somar os dois resultados;
- Calcular a % de cada dígito sobre a volume total;
- Diferença de 10% será o índice. O índice resumo será a soma de todos os índices por dígito dividido por 2;

Suposto: hipótese de retangularidade - é difícil imaginar uma razão para que alguns dos dígitos estão diferentemente representados na população;

Variação: mede o desvio de 10% para cada idade – Quanto mais próximo de 10% melhor a declaração. O índice geral é obtido da soma dos índices por idade (em módulo) dividido por 2. A variação desse geral é de 0 (sem problemas, até 90 todos declaram o mesmo dígito)

Qualidade e ajuste de dados demográficos

Índice de Myers (exemplo)

AGE COMPOSITION

117

Table 8-2.—Calculation of Preference Indexes for Terminal Digits by Myers' Blended Method, for the Philippines: 1960

[Age range covered here is 10 to 89 years. Commonly the same number of ages is included in the two sets of populations being weighted (cols. 1 and 2), and the second set of populations (col. 2) could be extended to age 99 if the figures for single ages were available. Above age 99 the population may be disregarded]

Terminal digit, a	Population with terminal digit a		Weights for--		Blended population		Deviation of percent from 10.00 ¹ (6)-10.00=
	Starting at age 10+a (1)	Starting at age 20+a (2)	Column 1 (3)	Column 2 (4)	Number (1)x(3)+(2)x(4)= (5)	Percent distribution (6)	
0.....	3,176,821	2,335,465	1	9	24,196,006	16.06	6.06
1.....	1,553,378	971,978	2	8	10,882,580	7.22	2.78
2.....	2,064,888	1,268,102	3	7	15,071,378	10.00	-
3.....	1,647,360	1,028,067	4	6	12,757,842	8.47	1.53
4.....	1,556,321	959,729	5	5	12,580,250	8.35	1.65
5.....	2,143,666	1,577,952	6	4	19,173,804	12.72	2.72
6.....	1,462,491	895,549	7	3	12,924,084	8.58	1.42
7.....	1,443,063	904,172	8	2	13,352,848	8.86	1.14
8.....	1,762,082	1,110,764	9	1	16,969,502	11.26	1.26
9.....	1,278,691	787,250	10	0	12,786,910	8.49	1.51
Total.....	(X)	(X)	(X)	(X)	150,695,204	100.00	20.07
Summary index of age preference= Total÷2=.....	(X)	(X)	(X)	(X)	(X)	(X)	10.04

Qualidade e ajuste de dados demográficos

- **Interpolação/ Extrapolação:** formas de inferir valores dentro (interpol.) ou fora (extrapol.) de um conjunto de dados conhecidos.
- **Formas:** modelos matemáticos, formas gráficas.
 - **Modelos matemáticos:** funções linear, exponencial, Gompertz e Logística (casos especiais de exponencial) e osculatórias (vem de ósculo que tem a ver com ponto de encontro entre dois ajustes polinomiais);
 - Formar gráficas: ajustes “manuais” (ex. q_x nas tábuas de mortalidade)

Qualidade e ajuste de dados demográficos

Interpolação/ Extrapolação: alguns usos

- Interpolação linear;
 - Estimativas de população
- Decomposição de grupo etários:
 - Multiplicadores de Sprague e Karup-King

Table C-4. — Interpolation Coefficients Based on the Karup-King Formula

The Karup-King formula is a four-term third-difference osculatory formula. It maintains the given values. Given points or groups must be equally spaced]

A. FOR INTERPOLATION BETWEEN GIVEN POINTS AT INTERVALS OF 0.2					B. FOR SUBDIVISION OF GROUPS INTO FIFTHS—Continued			
Interpolated point	Coefficients to be applied to--				Interpolated subgroup	Coefficients to be applied to--		
	N _{1.0}	N _{2.0}	N _{3.0}	N _{4.0}		G ₁	G ₂	G ₃
First interval					Last panel			
N _{1.0}	+1.000	.000	.000	.000	First fifth of G ₃	-.016	+.112	+.104
N _{1.2}	+.656	+.552	-.272	+.064	Second fifth of G ₃	-.032	+.104	+.128
N _{1.4}	+.408	+.856	-.336	+.072	Third fifth of G ₃	-.024	+.048	+.176
N _{1.6}	+.232	+.984	-.264	+.048	Fourth fifth of G ₃	+.008	-.056	+.248
N _{1.8}	+.104	+1.008	-.128	+.016	Last fifth of G ₃	+.064	-.208	+.344
Middle interval					C. FOR SUBDIVISION OF GROUPS INTO TENTHS OR HALVES			
N _{2.0}000	+1.000	.000	.000	Interpolated subgroup	Coefficients to be applied to--		
N _{2.2}	-.064	+.912	+.168	-.016		G ₁	G ₂	G ₃
N _{2.4}	-.072	+.696	+.424	-.048	First tenth of G ₂	+.0405	+.0640	-.0045
N _{2.6}	-.048	+.424	+.696	-.072	Second tenth of G ₂	+.0235	+.0880	-.0115
N _{2.8}	-.016	+.168	+.912	-.064	Third tenth of G ₂	+.0095	+.1060	-.0155
Last interval					Fourth tenth of G ₂	-.0015	+.1180	-.0165
N _{3.0}000	.000	+1.000	.000	Fifth tenth of G ₂	-.0095	+.1240	-.0145
N _{3.2}	+.016	-.128	+1.008	+.104	Sum of coefficients for first			
N _{3.4}	+.048	-.264	+.984	+.232	five-tenths = coefficients for			
N _{3.6}	+.072	-.336	+.856	+.408	first half of G ₂	+.0625	+.5000	-.0625
N _{3.8}	+.064	-.272	+.552	+.656	Sixth tenth of G ₂	-.0145	+.1240	-.0095
N _{4.0}000	.000	.000	+1.000	Seventh tenth of G ₂	-.0165	+.1180	-.0015
B. FOR SUBDIVISION OF GROUPS INTO FIFTHS					Eighth tenth of G ₂	-.0155	+.1060	+.0095
Interpolated subgroup	Coefficients to be applied to--			Interpolated subgroup	Ninth tenth of G ₂	-.0115	+.0880	+.0235
	G ₁	G ₂	G ₃		Last tenth of G ₂	-.0045	+.0640	+.0405
First panel				Sum of coefficients for last				
First fifth of G ₁	+.344	-.208	+.064	five-tenths = coefficients for				
Second fifth of G ₁	+.248	-.056	+.008	last half of G ₂	-.0625	+.5000	+.0625	
Third fifth of G ₁	+.176	+.048	-.024					
Fourth fifth of G ₁	+.128	+.104	-.032					
Last fifth of G ₁	+.104	+.112	-.016					
Middle panel								
First fifth of G ₂	+.064	+.152	-.016					
Second fifth of G ₂	+.008	+.224	-.032					
Third fifth of G ₂	-.024	+.248	-.024					
Fourth fifth of G ₂	-.032	+.224	+.008					
Last fifth of G ₂	-.016	+.152	+.064					

Karup
King

Table C-5.—Interpolation Coefficients Based on the Sprague Formula

[The Sprague formula is a six-term fifth-difference osculatory formula. It maintains the given values. Given points or groups must be equally spaced]

A. FOR INTERPOLATION BETWEEN GIVEN POINTS AT INTERVALS OF 0.2							B. FOR SUBDIVISION OF GROUPS INTO FIFTHS--Continued						
Interpolated point	Coefficients to be applied to--						Interpolated subgroup	Coefficients to be applied to--					
	N _{1.0}	N _{2.0}	N _{3.0}	N _{4.0}	N _{5.0}	N _{6.0}		G ₁	G ₂	G ₃	G ₄	G ₅	
First interval							First fifth of G ₃ Second fifth of G ₃ Third fifth of G ₃ Fourth fifth of G ₃ Last fifth of G ₃	Middle panel					
N _{1.0}	+1.0000	.0000	.0000	.0000	.0000			-.0128	+.0848	+.1504	-.0240	+.0016	
N _{1.2}	+.6384	+.6384	-.4256	+.1824	-.0336			-.0016	+.0144	+.2224	-.0416	+.0064	
N _{1.4}	+.3744	+.9984	-.5616	+.2304	-.0416			+.0064	-.0336	+.2544	-.0336	+.0064	
N _{1.6}	+.1904	+.1424	-.4896	+.1904	-.0336			+.0064	-.0416	+.2224	+.0144	-.0016	
N _{1.8}	+.0704	+.1264	-.2816	+.1024	-.0176			+.0016	-.0240	+.1504	+.0848	-.0128	
Next-to-first interval							First fifth of G ₄ Second fifth of G ₄ Third fifth of G ₄ Fourth fifth of G ₄ Last fifth of G ₄	Next-to-last panel					
N _{2.0}0000	+1.0000	.0000	.0000	.0000				-.0144	+.0912	+.1408	-.0176	
N _{2.2}	-.0336	+.8064	+.3024	-.0896	+.0144				-.0080	+.0400	+.1840	-.0160	
N _{2.4}	-.0416	+.5824	+.5824	-.1456	+.0224				.0000	-.0080	+.2160	-.0080	
N _{2.6}	-.0336	+.3584	+.8064	-.1536	+.0224				+.0080	-.0480	+.2320	+.0080	
N _{2.8}	-.0176	+.1584	+.9504	-.1056	+.0144				+.0144	-.0752	+.2272	+.0336	
Middle interval							First fifth of G ₅ Second fifth of G ₅ Third fifth of G ₅ Fourth fifth of G ₅ Last fifth of G ₅	Last panel					
N _{3.0}0000	.0000	+1.0000	.0000	.0000	.0000			+.0176	-.0848	+.1968	+.0704	
N _{3.2}	+.0128	-.0976	+.9344	+.1744	-.0256	+.0016			+.0160	-.0720	+.1360	+.1200	
N _{3.4}	+.0144	-.1136	+.7264	+.4384	-.0736	+.0080			+.0080	-.0320	+.0400	+.1840	
N _{3.6}	+.0080	-.0736	+.4384	+.7264	-.1136	+.0144			-.0080	+.0400	-.0960	+.2640	
N _{3.8}	+.0016	-.0256	+.1744	+.9344	-.0976	+.0128			-.0336	+.1488	-.2768	+.3616	
Next-to-last interval							C. FOR SUBDIVISION OF GROUPS INTO TENTHS OR HALVES						
N _{4.0} N _{4.2} N _{4.4} N _{4.6} N _{4.8}0000	.0000	+1.0000	.0000	.0000	Interpolated subgroup	Coefficients to be applied to--					
		+.0144	-.1056	+.9504	+.1584	-.0176		G ₁	G ₂	G ₃	G ₄	G ₅	
		+.0224	-.1536	+.8064	+.3584	-.0336	First tenth of G ₃ Second tenth of G ₃ Third tenth of G ₃ Fourth tenth of G ₃ Fifth tenth of G ₃ Sum of coefficients for first five-tenths = coefficients for first half of G ₃ Sixth tenth of G ₃ Seventh tenth of G ₃ Eighth tenth of G ₃ Ninth tenth of G ₃ Last tenth of G ₃ Sum of coefficients for last five-tenths = coefficients for second half of G ₃						
		+.0224	-.1456	+.5824	+.5824	-.0416		-.0076	+.0510	+.0660	-.0096	+.0002	
		+.0144	-.0896	+.3024	+.8064	-.0336		-.0052	+.0338	+.0844	-.0144	+.0014	
	Last interval							-.0022	+.0154	+.1036	-.0195	+.0027	
N _{5.0}0000	.0000	.0000	+1.0000	.0000		+.0006	-.0010	+.1188	-.0221	+.0037	
N _{5.2}		-.0176	+.1024	-.2816	+.1264	+.0704		+.0027	-.0133	+.1272	-.0203	+.0037	
N _{5.4}		-.0336	+.1904	-.4896	+.1424	+.1904							
N _{5.6}		-.0416	+.2304	-.5616	+.9984	+.3744							
N _{5.8}		-.0336	+.1824	-.4256	+.6384	+.6384							
N _{6.0}0000	.0000	.0000	.0000	+1.0000							
B. FOR SUBDIVISION OF GROUPS INTO FIFTHS								Interpolated subgroup	Coefficients to be applied to--				
First fifth of G ₁ Second fifth of G ₁ Third fifth of G ₁ Fourth fifth of G ₁ Last fifth of G ₁									G ₁	G ₂	G ₃	G ₄	G ₅
First panel								First fifth of G ₂ Second fifth of G ₂ Third fifth of G ₂ Fourth fifth of G ₂ Last fifth of G ₂					
Next-to-first panel													

Qualidade e ajuste de dados demográficos

Suavização:

- Forma de tornar as distribuição dos pontos “melhor comportada”. Pode ser feita manualmente (forma gráfica) ou por funções matemáticas;
- Os procedimentos de interpolação também podem ser utilizados como forma de suavizar essas distribuições. Considerar, por exemplo, a distribuição etária onde os erros de declaração, omissão, preferência digital podem ser “corrigidos”;
- Um método muito utilizado para suavizar os dados é a “**Média Móvel**” (explicar);
- Outros métodos: “splines”