

REVISTA BRASILEIRA DE ESTATÍSTICA

Órgão oficial do IBGE
e Sociedade Brasileira de Estatística

Redação:

Av. Franklin Roosevelt, 166 — ZC-39
Rio de Janeiro, RJ — Brasil — Tel: 252-3605

Diretor responsável:

Amaro da Costa Monteiro

Secretário:

Mário Fernandes Paulo

A Revista não se responsabiliza
pelos conceitos emitidos
em artigos assinados

Preço:

assinatura anual: Cr\$ 36,00
número avulso: Cr\$ 10,00

Vendas:

Av. Franklin Roosevelt, 146-A — loja B
Tel: 242-7142

SUMÁRIO

Artigos

- Considerações sobre modelos causais
— Elza Berquó e Rubens Murillo Marques 5
- Notas para o cálculo dos índices característicos de uma distribuição de frequência
— Prof. Lauro Sodré Viveiros de Castro 29

Transcrições

- Métodos de análise e avaliação
— William Brass e Ansley J. Coale 45

Noticiário

- Pesquisa Especial de Bovinos 117
- Calendário das reuniões internacionais de estatística 125
- Conselho Federal de Estatística amplia regulamentação profissional 127
- Sistema Estatístico Nacional 135
- Normas técnicas para numeração progressiva das seções de documentos 136

Bibliografia

- Resenha bibliográfica 141
- Anuário Estatístico do Brasil — 1974 ... 152
- Publicações editadas pelos órgãos de estatística do IBGE no trimestre janeiro-março de 1975 153
- Censo Agropecuário: Alagoas, Sergipe, Santa Catarina, Rio Grande do Sul e Distrito Federal 155
- RBE: Índice alfabético e remissivo — Ano 1974 157

R. bras. Estat.	Rio de Janeiro	v. 36	n.º 141	p. 1 a 164	jan./mar. 1975
-----------------	----------------	-------	---------	------------	----------------

Revista Brasileira de Estatística. v. 1- n. 1- jan./mar.
1940- Rio de Janeiro, Fundação Instituto Brasileiro de Geografia
e Estatística. Diretoria Técnica. Departamento de Divulgação Estatística,
1940-

v. 27cm Trimestral

Antiga estrutura deste órgão: Brasil. Conselho Nacional de Estatística. Diretoria de Documentação e Divulgação e Instituto Brasileiro de Estatística. Departamento de Divulgação Estatística.

1. Estatística — Periódicos. I. Fundação Instituto Brasileiro de Geografia e Estatística. Diretoria Técnica. Departamento de Divulgação Estatística, ed.

CDD 310.5



CDU 31:05(81)

A evolução das pesquisas estatísticas, como um imperativo do desenvolvimento nacional e da nova dinâmica de progresso do mundo moderno, leva a REVISTA BRASILEIRA DE ESTATÍSTICA, a partir deste número, a dar nova distribuição a sua matéria, objetivando refletir, com maior destaque e fidelidade, as diversas pesquisas dos órgãos que integram o IBGE.

Assim, dentro dessa nova sistemática, abriremos maior espaço à divulgação dos estudos elaborados pelas Superintendências Técnicas, que abrangem indicadores econômicos e sociais, áreas metropolitanas, pesquisas populacionais e urbanas, estudos demográficos, educacionais, agropecuários, pesquisas domiciliares, comerciais, industriais e de serviços, planejamento dos diversos censos, pesquisas sobre recursos naturais, etc.

Não obstante a diversificação da matéria enumerada, e sem prejuízo de suas outras seções, a RBE continuará recebendo, como até aqui, a colaboração de quantos queiram contribuir para o aprimoramento do seu nível técnico.

CONSIDERAÇÕES SOBRE MODELOS CAUSAIS*

Elza Berquó

do Centro Brasileiro de Análise e Planejamento
(CEBRAP) de São Paulo

Prof. Rubens Murillo Marques

Titular de Estatística da Universidade
Estadual de Campinas, São Paulo

SUMÁRIO

Introdução

1. *Objetivos*
2. *Modelo Linear Geral — Regressão Linear Múltipla*
3. *Modelo Causal Completo*
4. *Análise de Trajetórias com um Modelo Causal Completo*
5. *Modelos Causais com mais de uma Equação*

INTRODUÇÃO

Um estudo retrospectivo da literatura nos leva diretamente aos trabalhos do geneticista Sewall Wright surgidos nos idos de 1918¹, 1920² e 1921³, nos quais o autor declara ter encontrado um método flexível de relacionar, num sistema multidimensional, os coeficientes de correlação entre variáveis às relações funcionais existentes entre elas. A este método, Wright denominou "The method of the path coefficients"

* Estudo apresentado na III Reunião do Grupo de Trabalho sobre Reprodução da População do Conselho Latinoamericano de Ciências Sociais — CLACSO

¹ WRIGHT, S. — On the nature of size factors. *Genetics* (3): 367-74, 1918

² ———. The relative importance of heredity and environment in determining the piebald pattern of guinea pigs. *Proceedings National Academy Science* (6): 320-32, 1920

³ ———. Causation and conelation *Journal Ag Res.* (20): 557-85, 1921

e apresenta em seus trabalhos não só uma exposição metodológica mas também alguns exemplos ilustrativos no campo da genética de população. Essas idéias, envolvendo noções de causação, no entanto, não foram muito bem recebidas num período em que a noção de correlação entre variáveis, devida a Karl Pearson, representava a mais recente contribuição à estatística descritiva. De fato, o surgimento em 1922⁴ do trabalho de Niles ilustra esta reação contrária, nele se lê: "we therefore conclude that philosophically the basis of the method of path coefficients is faulty, while practically the results of applying it where it can be checked prove it to be wholly unreliable".

As réplicas de Wright às críticas de Niles não tardaram, aparecendo em 1923⁵ mais um trabalho seu, no qual tenta mostrar que a condenação de Niles está baseada em uma concepção muito pessoal sobre causação e leis da natureza como conceitos científicos.

Diz ainda Wright que a segunda crítica no que se refere à aplicabilidade do método, está baseada no erro de Niles de não ter sido capaz de reconhecer que um "path coefficient" não é uma função simétrica de duas variáveis, pois encerra necessariamente uma noção de sentido.

Wright não ficou sem resposta e ainda em 1923⁶ Niles escreve um pequeno trabalho resumindo suas críticas às idéias de Wright e que a seu ver ele não conseguira remover. Os vícios da teoria continuam residindo no fato de que não existe maneira de saber *a priori* qual o sistema de ação de uma ou mais variáveis sobre outras variáveis, no sentido de se ter uma visão perfeita do universo tornando, portanto, impossível a pressuposição de um sistema explicativo completo, o que implica necessariamente na dificuldade em se atribuir a uma variável (efeito) a causa da outra.

Conquanto, Wright seguisse aplicando seu método em problemas de genética de população ao último trabalho de Niles seguiu-se um longo silêncio, de aproximadamente dez anos, quebrado em 1934⁷ por uma longa monografia em "The Annals of Mathematical Statistics", já em seu quinto número de existência. Nela Wright restabelece seu método dos "path coefficients" e faz uma revisão dos tipos de aplicação, especialmente naqueles casos onde, a seu ver, teria havido mal entendidos no passado, tanto no que se refere aos propósitos quanto aos procedimentos.

Após este trabalho poder-se-ia dizer que novamente na literatura estatística nenhuma discussão maior surge por um período de aproximadamente vinte anos durante os quais, além disso, é marcante o florescimento da estatística matemática baseada na axiomatização da teoria das probabilidades.

⁴ NILES, H. E. — Correlation, Causation and Wright's Theory of "Path Coefficients". *Genetics* (7): 258, 1922

⁵ WRIGHT, S. — The theory of path coefficients. A reply to Niles criticism. *Genetics* (8): 239-55, 1923

⁶ NILES, H. E. — The method of path coefficients An Answer to Wright *Genetics* (8): 261, 1923

⁷ WRIGHT, S. — The method of path coefficients *Annals of Mathematical Statistics* (5): 161-215, 1934

Em 1954 aparece assim a importante contribuição de Tukey⁸. É interessante observar que o autor, na introdução de seu trabalho, confessa que, ao ser convidado para preparar o Capítulo 3 do livro "Statistics and Mathematics in Biology", editado por Kempthorne, desconhecia completamente o que fosse "path analysis". Diz, ainda, que ao chegar a um ponto em que lhe parecia compreender o assunto razoavelmente bem, passou a se indagar das razões desta ignorância, concluindo que se devera muito mais ao fato de não ter manipulado problemas nesta área do que à falta de utilidade do método. Após um pouco mais de meditação sobre o assunto, acabou por achar que a tendência a se confundir correlação e causação como conceitos biométricos tinha também muito que ver com este estado de coisas.

Para a época, o trabalho de Tukey pode ser considerado de fôlego teórico. De fato, classificando os métodos de análise de dados existentes até aquele momento, em descritivo, tangencial e funcional, estabelece uma ponte entre este último e o princípio de causalidade.

Para Tukey correlação desempenha um certo papel apenas na fase descritiva, enquanto a regressão é a técnica no enfoque tangencial que, a seu ver, tem caráter de inferência quando as populações estudadas estão sujeitas a "pequenas mudanças". Quando os coeficientes de regressão estiverem sujeitos a uma grande estabilidade, como no caso clássico da regressão da estatura dos filhos sobre a estatura dos pais, em que o coeficiente de regressão é praticamente igual a 0,5 independentemente da população que se considere, então, estamos diante de uma situação funcional.

Diz ainda que, quando a "path analysis" foi proposta, o ataque sobre causalidade como um profundo conceito filosófico foi longe demais. Agora que podemos reconhecer na causação um papel de guia e marco de referência para análise funcional, estamos certos de tratar-se de um conceito útil e vemos pouca conexão entre ele e os coeficientes de correlação. O coeficiente de correlação é uma medida puramente descritiva; já o quadrado deste coeficiente pode muitas vezes ser interpretado como um coeficiente de determinação e, como tal, é uma medida tangencial. O método da "path analysis" está num nível superior! Quando construído em termos de coeficiente de correlação (como o foi originalmente) a intenção foi tomá-los como coeficiente de determinação, obviamente um método tangencial, baseado em um esquema causal. Reformulada em termos de coeficiente de regressão, a "path analysis" muito freqüentemente produz resultados invariantes, podendo, portanto, ser encarado como um método funcional.

Para Tukey, Wright deu um grande passo da etapa descritiva para a tangencial e só se deteve neste degrau sem atingir o seguinte, isto é, o funcional, devido às limitações da época em que propôs sua teoria e também devido à natureza dos problemas que manipulou.

⁸ TUKEY, J W — Causation, regression and path analysis In: *Statistics and Mathematics in Biology* Iowa, Ed Kempthorne, 1954 p. 35-66

Ao trabalho de Tukey seguiu-se uma monografia de Li, 1956⁹, de caráter expositivo, na qual ele situa a “path analysis” como um caso especial de análise multidimensional para tratar com sistemas “fechados” de variáveis que são linearmente relacionadas.

Por um sistema fechado entende que cada variável no sistema ou é uma combinação linear de outras variáveis do próprio sistema ou é um dos fatores básicos, que poderão ou não ser correlacionados com outros fatores básicos do sistema. O sistema é formalmente completo, incluindo todos os fatores básicos (causas) e suas variáveis resultantes (efeitos). Chama ainda a atenção Li para o fato de que o método de “path analysis” não deve ser visto como um meio de inferir relações causais a partir de correlações observadas entre um conjunto de variáveis interligadas; muito pelo contrário, o emprego do método deve ser precedido pela formulação de um esquema causal baseado ou no conhecimento *a priori* das relações causais, ou sobre as hipóteses quer aceites quer a serem testadas pelo investigador. Com estas definições e observações, Li toma equações de regressão múltipla, compatíveis segundo ele, com esquemas propostos, para os problemas concretos em questão, sempre na área da genética, e resolve o sistema de equações para encontrar os “path coefficients”.

Acreditamos que motivados por este trabalho, Turner e Stevens, 1959¹⁰ foram levados a enquadrar os esquemas causais como casos particulares da regressão múltipla generalizada para o caso de um sistema de duas ou mais equações de regressão, mostrando que, como tal, a solução das equações, para a obtenção dos “path coefficients”, não é sempre trivial, havendo situações que levam à indeterminação. Wright contesta logo em seguida, 1960¹¹, insistindo não só em que para ele os dois enfoques apresentam certas diferenças mas também nas vantagens de se trabalhar com os “path coefficients” em lugar dos coeficientes de regressão pelo fato daqueles serem números abstratos.

Ao que pudemos constatar, Wright abre e fecha as cortinas de um período que durou praticamente vinte anos e cuja história está documentada nas monografias publicadas em livros e revistas, especializados no domínio da biologia.

Porque se esquecem dessas contribuições, nos trabalhos que começam a surgir logo em seguida no cenário das ciências sociais, sugerem o aparecimento de um movimento, até certo ponto independente, à procura de novas formas de esquemas capazes de darem conta da análise de um número cada vez maior de variáveis presentes nos “surveys”

O trabalho de Simon, em 1954¹², é um exemplo disso, parece não ter tomado conhecimento dos ensaios de Wright ou, se o fez, parece ter

⁹ LI, C. C. — The concept of path coefficient and its impact on population genetics *Biometrics* (2) June 1956

¹⁰ TURNER, M. E. & STEVENS, C. D. — The regression analysis of causal paths *Biometrics* (2) June 1959

¹¹ WRIGHT, S. — Path coefficients and path regressions: alternative or complementary concepts? *Biometrics* (2) June 1960

¹² SIMONS, H. A. — Spurious correlation: a causal interpretation *American Statistical Association Journal* (267): 467-70, Sept 1954

sofrido mais diretamente a influência das contribuições de Koopman¹³, no campo de econometria. Simon retoma o problema da correlação espúria amplamente examinado nos anos 30 por renomados estatísticos da época como G. U. Yule e revivido mais tarde por Kendall e Lazarsfeld¹⁴. Para o entendimento da relação entre duas variáveis x e y introduz uma terceira z e tenta uma interpretação de tipo causal, em conformidade com suas idéias de causalidade expressas em trabalhos anteriores¹⁵. O que se fazia tradicionalmente, diz Simon, é calcular a correlação entre x e y , rx_y e sempre que este coeficiente é significativamente diferente de zero, saber o que concluir sobre a relação causal entre x e y . Em seguida calcular a correlação parcial entre x e y , mantida z constante $rx_y.z$ e comparar com rx_y . Se $rx_y.z$ é próxima de zero, enquanto rx_y não o é então: (a) z é uma variável interveniente — o efeito causal de x sobre y (ou vice-versa) opera através de z ; ou (b) a correlação entre x e y resulta do efeito causal conjunto de z sobre x e y e, portanto, esta correlação é espúria. Deve-se notar entretanto que em (a) não se sabe se x é causa de y ou se y é causa de x (através de z em ambos os casos); e de qualquer maneira as correlações não nos dizem se estamos diante da situação (a) ou (b).

Por isso a técnica desenvolvida por Simon envolve o uso de equações simultâneas do tipo de regressão em vez de correlações parciais.

Diz ele, se supuzermos que um sistema de, digamos, três variáveis x , y e z é determinado por um mecanismo linear, precisaremos de três equações: uma contendo x como variável dependente e y e z como independentes; na segunda equação y aparecerá como dependente e x e z como independentes; na terceira, x e y como independentes e z aparecerá como variável dependente. As equações teriam o aspecto:

$$\begin{aligned} x + a_{12}y + a_{13}z &= u_1 \\ a_{21}x + y + a_{23}z &= u_2 \\ a_{31}x + a_{32}y + z &= u_3 \end{aligned}$$

onde os u são erros que medem os efeitos sobre o sistema de todas as outras variáveis não introduzidas explicitamente.

Em seguida, tenta distinguir uma correlação genuína entre x e y de uma espúria, pela consideração da nulidade de alguns dos coeficientes do sistema supra. De fato, se $a_{31} = a_{32} = a_{33} = 0$, isto quer dizer, para Simon, que uma mudança em u_3 alterará z diretamente e x e y indiretamente; uma mudança em u_2 alterará y diretamente e x indiretamente, mas deixará z inalterada. Finalmente, uma mudança em u_1 afetará somente x . Portanto, diz ele, y é causalmente dependente de

¹³ KOOPMAN, T. C. — Identification problems in Economic Model Construction *Econometrica* (17): 125-44, April 1949 Reprinted as Chapter II In: *Studies in Econometric Methods*, Corolless Commission Monograph 14

¹⁴ KENDALL, P. L. & LAZARSELD, Paul F. — Problems of survey analysis In: MERTON & LAZARSELD ed *Continuities in Social Research*, Free Press, 1950, p 133-96

¹⁵ SIMON, H. A. — On the definition of causal relations *Journal Philosophy* (49): 517-28, 1952

z e x é causalmente dependente de y e z . Se x e y apresentarem correlação, esta correlação é verdadeira se ainda $a_{12} \neq 0$.

Entretanto, prossegue Simon, se todas as demais condições acima referidas forem satisfeitas, embora $a_{12} = 0$, teríamos que interpretar a correlação entre x e y como espúria pois ela seria devida exclusivamente à influência de z sobre as variáveis x e y .

Na prática, todavia, os coeficientes não são conhecidos e precisam ser estimados o mesmo sucedendo com os u . Para tanto, depara-se com um sistema de $(3n + 6)$ incógnitas e apenas $3n$ equações (3 para cada ponto amostral), o que torna o sistema não identificável. A fim de contornar esta dificuldade, o autor propõe duas classes de pressuposições: a primeira, já mencionada, consiste em supor uma ordenação das variáveis, isto é, pensar que alguns coeficientes são nulos de início; a segunda, em geral, mais implícita do que explícita no tratamento do problema, reside na suposição de que as perturbâncias causais são não correlatas, isto é, no dizer de Boudon¹⁶, os “fatores implícitos” são específicos — não agem simultaneamente sobre duas variáveis.

A Simon seguiram-se as contribuições de Blalock^{17 e 18}, praticamente na mesma linha, surgindo assim na literatura o conhecido modelo Simon-Blalock.

Blalock¹⁹ aceita a idéia de que há dois critérios para determinar a importância relativa de um conjunto de variáveis. Um critério baseado no valor observado das correlações entre as variáveis e outro, um “critério causal” usado na argumentação teórica na tentativa de levar em conta uma ordem causal entre as variáveis. Para Blalock, o que se necessita é de uma técnica quantitativa que leve em consideração as relações causais entre as variáveis independentes e para ele, o método de Simon provê a maneira de testar a existência ou não de um esquema causal. De fato, diante de um conjunto de variáveis cuja estrutura causal é procurada, postula-se a ordem causal, a qual dá como consequência, valores teóricos esperados para os coeficientes de correlação. Calcula-se, em seguida, as correlações e compara-se os valores observados com os teóricos correspondentes. A aderência entre estas duas séries de valores nos leva a aceitar este ou aquele modelo causal.

Este tipo de enfoque provoca reações muito semelhantes às surgidas anos atrás, conforme já salientamos nessa Introdução, ou seja, sobre a noção de assimetria que está presente nos esquemas causais *versus* simetria nas correlações. O trabalho de Polk²⁰ critica o modelo de Simon-Blalock pelo fato de conter implicitamente a noção de correlação na qual z prediz y mas y não prediz z . Diz Polk que toda a dificul-

¹⁶ BOUDON, R — *L'analyse mathématique des faits sociaux*. Paris, Librairie Plon, 1967

¹⁷ BLALOCK, JR., H. M. — Correlational analysis and causal inference *American Anthropologist* (62): 624-31, 1960

¹⁸ ———. Correlation and Causality: The multivariate case *Social Forces* (39): 1246-51, 1961

¹⁹ ———. Evaluating the relative importance of variables. *American Sociological Review*, 26: 866-74, 1961

²⁰ POLK, K. — A note on asymmetric causal models *American Sociological Review*, 27 (4) 1962

dade em aplicar este método está na contradição subjacente, pois no campo da correlação a simetria é condição básica. O procedimento usado por Simon, empregando equações simultâneas é uma tentativa de incorporar na estrutura das equações matemáticas de predição esta assimetria que define as relações causais. No entretanto, Polk vê as seguintes dificuldades na aplicação do modelo de Simon: “os modelos de regressão linear serão simétricos mesmo quando a relação entre as variáveis é assimétrica no sentido causal”; “uma relação causal não pode ser estabelecida por uma equação matemática sugerida por Simon, desde que causação só pode ser estabelecida pelos conhecimentos teóricos que tenha o pesquisador da natureza das variáveis incluídas na análise”.

Felizmente, Blalock²¹ contesta Polk sobre o erro cometido quando diz que as equações de regressão são simétricas. Enfatiza que esta simetria só existe em equações matemáticas, aceita a “crítica” de que seu método é empírico e prossegue advogando que as equações de regressão podem ser usadas em dois sentidos fundamentalmente distintos: como equações de predição ou estimação e como modelos causais.

Em 1966, Ducan²² apresenta uma breve revisão da literatura sobre esquemas causais e, para satirizar Boudon quando diz que o modelo de Simon-Blalock é um “special case” ou “weak form” de “path analysis” ou de sua “dependence analysis” e que além disso os exemplos apresentados não são “convincing”, tenta mostrar alguns exemplos de trabalhos já publicados, vistos agora do ângulo da “path analysis”. Ducan parece muito sereno quando afirma que no caso em que o sistema não apresenta variáveis não mensuráveis, a “path analysis” se confunde com a teoria da regressão.

Baseado no trabalho de Ducan surgem, até onde temos notícia, duas tentativas de empregar a “path analysis” em estudos de fecundidade. Uma, trabalhando com dados censitários²³ e a outra com dados de *survey*²⁴.

1. OBJETIVOS

A revisão da literatura da qual nos ocupamos até agora teve por finalidade assinalar as controvérsias que ainda subsistem neste campo, agravadas pelo fato de que cada vez mais os especialistas de uma área têm menos contato com as contribuições afins do ponto de vista da metodologia, as quais vão surgindo em outros setores do conhecimento. No momento em que estes procedimentos começam a ser aplicados em

²¹ BLALOCK, JR., H. M. — Further observations on asymmetric causal models. *American Sociological Review*, 27 (4) 1962

²² DUCAN, O. D. — Path Analysis: Sociological. *American Journal Sociology*, 72 (1): 1-15

²³ BLANCH, J. M. — *Factores Estructurales y ecológicos en la fecundidad de centroamérica y Panamá*. Apresentado à Conferência Regional Latinoamericana de Población, México, ago. 1970

²⁴ PATARRA, N. L. — *O estudo sobre Reprodução Humana no Distrito de São Paulo. Um enfoque sociológico*. Tese de doutoramento apresentada à Faculdade de Saúde Pública da Universidade de São Paulo. São Paulo, 1972

modelos de reprodução da população, nos parece de importância tentar esclarecer as bases metodológicas sobre as quais eles repousam. Mais especificamente, como mostraremos nas seções seguintes, pontos fundamentais como, distinção entre:

- tratamento matemático e tratamento estatístico dado às variáveis;
 - correlação como conceito e como mera medida descritiva;
 - regressão, modelos causais completos e “path analysis”;
- devem estar presentes tanto na escolha do modelo de análise mais adequado a cada caso como na sua aplicação.

Com este objetivo é que se justifica, a nosso ver, o detalhe com que foram apresentados alguns aspectos da metodologia analisada.

2. MODELO LINEAR GERAL — REGRESSÃO LINEAR MÚLTIPLA

Seja X_0 uma variável aleatória e X_1, X_2, \dots, X_k ($k - 1$) variáveis tais que:

$$E(X_0) = \sum_{j=1}^k \beta_{0j} X_j \quad (2.1)$$

com $X_i = 1$

Os valores dos parâmetros β_{0j} são, em geral, desconhecidos nos problemas práticos e precisam ser estimados a fim de se conhecer explicitamente a relação linear (2.1). É claro que se eles fossem conhecidos a (2.1) permitiria conhecer exatamente os valores de $E(X_0)$ dados os valores de X_1, X_2, \dots, X_k .

Se pensarmos em termos de uma amostra de n observações sobre X_0 e os X_j ; podemos escrever:

$$X_{0i} = \sum_{j=1}^k \beta_{0j} X_{ij} + \mu_i \quad i = 1, 2, \dots, n \quad (2.2)$$

onde μ_i é uma perturbação, ou erro, ou resíduo devido à amostragem-erros de medida na variável dependente e ocasionados pelos instrumentos de medida utilizados. É claro que se a (2.1) não incluísse todas as variáveis X_j , então μ_i estaria também contendo as discrepâncias sobre X_0 advindas da não inclusão dessas variáveis. Esta é em geral a situação mais corrente nos problemas práticos. Ou estamos diante da verificação experimental de uma lei (por exemplo, para uma determinada massa gasosa, a uma temperatura constante, o produto de sua pressão por seu volume é constante — lei de Boyle-Mariotte), onde os μ_i representam apenas as discrepâncias provenientes dos erros de leitura dos instrumentos empregados. Suponhamos ainda que estamos à procura de um modelo capaz de explicar da melhor forma possível o compor-

tamento de uma variável em função de um conjunto de outras que, acreditamos, devem intervir no problema. Neste último caso os μ_i medem, além dos erros experimentais, os efeitos de nossa falha na escolha das variáveis explicativas.

Em notação matricial, a (2.2) pode ser escrita:

$$\underline{X}_o + \underline{X}\underline{\beta} + \underline{\mu} \quad (2.3)$$

onde

$$\underline{X}_o = \begin{bmatrix} X_{o1} \\ X_{o2} \\ \vdots \\ X_{on} \end{bmatrix} \quad \underline{X} = \begin{bmatrix} 1 & X_{21} & \dots & X_{k1} \\ 1 & X_{22} & \dots & X_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & X_{2n} & \dots & X_{kn} \end{bmatrix} \quad \underline{\beta} = \begin{bmatrix} \beta_{o2} \\ \beta_{o3} \\ \vdots \\ \beta_{ok} \end{bmatrix} \quad \underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$$

Para a estimação do vetor $\underline{\beta}$ algumas pressuposições são necessárias:

- (a) $E(\underline{\mu}) = 0$, isto é, os μ_i são variáveis aleatórias com média zero;
- (b) $E(\underline{\mu}\underline{\mu}') = \sigma^2 \underline{I}_n$, ou seja, os μ_i têm variância constante e igual a σ^2 para todo i e são não correlatos dois a dois;
- (c) \underline{X} é um conjunto de números fixados, isto é, em repetidas amostragens a única fonte de variação no vetor \underline{X}_o é a variação no vetor $\underline{\mu}$, isto é, os \underline{X} não estão supostos aqui como variáveis aleatórias;
- (d) \underline{X} tem posto $k < n$ o que significa que o número de parâmetros a serem estimados (k) não pode exceder o tamanho da amostra (n) e que nenhuma relação linear *exata* existe entre qualquer uma das variáveis \underline{X} . Se isto acontecesse, o posto de \underline{X} seria menor do que k e também o posto de $\underline{X}'\underline{X}$ e sendo esta simétrica de ordem k , significaria que a inversa $(\underline{X}'\underline{X})^{-1}$ não estaria definida, impedindo a determinação dos $\underline{\beta}$.

Seja $\hat{\underline{\beta}} = \{\hat{\beta}_{o1}, \hat{\beta}_{o2}, \dots, \hat{\beta}_{ok}\}$ um vetor coluna dos estimadores de $\underline{\beta}$. Com isto, podemos escrever:

$$\underline{X}_o = \underline{X} \hat{\underline{\beta}} + \underline{e} \quad (2.4)$$

onde \underline{e} denota o vetor coluna de n resíduos $(\underline{X}_o - \underline{X}\hat{\underline{\beta}})$. A (2.4) é denominada equação de regressão de \underline{X}_o em X_1, X_2, \dots, X_k e $\hat{\beta}_{oj}$ é o coeficiente de regressão de X_o em X_j ($j = 1, 2, \dots, k$) e mede o número de unidades de que se espera que X_o se altere para cada aumento de uma unidade em X_j . Portanto, $\hat{\beta}_{oj}$ depende não só do grau de correla-

ção de X_0 com relação a X_1 , mas também da unidade física na qual a variável X_1 é medida.

A diferença entre a (2.3) e a (2.4) reside no fato de que na (2.3) $\underline{\mu}$ se refere a resíduos desconhecidos, uma vez que o vetor $\underline{\beta}$ é desconhecido, enquanto que na (2.4) tem-se um conjunto de estimativas $\hat{\beta}$ e o conjunto correspondente de resíduos.

Pela (2.4), a soma de quadrados dos resíduos vale:

$$\sum_{i=1}^n e_i^2 = \underline{e}'\underline{e} = (\underline{X}_0 - \underline{X}\hat{\beta})'(\underline{X}_0 - \underline{X}\hat{\beta})$$

O valor do vetor $\hat{\beta}$ é obtido minimizando $\sum_{i=1}^n e_i^2$, isto é, tal que

$$\underline{e}'\underline{e} = \text{mínima} \quad (2.5)$$

ou seja: $\hat{\beta}$ é a solução da equação

$$\frac{\partial}{\partial \hat{\beta}} (\underline{e}'\underline{e}) = 0$$

$$\text{ou de: } \frac{\partial}{\partial \hat{\beta}} (\underline{X}_0' \underline{X}_0 - 2\hat{\beta}' \underline{X}' \underline{X}_0 + \hat{\beta}' \underline{X}' \underline{X} \hat{\beta}) = 0$$

$$\text{que vale: } \hat{\beta} = (\underline{X}' \underline{X})^{-1} \underline{X}' \underline{X}_0 \quad (2.6)$$

Pode-se demonstrar que:

$$\hat{\beta} = \underline{\beta} + (\underline{X}' \underline{X})^{-1} \underline{X}' \underline{\mu}$$

$$E(\hat{\beta}) = \underline{\beta} \quad (2.7)$$

$$V_{ar}(\hat{\beta}) = \sigma^2 (\underline{X}' \underline{X})^{-1} \quad (2.8)$$

isto é, que os estimadores de mínimos quadrados são lineares, não viciados e de menor variância dentre todos os demais estimadores.

Pode-se demonstrar também que o coeficiente de correlação múltipla $R_{0 \ 23 \ k}^2$ definido como a proporção da variância total de X_0 explicada pela relação linear ajustada vale:

$$R_{0 \ 23 \ k}^2 = \frac{\hat{\beta}' \underline{X}' \underline{X}_0 - \frac{1}{n} (\sum X_0)^2}{\underline{X}_0' \underline{X}_0 - \frac{1}{n} (\sum X_0)^2} \quad (2.9)$$

Convém notar que falar aqui em coeficiente de correlação constitui um *abuso de linguagem* uma vez que a noção de correlação pressupõe uma distribuição conjunta de variáveis aleatórias. Isto apenas se justifica no sentido de que ainda assim ele se reveste de utilidade prática para efeito de interpretação da adequabilidade do modelo. Por esse motivo é preferível denominar $R_{0 \ 23 \ k}^2$ de coeficiente de determinação, cujo valor máximo é obviamente a unidade em cujo caso haja uma

completa determinação de X_o pelo conjunto X_2, \dots, X_k . Neste caso estaremos diante de um modelo determinístico, matemático, sem distúrbâncias, isto é: $\underline{X}_o = \underline{X}\hat{\beta} = \underline{X}\beta$.

Todas as fórmulas até aqui deduzidas o foram a partir da (2.2) onde as variáveis estão medidas a partir da origem zero.

Tomando a média das observações da (2.2) e da (2.4) vem:

$$\bar{X}_o = \beta_{o1} + \beta_{o2} \bar{X}_2 + \dots + \beta_{ok} \bar{X}_k + \bar{\mu} \quad (2.10)$$

$$= \hat{\beta}_{o1} + \hat{\beta}_{o2} \bar{X}_2 + \dots + \hat{\beta}_{ok} \bar{X}_k \quad (2.11)$$

uma vez que $\bar{e} = 0$.

Subtraindo-se da (2.2) a (2.10) e da (2.4) a (2.11) e usando-se letras minúsculas para representar os desvios, vem:

$$x_{oi} = \beta_{o2} x_{2i} + \dots + \beta_{ok} x_{ki} + (\mu_i - \bar{\mu}) \quad (2.12)$$

$$= \hat{\beta}_{o2} x_{2i} + \dots + \hat{\beta}_{ok} x_{ki} + e_i \quad (2.13)$$

para $i = 1, 2, \dots, n$.

Em notação matricial:

$$\underline{X}_o = \underline{X}\beta + \underline{\mu} - \underline{\mu} = \underline{X}\hat{\beta} + \underline{e} \quad (2.14)$$

onde:

$$\underline{X}_o = \begin{bmatrix} x_{o1} \\ x_{o2} \\ \vdots \\ x_{on} \end{bmatrix} \quad \underline{X} = \begin{bmatrix} x_{21} & x_{31} & \dots & x_{k1} \\ x_{22} & x_{32} & \dots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{2n} & x_{3n} & \dots & x_{kn} \end{bmatrix} \quad \underline{\beta} = \begin{bmatrix} \beta_{o2} \\ \beta_{o3} \\ \vdots \\ \beta_{ok} \end{bmatrix}$$

$$\underline{\hat{\beta}} = \begin{bmatrix} \hat{\beta}_{o2} \\ \hat{\beta}_{o3} \\ \vdots \\ \hat{\beta}_{ok} \end{bmatrix} \quad \underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} \quad \underline{\bar{\mu}} = \begin{bmatrix} \bar{\mu} \\ \bar{\mu} \\ \vdots \\ \bar{\mu} \end{bmatrix} \quad \underline{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

Continuam sendo válidos

$$\hat{\beta} = (\underline{X}'\underline{X})^{-1} \underline{X}'\underline{X}_o$$

$$E(\hat{\beta}) = \beta$$

$$V_{ar}(\hat{\beta}) = \sigma^2(\underline{X}'\underline{X})^{-1}$$

A única mudança é no coeficiente de determinação, isto é,

$$R_{o\text{ }23\text{ }k}^2 = \frac{\beta' \underline{X}' \underline{X}_o}{\underline{X}_o' \underline{X}_o}$$

Se agora dividirmos ambos os membros da (2.13) por s_o , desvio padrão de \underline{X}_o , vem:

$$\frac{x_{oi}}{s} = \hat{\beta}_{o2} \frac{s_2}{s_o} \frac{x_{2i}}{s_2} + \hat{\beta}_{o3} \frac{s_3}{s_o} \frac{x_{3i}}{s_3} + \dots + \hat{\beta}_{ok} \frac{s_k}{s_o} \frac{x_{ki}}{s_k} + \frac{e_i}{s_o} \quad (2.15)$$

onde s_2, s_3, \dots, s_k são, respectivamente, os desvios padrão de $\underline{X}_2, \underline{X}_3, \dots, \underline{X}_k$

Definindo:

$$x_{ji}^* = \frac{x_{ji}}{s_j}$$

$$x_{oi}^* = \frac{x_{oi}}{s_o}$$

$$\hat{\beta}_{oj}^* = \hat{\beta}_{oj} \frac{s_j}{s_o} \quad j = 1, 2, \dots, k$$

$$e_i^* = \frac{e_i}{s_o}$$

* a (2.15) toma-se:

$$x_{oi}^* = \hat{\beta}_{o2}^* x_{2i}^* + \hat{\beta}_{o3}^* x_{3i}^* + \dots + \hat{\beta}_{ok}^* x_{ki}^* + e_i^* \quad (2.16)$$

$\hat{\beta}_{oj}^*$ são, agora, *coeficientes de regressão padronizados* e, portanto, como já vimos anteriormente, seus valores numéricos são adimensionais.

A fim de escrever mais explicitamente $\hat{\beta}^*$, vejamos a forma que as matrizes $\underline{X}'\underline{X}$ e $\underline{X}'\underline{X}_o$ tomam nesse caso.

$$\begin{aligned} \underline{X}'\underline{X} &= \begin{bmatrix} x_{21} & x_{22} & \dots & x_{2n} \\ x_{31} & x_{32} & \dots & x_{3n} \\ \dots & \dots & \dots & \dots \\ x_{k1} & x_{k2} & \dots & x_{kn} \end{bmatrix} \begin{bmatrix} x_{21} & x_{31} & \dots & x_{k1} \\ x_{22} & x_{32} & \dots & x_{k2} \\ \dots & \dots & \dots & \dots \\ x_{2n} & x_{3n} & \dots & x_{kn} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{j=1}^n x_{2j}^2 & \sum_{j=1}^n x_{2j} x_{3j} & \dots & \sum_{j=1}^n x_{2j} x_{kj} \\ \sum_{j=1}^n x_{2j} x_{3j} & \sum_{j=1}^n x_{3j}^2 & \dots & \sum_{j=1}^n x_{3j} x_{kj} \\ \dots & \dots & \dots & \dots \\ \sum_{j=1}^n x_{2j} x_{kj} & \sum_{j=1}^n x_{3j} x_{kj} & \dots & \sum_{j=1}^n x_{kj}^2 \end{bmatrix} \end{aligned}$$

$$\underline{X}'\underline{X} = \begin{bmatrix} n & nr_{23} & nr_{24} & \dots & nr_{2k} \\ nr_{23} & n & nr_{34} & \dots & nr_{3k} \\ nr_{24} & nr_{34} & n & \dots & nr_{4k} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ nr_{2k} & nr_{3k} & nr_{4k} & \dots & n \end{bmatrix}$$

$$X'X = n^k \begin{bmatrix} 1 & r_{23} & r_{24} & \dots & r_{2k} \\ r_{32} & 1 & r_{34} & \dots & r_{3k} \\ r_{42} & r_{43} & 1 & \dots & r_{4k} \\ \dots & \dots & \dots & \dots & \dots \\ r_{k2} & r_{k3} & r_{k4} & \dots & 1 \end{bmatrix}$$

$$\underline{X}'\underline{X}_o = \begin{bmatrix} x_{21} & x_{22} & \dots & x_{2n} \\ x_{31} & x_{32} & \dots & x_{3n} \\ \dots & \dots & \dots & \dots \\ x_{k1} & x_{k2} & \dots & x_{kn} \end{bmatrix} \quad \begin{bmatrix} x_{o1} \\ x_{o2} \\ \cdot \\ \cdot \\ x_{on} \end{bmatrix}$$

$$= \begin{bmatrix} nr_{o2} \\ nr_{o3} \\ \cdot \\ \cdot \\ nr_{ok} \end{bmatrix}$$

onde $nr_{oi} = n \sum_{i=1}^n x_{oi} x_{ii}$

$$= n^k \begin{bmatrix} r_{o2} \\ r_{o3} \\ \cdot \\ \cdot \\ r_{ok} \end{bmatrix}$$

Portanto:
$$\hat{\underline{\beta}} = \begin{bmatrix} 1 & r_{23} & \dots & r_{2k} \\ r_{32} & 1 & \dots & r_{3k} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ r_{k2} & r_{k3} & \dots & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_{o2} \\ r_{o3} \\ \cdot \\ \cdot \\ r_{ok} \end{bmatrix}$$

(2.17)

3. MODELO CAUSAL COMPLETO

Consideremos um sistema completo ou fechado de variáveis linearmente relacionadas, no sentido de Wright²⁵ e de Li²⁶. Neste sentido, deve-se entender um sistema no qual cada variável ou é uma combinação linear de outras variáveis no sistema ou é um dos fatores básicos, que podem ou não ser correlacionados com outros fatores básicos do

²⁵ WRIGHT, S — *op. cit.* p 6

²⁶ LI, C. C. — *op. cit.* p 8

mesmo sistema. O sistema é formalmente fechado incluindo *todos* os fatores básicos (*causas*) e suas variáveis resultantes (*efeitos*). Por exemplo, a figura 1 ilustra uma situação na qual a variável X_0 é totalmente determinada pelos fatores básicos X_2, X_3, \dots, X_k os quais, por sua vez, não possuem nenhuma correlação entre si. Nesse esquema, a influência de X_2 sobre X_0 é representada por uma seta, partindo de X_2 e apontando na direção de X_0 , o mesmo sucedendo com X_3, \dots, X_k . Admitindo-se que os graus de influência de X_2, X_3, \dots, X_k sobre X_0 possam ser quantificados, o passo seguinte consiste na quantificação dessas influências que denotaremos por $C_{o2}, C_{o3}, \dots, C_{ok}$, respectivamente. Uma vez que X_2, X_3, \dots, X_k são não correlatas, a única maneira delas determinarem X_0 é diretamente.

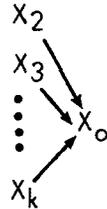


fig 1

No caso da figura 2 em que se admite que os fatores básicos X_2, \dots, X_k estão correlacionados entre si, X_0 pode ser determinada por X_2 diretamente e também por X_2 via X_3 , por X_3 e também por X_3 via X_2 e assim por diante.

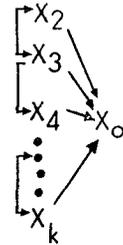


fig 2

Convém aqui observar que as hipóteses concernentes à existência de correlações podem, em muitas situações, ser removidas em parte ou totalmente pela adoção de um delineamento que conduza a uma situação em que as correlações entre X_2, X_3, \dots, X_k são nulas, uma vez que os valores dos X são fixados *a priori*.

Em ambos os casos podemos escrever:

$$X_0 = C_{o1} X_1 + C_{o2} X_2 + \dots + C_{ok} X_k \quad \text{com } X_1 = 1 \quad (3.1)$$

A (3.1) pode ainda ser escrita:

$$X_0 - \mu_0 = C_{o2} (X_2 - \mu_2) + \dots + C_{ok} (X_k - \mu_k)$$

onde $\mu_0, \mu_2, \dots, \mu_k$ são as médias de X_0, X_2, \dots, X_k , respectivamente. Ou ainda:

$$\frac{X_0 - \mu_0}{\sigma_0} = C_{o2} \frac{\sigma_2}{\sigma_0} \frac{X_2 - \mu_2}{\sigma_2} + \dots + C_{ok} \frac{\sigma_k}{\sigma_0} \frac{X_k - \mu_k}{\sigma_k}$$

onde $\sigma_0, \sigma_2, \dots, \sigma_k$ são os desvios padrão de X_0, X_2, \dots, X_k , respectivamente.

Definindo $P_{oj} = C_{oj} \frac{\sigma_j}{\sigma_0}$ ($j = 2, 3, \dots, k$) e usando letras minúsculas para representar as variáveis reduzidas, tem-se:

$$x_0 = p_{o2} x_2 + p_{o3} x_3 + \dots + p_{ok} x_k \quad (3.2)$$

Considerando-se, agora, uma série de $(k - 1)$ observações feitas sobre X_0, X_2, \dots, X_k então, o valor de x_0 para a i -ésima observação será, de acordo com a (3.2):

$$x_{0i} = p_{02} x_{2i} + p_{03} x_{3i} + \dots + p_{0k} x_{ki} \quad (3.3)$$

Ao escrevermos a (3.3) estamos supondo uma situação onde não só o modelo é fechado mas ainda admitimos que as observações feitas sobre X_0, X_2, \dots, X_k satisfazem exatamente à (3.3), isto é, não existem perturbações devidas a erros de observação. A (3.3) é, portanto, um modelo causal completo ou determinístico. Conquanto esta situação possa parecer de pouco interesse, ela é aqui considerada e analisada em detalhe, uma vez que, como mostraremos a seguir, os modelos causais encontram aqui suas raízes. Multiplicando-se ambos os membros de (3.3), sucessivamente, por $x_{2i}, x_{3i}, \dots, x_{ki}$ e somando-se para $i = 1, 2, \dots, k - 1$, pois existem $(k - 1)$ incógnitas P_{02}, \dots, P_{0k} , vem:

$$\begin{aligned} \sum_{i=1}^{k-1} x_{0i} x_{2i} &= \sum_{i=1}^{k-1} \sum_{j=2}^k p_{0j} x_{ji} x_{2i} \\ \sum_{i=1}^{k-1} x_{0i} x_{3i} &= \sum_{i=1}^{k-1} \sum_{k=2}^k p_{0j} x_{ji} x_{3i} \\ &\vdots \\ \sum_{i=1}^{k-1} x_{0i} x_{ki} &= \sum_{i=1}^{k-1} \sum_{j=2}^k p_{0j} x_{ji} x_{ki} \end{aligned} \quad (3.4)$$

Ou seja:

$$\begin{aligned} r_{02} &= \sum_{j=2}^k p_{0j} r_{2j} \\ r_{03} &= \sum_{j=2}^k p_{0j} r_{3j} \\ &\vdots \\ r_{0k} &= \sum_{j=2}^k p_{0j} r_{kj} \end{aligned} \quad (3.5)$$

que constitui um sistema de $(k - 1)$ equações a $(k - 1)$ incógnitas P_{02}, \dots, P_{0k} .

No primeiro caso considerado em que os fatores x_2, \dots, x_k não apresentam correlação entre si, a (3.5) torna-se:

$$\begin{aligned} P_{02} &= r_{02} \\ P_{03} &= r_{03} \\ &\vdots \\ P_{0k} &= r_{0k} \end{aligned} \quad (3.6)$$

ou seja, os coeficientes que medem a intensidade com que cada um dos fatores X_1, \dots, X_k determinam X_0 são os próprios coeficientes de correlação entre estes X_0 . Entretanto, é preciso aqui não esquecer o que já foi salientado na Seção 2 ou seja que a consideração de correlações só tem sentido quando se está considerando distribuição conjunta de variáveis. Os r têm, portanto, aqui apenas um significado de números calculados segundo a fórmula genérica

$$r_{zw} = \frac{\sum zw}{n}$$

onde z e w são variáveis reduzidas.

A falta desta consideração é que levou vários autores, como já mencionamos na Introdução, a questionar os coeficientes presentes nos esquemas causais pelo fato deles, contendo uma noção de sentido da influência de uma variável sobre outra que os torna não simétricos, serem iguais a coeficientes de correlação que, ao contrário, são simétricos parecendo paradoxal, portanto, o sistema (3.6).

Tomando-se a (3.3) e calculando-se a variância de x_0 tem-se para o caso de não correlação entre as variáveis:

$$V_{ar}(x_{0i}) = \sum_{j=2}^k p_{0j}^2 V_{ar}(x_{ji})$$

uma vez que P_{0j} são parâmetros e não variáveis aleatórias como é o caso dos $\hat{\beta}$ do modelo da regressão. Mas desde que os x são variáveis reduzidas, tem-se o conhecido teorema de Wright²⁷.

$$1 = \sum_{j=2}^k p_{0j}^2 \quad (3.7)$$

isto é, p_{0j}^2 representa a parcela ou fração da variância de x_0 que é explicada diretamente pela variável x_j . Isto fica claro também se pensarmos em p como sendo numericamente igual a um coeficiente de correlação, conforme (3.6), isto é:

$$1 = \sum_{j=2}^k r_{0j}^2$$

onde o quadrado de cada coeficiente de correlação de ordem zero entre duas variáveis representa um coeficiente de determinação, isto é, que fração da variação de uma das variáveis é explicada pela outra. O resultado (3.7) enfatiza mais uma vez o fato de que em sendo a (3.1) um modelo determinístico, X_0 é totalmente determinada por X_1, \dots, X_k e por isso sua variabilidade depende exclusivamente da variabilidade de X_1, \dots, X_k .

²⁷ WRIGHT, S — op. cit p 6

No segundo caso, isto é, em que as correlações entre os fatores básicos não são necessariamente nulas, a (3.5) mostra de forma explícita os efeitos diretos e indiretos dos fatores sobre X_o . De fato, a primeira equação da (3.5):

$$r_{o2} = p_{o2} + p_{o3} r_{23} + p_{o4} r_{24} + \dots + p_{ok} r_{2k}$$

mostra que a correlação entre x_o e x_2 é a soma dos valores de $(k - 1)$ trajetórias ou vias que as conectam, isto é, r_{o2} foi decomposta em $(k - 1)$ componentes: P_{o2} que mede a influência direta de X_2 sobre X_o e a soma restante que mede o efeito indireto, via X_3, X_4, \dots, X_k . O mesmo pode ser pensado com relação a X_3 e depois X_4 , etc., cuja partição dos efeitos encontra-se na segunda, terceira, etc. equações do sistema (3.5). Segundo Li²⁸, esta possibilidade de decomposição de uma correlação em suas várias componentes, isto é, uma "análise da correlação" constitui, no campo da correlação, o correspondente da "análise de variância" no campo da comparação de médias e torna a análise de tipo causal um método bastante desejável.

Neste caso, podemos escrever para $V_{ar.}(x_{oi})$ pela (3.2):

$$V_{ar.}(X_{oi}) = \sum_{j=2}^k p_{oj}^2 + 2 \sum_{j \neq l=2}^k p_{oj} p_{ol} r_{jl}$$

ou:

$$1 = \sum_{j=2}^k p_{oj}^2 + 2 \sum_{j \neq l=2}^k p_{oj} p_{ol} r_{jl} \quad (3.8)$$

Como se vê, agora a variância total de x_o é explicada em parte pelos efeitos diretos de x_2, \dots, x_k sobre x_o e em parte pelos efeitos indiretos via correlações entre x_2, \dots, x_k . É claro que neste caso o efeito direto de cada um dos fatores sobre x_o é menor do que no caso anterior sendo no máximo igual àquele.

A solução da (3.5) dará os valores de $P_{o2}, P_{o3}, \dots, P_{ok}$. De fato, em notação matricial

$$\underline{P} = \underline{R}^{-1} \underline{R}_o \quad (3.9)$$

onde

$$\underline{R}_o = \begin{bmatrix} r_{o2} \\ r_{o3} \\ \cdot \\ \cdot \\ r_{ok} \end{bmatrix} \quad p = \begin{bmatrix} p_{o2} \\ p_{o3} \\ \cdot \\ \cdot \\ p_{ok} \end{bmatrix}$$

$$\underline{R} = \begin{bmatrix} 1 & r_{23} & \dots & r_{2k} \\ r_{32} & 1 & \dots & r_{3k} \\ \dots & \dots & \dots & \dots \\ r_{k2} & r_{k3} & \dots & 1 \end{bmatrix}$$

²⁸ LI, C. C. — op cit p. 8

Como se vê, a (3.9) nada mais é do que (2.17) do modelo de regressão múltipla, isto é, o vetor p é idêntico ao vetor $\hat{\beta}$, isto é, suas componentes correspondentes são numericamente iguais.

Este fato levou muitos autores a afirmarem que os coeficientes de um modelo causal nada mais são do que coeficientes de regressão padronizados. Entretanto, não se pode pensar em regressão neste caso em que, pela própria natureza do problema, $e'e$ da Seção 2 seria necessariamente:

$$e'e = 0$$

logo

$$\underline{X}'_o \underline{X}_o - \hat{\beta}' \underline{X}' \underline{X}_o = 0$$

$$\hat{\beta}' \underline{X}' \underline{X}_o = \underline{X}'_o \underline{X}_o$$

e, portanto, o coeficiente de determinação é necessariamente:

$$R_{o \text{ vs } k}^2 = 1$$

que caracteriza um sistema fechado.

4. ANÁLISE DE TRAJETÓRIAS²⁹ COMO UM MODELO CAUSAL COMPLETO

As situações reais que ocorrem na natureza são, sem dúvida, as que conduzem à consideração de modelos estatísticos de regressão, ao invés de modelos matemáticos ou determinísticos. Isto decorre do fato de o nosso conhecimento ou a disponibilidade das informações a respeito de uma variável nunca ser total, no sentido de que não somos capazes de estabelecer *a priori* o conjunto completo de todas as causas que a determinam, nem sequer de evitar a inerente variabilidade introduzida pela mensuração.

Assim, ao estabelecer um modelo de regressão, a variabilidade da variável dependente que ainda persiste, ou seja, que não pode ser explicada pelo conjunto fixado de variáveis independentes ou é decorrente da não enumeração exaustiva destas ou advém de erros de mensuração atribuíveis aos instrumentos de medidas utilizados ou, o que é mais real, à combinação destas duas fontes de variação. Isto quando, por hipótese, a forma do modelo (linear, aditivo, etc.) não está em discussão. Caso contrário, desvios desta pressuposições afetam também a referida variabilidade. De fato, se uma variável X_o (reduzida) é expressa como

²⁹ "Path Analysis"

dependente, linearmente, de outras variáveis (reduzidas) X_2, X_3, \dots, X_k e se o modelo fosse completo, teríamos:

$$X_o = p_{o2} X_2 + \dots + p_{ok} X_k \quad \text{conforme (3.2).}$$

A realidade, todavia, mostra que, pelas razões acima referidas, a variável X_o apresenta uma variabilidade não explicável por X_2, \dots, X_k , portanto devemos ter:

$$X_o = p_{o2} X_2 + \dots + p_{ok} X_k + \mu$$

conforme (2.2), onde μ é exatamente a medida desta perturbação adjudicável aos aspectos supramencionados. Num certo sentido, então, μ é uma medida da nossa “desinformação” e da falta de fidedignidade de nossas medidas, ou ainda da nossa “ignorância” a respeito dos mecanismos determinantes de X_o , a saber, uma medida do mundo exterior ao considerado através de X_2, \dots, X_k e que não foi explicitado.

Neste ponto satisfazem-se os estatísticos que nada mais tentam senão aprimorar o modelo, tentando minimizar o desconhecido, pela procura de mais variáveis explicativas.

Uma posição pragmática seria a de imaginar que o universo do desconhecido, conquanto inatingível, possa ser expresso pela introdução de uma variável que o meça, X_μ , e completar, formalmente, o modelo. Isto é, introduzir a componente $p_{o\mu} X_\mu$ tornando:

$$X_o = p_{o2} X_2 + \dots + p_{ok} X_k + p_{o\mu} X_\mu \quad (4.1)$$

Isto significa, em última análise, um retorno a um modelo determinístico pois preferem os pragmáticos, em vez de admitir a nossa ignorância a respeito do fenômeno, supor a existência de uma causa unificada de influência do mundo externo, expressa por uma variável X_μ conquanto, obviamente, esta não seja mensurável diretamente.

Este procedimento, de completar artificialmente o modelo causal, é característica fundamental da Análise de Trajetórias.

Com isto, estamos exatamente diante da situação descrita na Seção anterior, pois o sistema foi tornado completo e sem erros experimentais já que esses estão incluídos em X_μ . Podem, portanto, ser aplicados todos os resultados obtidos, introduzindo formalmente mais uma variável ao modelo.

Todavia, a determinação de $p_{o\mu}$ não é possível, como acontece para os demais p , pelo fato de que agora o sistema (3.5) assume a forma:

$$\begin{aligned} r_{o2} &= \sum_{j=2}^k p_{oj} r_{2j} + p_{o\mu} r_{2\mu} \\ &\vdots \\ r_{ok} &= \sum_{j=2}^k p_{oj} r_{kj} + p_{o\mu} r_{k\mu} \\ r_{o\mu} &= \sum_{j=2}^k p_{oj} r_{\mu j} + p_{o\mu} \end{aligned} \quad (4.2)$$

na, qual os valores de $r_{0\mu}$, $r_{2\mu}$, ..., $r_{k\mu}$ são desconhecidos, uma vez que X_μ não é mensurável diretamente. Teríamos, assim, um sistema de k equações a $2k$ incógnitas:

$$p_{02}, p_{03}, \dots, p_{0k}, p_{0\mu}, r_{0\mu}, r_{2\mu}, \dots, r_{k\mu}.$$

O expediente de recorrer à equação (3.8), que, neste caso, explicitando os termos que contém o subscrito μ , torna-se:

$$1 = \sum_{j=2}^k p_{0j}^2 + p_{0\mu}^2 + 2 \sum_{j=l=2}^k p_{0j} p_{0l} r_{jl} + 2 \sum_{j=2}^k p_{0j} p_{0\mu} r_{j\mu} \quad (4.3)$$

também não resolve o problema pois o valor de $p_{0\mu}$ continuaria dependendo dos valores das correlações entre X_μ e as demais variáveis.

É para contornar este tipo de dificuldade que é imposta a condição de que X_μ seja não correlacionada com X_2, X_3, \dots, X_k , isto é,

$$r_{2\mu} = r_{3\mu} = \dots = r_{k\mu} = 0 \quad (4.4)$$

com isto, a (4.3) permite, agora, a obtenção do valor de $p_{0\mu}^2$ e, portanto, de $p_{0\mu}$.

De fato, da (4.3) segue:

$$p_{0\mu}^2 = 1 - \sum_{j=2}^k p_{0j}^2 - 2 \sum_{j=l=2}^k p_{0j} p_{0l} r_{jl} \quad (4.5)$$

onde $p_{02}, p_{03}, \dots, p_{0k}$ são dados pela (4.2), após substituir $r_{2\mu}, r_{3\mu}, \dots, r_{k\mu}$ por zero.

A imposição da nulidade das correlações entre X_μ e X_2, \dots, X_k torna a estrutura causal completamente identificável. É neste sentido que Boudon³⁰ afirma que "quand les correlations entre les facteurs implicites sont nulles, les coefficients de dépendence sont toujours identifiables".

A interpretação de $P_{0\mu}^2$ permite fornecer uma medida de toda a influência externa não explicitada em X_2, \dots, X_k , se o que restou no mundo exterior ao sistema *não for correlacionado* com o já explicitado. Em outras palavras, X_μ estaria contendo toda uma outra classe de informações não explicitadas no modelo.

As implicações da (4.4) tornam, a nosso ver, extremamente vulneráveis os modelos causais completos. Por esse motivo, suas aplicações devem estar condicionadas a sérias e exaustivas considerações preliminares específicas a cada caso.

³⁰ BOUDON, R — *op. cit.* p 10

INTERPRETAÇÃO DOS MODELOS

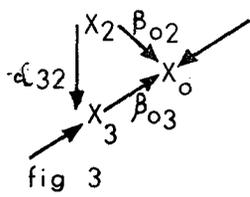
MODELO	UNIVERSO CONHECIDO DESCONHECIDO	EFEITO INTRO- DUZIDO PELA DEXPERIMEN- TAÇÃO DISTUR- BÂNCIA u	CARACTE- RÍSTICAS Variabilidade residual e coef. de de- terminação
LEI (DETERMINÍSTICO)	$\begin{array}{ c c } \hline X_2, & X_k \\ \hline \end{array} \quad \left \quad \begin{array}{ c } \hline \{\phi\} \\ \hline \end{array} \right.$	$\sigma^2 = 0$ X_o $R^2 = 1$	$\sigma^2 > 0$ $R^2 < 1$
ESTATÍSTICO (REGRESSÃO)	$\begin{array}{ c c } \hline X_2, & X_k \\ \hline \end{array} \quad \left \quad \begin{array}{ c } \hline \{?\} \\ \hline \end{array} \right.$	$\sigma^2 > 0$ X_o $R^2 < 1$	$\sigma^2 > 0$ $R^2 < 1$
CAUSAL (ANÁLISE DE TRAJETÓRIA)	$\begin{array}{ c c } \hline X_2, & X_k \\ \hline \end{array} \quad \left \quad \begin{array}{ c } \hline X_u \\ \hline \end{array} \right.$	$\sigma^2 = 0$ X_o $R^2 = 1$	$\sigma^2 = 0$ $R^2 = 1$

5. MODELOS CAUSAIS COM MAIS DE UMA EQUAÇÃO

Como já salientamos quando tratamos dos objetivos deste trabalho, daríamos ênfase, no estudo substantivo, ao modelo descrito através de apenas uma equação.

A fim de frisar algumas de nossas divergências com a literatura corrente nesta área, trataremos, a título de exemplo, do caso em que o modelo proposto inclui apenas duas equações simultâneas.

1.º Caso — O modelo compreende duas equações estatísticas



Suponhamos a situação representada pela figura 3 e que se traduz em termos de variáveis reduzidas, de acordo com a (2.16), por:

$$x_{0i} = \hat{\beta}_{02} x_{2i} + \hat{\beta}_{03} x_{3i} + e_{0i} \quad (5.1)$$

$$x_{3i} = \hat{\alpha}_{32} x_{2i} + e_{3i} \quad (5.2)$$

Os valores dos coeficientes de regressão da (5.1), de acordo com a (2.17) valem:

$$\hat{\beta}_{02} = \frac{r_{02} - r_{03} r_{23}}{1 - r_{23}^2} \quad (5.3)$$

$$\hat{\beta}_{03} = \frac{r_{03} - r_{02} r_{23}}{1 - r_{23}^2} \quad (5.4)$$

Por sua vez, na equação de regressão de x_3 em x_2 (5.2), $\hat{\alpha}_{32}$ é dado por:

$$\hat{\alpha}_{32} = r_{32} \quad (5.5)$$

Substituindo-se, agora, em (5.1) x_{3i} expressa pela (5.2) vem:

$$x_{oi} = (\hat{\beta}_{02} + \hat{\beta}_{03} \hat{\alpha}_{32}) x_{2i} + (\hat{\beta}_{03} e_{3i} + e_{oi})$$

fazendo

$$\hat{\beta}_{02} + \hat{\beta}_{03} \hat{\alpha}_{32} = \hat{\gamma}_{02}$$

$$\hat{\beta}_{03} e_{3i} + e_{oi} = \delta_{oi}$$

tem-se:

$$x_{oi} = \hat{\gamma}_{02} x_{2i} + \delta_{oi}$$

que é, agora, uma equação de regressão de x_o em x_2 e, portanto, o valor de mínimos quadrados de $\hat{\gamma}_{02}$ é r_{02} . Assim sendo, os coeficientes $\hat{\beta}_{02}$ e $\hat{\beta}_{03}$ satisfazem à condição:

$$\hat{\beta}_{02} + \hat{\beta}_{03} r_{32} = r_{02}$$

De fato, substituindo-se na expressão supra, $\hat{\beta}_{02}$ e $\hat{\beta}_{03}$ pelos valores (5.3) e (5.4), respectivamente, tem-se uma identidade:

$$\frac{r_{02} - r_{03} r_{23}}{1 - r_{23}^2} + \frac{r_{03} - r_{02} r_{23}}{1 - r_{23}^2} r_{32} = r_{02}$$

$$\frac{r_{02} - r_{03} r_{23} + r_{03} r_{23} - r_{02} r_{23}^2}{1 - r_{23}^2} = r_{02}$$

$$\frac{r_{02} (1 - r_{23}^2)}{1 - r_{23}^2} = r_{02}$$

$$r_{02} = r_{02}$$

Logo, o problema proposto sempre tem solução quando as duas equações do modelo são equações de regressão, ao contrário do que diz Turner (10) ³¹.

³¹ TURNER, M E & STEVENS, C D — *op cit* p 8

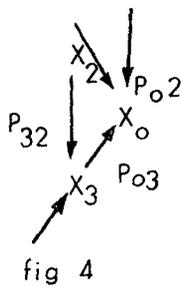
É importante observar que:

$$1) \sum_{i=1}^n e_{oi} e_{si} = 0$$

$$2) \hat{\beta}_{os}^2 \sum_{i=1}^n e_{si}^2 + \sum_{i=1}^n e_{oi}^2 = \sum_{i=1}^n \delta_{oi}^2$$

3) O coeficiente de determinação é igual a $(\hat{\beta}_{o2} + \hat{\beta}_{os} \hat{\alpha}_{s2})^2$

2.º Caso — O modelo causal completo inclui duas equações



Suponhamos que a vinculação entre as variáveis X_0 , X_2 e X_3 seja traduzida, conforme a figura 4, pelo sistema (com as variáveis já na forma reduzida):

$$x_{0i} = p_{02} x_{2i} + p_{03} x_{3i} + p_{01} x_{1i}$$

$$x_{3i} = p_{32} x_{2i} + p_{31} x_{1i}$$

Os coeficientes p_{02} e p_{03} são obtidos resolvendo-se as equações:

$$r_{02} = p_{02} + p_{03} r_{23}$$

$$r_{03} = p_{02} r_{23} + p_{03}$$

que provém do sistema (3.5) pois $r_{21} = 0$, conforme Seção 4. Isto é,

$$p_{02} = \frac{r_{02} - r_{03} r_{23}}{1 - r_{23}^2}, \quad p_{03} = \frac{r_{03} - r_{02} r_{23}}{1 - r_{23}^2}$$

Por sua vez:

$$p_{01}^2 = 1 - p_{02}^2 - p_{03}^2 - 2p_{02} p_{03} r_{23}$$

Finalmente,

$$p_{32} = r_{23}$$

$$p_{31} = \sqrt{1 - p_{32}^2}$$

$$\text{pois } r_{31} = 0$$

3.º Caso — O modelo inclui uma equação estatística e uma equação matemática.

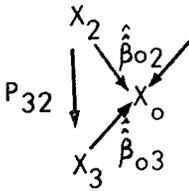


fig 5

Suponhamos que as variáveis X_0 , X_2 e X_3 estejam vinculadas pelo sistema (em forma reduzida)

$$x_{0i} = \hat{\beta}_{02} x_{2i} + \hat{\beta}_{03} x_{3i} + e_{0i}$$

$$x_{3i} = p_{32} x_{2i}$$

Neste caso, para a equação de regressão de x_0 em x_2 e x_3 , o método de mínimos quadrados dá como solução para p_{02} e p_{03} :

$$p = (\underline{X}'\underline{X})^{-1} \underline{X}'\underline{X}_0$$

onde:

$$\underline{X}'\underline{X} = \begin{bmatrix} n & nr_{23} \\ nr_{23} & n \end{bmatrix}$$

Mas, pela segunda equação do sistema, a relação entre x_3 e x_2 é exata, isto é, $r_{23} = 1$ e, portanto:

$$\underline{X}'\underline{X} = \begin{bmatrix} n & n \\ n & n \end{bmatrix}$$

cujo determinante é zero. Portanto, $(\underline{X}'\underline{X})^{-1}$ não está definida e o vetor coluna p dos coeficientes p_{02} e p_{03} não pode ser obtido, isto é, estes são indeterminados.

NOTAS PARA O CÁLCULO DOS ÍNDICES CARACTERÍSTICOS DE UMA DISTRIBUIÇÃO DE FREQUÊNCIA

Prof. Lauro Sodré Viveiros de Castro
da Escola Nacional de Ciências Estatísticas

Já vimos em artigo anterior* o cálculo dos Momentos de uma Distribuição de Frequência.

No exemplo dado, os quatro Momentos Centrados foram obtidos com os resultados abaixo:

X (gramas)	f (peças)	
1900 - 1930 .	8	$\left\{ \begin{array}{l} \bar{X} = 2\ 012,5 [g] \\ \sigma = \sqrt{n_2} = 35,1 [g] \\ n_1 = 0 \\ n_2 = 1.232,10 [g^2] \\ n_3 = 1.379,174, [g^3] \\ n_4 = 4\ 293\ 612,5333 [g^4] \end{array} \right.$
1930 - 1960 ...	55	
1960 - 1990 . .	192	
1990 - 2020 .	332	
2020 - 2050 .	275	
2050 - 2080	114	
2080 - 2110	24	
—	1 000	

* Revista Brasileira de Estatística n.º 135, jul/set 1973

Vamos ver agora que fazer com esses Momentos, para que as informações assim estabelecidas possam ter comparabilidade com as de outras distribuições e também, dentro da própria distribuição, sirvam para indicar a que tipo de curva a mesma se adapta. Além do mais, para possibilitar tabelamentos, é mister a passagem das unidades concretas para medidas abstratas.

Como decorre do trabalho já referido, os quatro "Momentos Centrados" (em relação à média aritmética) têm como expressão geral:

$$n_r = \frac{\sum x^r f}{\sum f}$$

e são medidos em unidades concretas, no caso *gramas* nas potências 1, 2, 3 e 4.

Para o objetivo que acabamos de enunciar, é necessário que os "braços de alavanca" \underline{x} não sejam medidos em unidades concretas, e sim tornados abstratos mediante a adoção de uma unidade-padrão universal, que será o desvio padrão σ (que também tem o nome de "unidade-padrão" justamente por isso).

Assim x será expresso como $t = \frac{x}{\sigma}$, isto é, não mais será medido em *gramas*, mas sim em *desvios-padrões*.

Nos dois primeiros Momentos (exprimem POSIÇÃO e DISPERSÃO) há necessidade da conservação de unidades concretas: média aritmética e desvio-padrão, a primeira porque é o elo que relaciona a distribuição com o meio externo, e o segundo porque vai ser a nova unidade de medida.

Para os índices subseqüentes (ASSIMETRIA e ACHATAMENTO), que vão ser simbolizados respectivamente por $\sqrt{\beta_1}$ e β_2 , teremos:

a) ÍNDICE DE ASSIMETRIA:

$$\sqrt{\beta_1} = \frac{\sum \left(\frac{x}{\sigma}\right)^3 f}{\sum f} = \frac{\sum x^3 f}{\sigma^3 \sum f} = \frac{n_3}{\sigma^3} = \frac{n_3}{n_2^{1,5}}$$

no caso presente: $\sqrt{\beta_1} = -0,0319$

isto é, a nossa distribuição está bem próxima de uma simetria perfeita, que ocorre quando $\sqrt{\beta_1} = 0$

É de se notar, também, que β_1 conservará o sinal de n_3 . O nosso β_1 será, então, igual a $-0,00102$.

b) ÍNDICE DE ACHATAMENTO (CURTOSE):

$$\beta_2 = \frac{\sum \left(\frac{x}{\sigma}\right)^4 f}{\sum f} = \frac{\sum x^4 f}{\sigma^4 \sum f} = \frac{n_4}{\sigma^4} = \frac{n_4}{n_2^2}$$

no caso em estudo $\beta_2 = 2,828$

O índice β_2 varia em torno de 3, que é a sua base de comparação por ser $\beta_2 = 3$ na Curva Normal de Probabilidades, como veremos mais tarde.

Assim, quando $\beta_2 = 3$, a distribuição terá uma curtose normal, e se dirá MESOCÚRTICA.

Se $\beta_2 < 3$, a distribuição é mais achatada (menos afilada) que a normal, e será dita PLATICÚRTICA.

Ao contrário, se $\beta_2 > 3$, a distribuição será menos achatada, e terá o nome de LEPTOCÚRTICA.

Dessas duas últimas condições estabelecidas se vê que o índice mede na verdade o afilamento, e não o achatamento, pois crescendo β_2 , é maior o afilamento e não o achatamento.

* * *

Uma notação diferente pode trazer confusões ao estudante. De fato, em alguns manuais se encontra a notação:

$$\alpha_3 = \sqrt{\beta_1}$$

$$\alpha_4 = \beta_2$$

sendo que, em outros livros, o índice de achatamento é reduzido para zero, fazendo-se

$$\alpha_4 = \beta_2 - 3$$

e ficando como constantes:

$$\alpha_1 = n_1 = 0 \quad (\text{Posição})$$

$$\alpha_2 = \frac{n_2}{\sigma^2} = 1 \quad (\text{Dispersão})$$

Isso dará lugar a denominação “zero-um” a uma distribuição com a origem na média aritmética e tendo como unidade de medida o desvio-padrão.

* * *

Temos então, resumindo, os quatro índices característicos da distribuição estudada:

POSIÇÃO: $\bar{X} = 2.015,5 \text{ gramas}$

DISPERSÃO: $\sigma = 35,1 \text{ gramas}$

ASSIMETRIA: $\sqrt{\beta_1} = -0,03 \text{ (adimensional)}$

ACHATAMENTO: $\beta_2 = 2,83 \text{ (adimensional)}$

* * *

Uma ligeira verificação pode ser feita quanto ao valor numérico do desvio-padrão: em distribuições em talho-*i*, com um número razoável de observações (maior do que 100), a amplitude total deve dar, aproximadamente, o sextuplo do desvio padrão.

Na tabela estudada, a amplitude total é

$$2.110 - 1.900 = 210 \text{ gramas}$$

dando, assim, uma ordem de grandeza para o desvio-padrão de 35 gramas. Essa consideração provém de uma analogia com a Curva Normal (como será visto mais tarde), onde seis desvios-padrões cobrem praticamente toda a área sob a curva.

* * *

Por oportuno, voltaremos a examinar o cálculo dos Momentos Centrados, pois o "3.º MODO" referido no trabalho anterior foi tratado, para aliviar a matéria, de um modo muito superficial. Entretanto, como veremos no 3.º MODO, aparecem, como trânsito para os Momentos desejados, as médias do 1.º, 2.º, 3.º e 4.º graus e que nos levam a considerações interessantes.

De fato, partindo da expressão geral dos Momentos Centrados que vimos no início deste trabalho:

$$n_r = \frac{\sum x^r f}{\sum f}$$

e substituindo \underline{x} pelo seu correspondente $X - \bar{X}$, temos para os 4 primeiros momentos:

$$1^\circ) r = 1 \quad n_1 = \frac{\sum (X - \bar{X})f}{\sum f} = 0$$

$$2^\circ) r = 2 \quad n_2 = \frac{\sum (X - \bar{X})^2 f}{\sum f} = \frac{\sum X^2 f}{\sum f} - \bar{X}^2$$

$$3^\circ) r = 3 \quad n_3 = \frac{\sum (X - \bar{X})^3 f}{\sum f} = \frac{\sum X^3 f}{\sum f} - 3\bar{X} \frac{\sum X^2 f}{\sum f} + 2\bar{X}^3$$

$$4^\circ) r = 4 \quad n_4 = \frac{\sum (X - \bar{X})^4 f}{\sum f} = \frac{\sum X^4 f}{\sum f} - 4\bar{X} \frac{\sum X^3 f}{\sum f} + 6\bar{X}^2 \frac{\sum X^2 f}{\sum f} - 3\bar{X}^4$$

sendo de se notar que as frações que aparecem são:

$$\frac{\sum X^2 f}{\sum f} \rightarrow \text{quadrado da média quadrática, ou do 2.º grau;}$$

$$\frac{\sum X^3 f}{\sum f} \text{ cubo da média cúbica, ou do 3.º grau;}$$

$$\frac{\sum X^4 f}{\sum f} \text{ 4.ª potência da média do 4.º grau.}$$

Calculados os valores que estão esquematizados no quadro de cálculo do "3.º MODO" do trabalho anterior, encontramos:

$$\sum f = 1.000$$

$$\sum X f = 2.012\ 470$$

$$\sum X^2 f = 4.051.267\ 600$$

$$\sum X^3 f = 8\ 158\ 012.252\ 750$$

$$\sum X^4 f = 16\ 432.721.020\ 555.000$$

o que nos permite calcular as médias:

1) média aritmética $\bar{X}_1 = 2\ 012,47 \text{ gramas}$

2) média quadrática $\bar{X}_2 = \sqrt{4.051.267,60} = 2.012,77 \text{ gramas}$

3) média cúbica $\bar{X}_3 = \sqrt[3]{8.158.012.252,750} = 2.013,08 \text{ gramas}$

4) média do 4.º grau $\bar{X} = \sqrt[4]{16.432\ 721\ 020.555} = 2.013,39 \text{ gramas}$

Sendo os índices característicos obtidos por meio dessas médias, como visto:

$$n_1 = 0 \qquad \bar{X} = 2\ 012,47 \text{ g}$$

$$n_2 = 1.232,10 \qquad \sigma = 35,1 \text{ g}$$

$$n_3 = -1\ 379,174 \qquad \sqrt{\beta_1} = -0,03$$

$$n_4 = 4\ 293\ 612,5333 \qquad \beta_2 = 2,83$$

Desse modo de calcular se conclui que, dadas as médias de 1.º, 2.º, 3.º e 4.º graus, que à primeira vista parecem indicar só a característica POSIÇÃO, na realidade e quando apresentadas como um grupo de medidas, exprimem também a DISPERSÃO, a ASSIMETRIA e o ACHATAMENTO.

* * *

Nesta altura, parece-nos interessante repetir o que já escrevemos há algum tempo (*):

“... medimos a Posição, a Dispersão, a Assimetria e o Achatamento de uma distribuição por meio das médias aritmética, quadrática, cúbica e biquadrática (...), admitir-se-á que a média de grau 1,5 possa medir um fato intermediário, ou de transição, entre Posição e Dispersão? Da mesma forma, as médias 2,1 2,2 ... irão exprimir características que seriam verdadeiras misturas de Dispersão e Assimetria?

Haverá uma transição lenta e contínua entre Assimetria e Achatamento?

* * *

Seguem-se cinco Exercícios práticos para melhor consolidação da matéria exposta.

EXERCÍCIO N.º 1

Calcular os índices característicos da distribuição ao lado:

X (cm)	X MÉDIO	f
0 † 2	1	15
2 † 4	3	60
4 † 6	5	88
6 † 8	7	34
8 † 10	9	3
—	—	200

Solução: Quando os valores de X se apresentam fáceis para o cálculo direto (em relação à origem natural), é vantajoso o modo de operar referido como “3.º MODO” no texto:

X_m	f	Xf	X^2f	X^3f	X^4f
1	15	15	15	15	15
3	60	180	540	1 620	4 860
5	88	440	2 200	11 000	55 000
7	34	238	1 666	11 662	81 634
9	3	27	243	2 187	19 683
—	200	900	4 664	26 484	161 192

* Revista Brasileira de Estatística — n.º 81/82 — 1960 — pág. 72

Embora não necessárias ao problema, podemos ter as médias de graus 1, 2, 3 e 4:

$$\left\{ \begin{array}{l} \text{m\u00e9dia aritm\u00e9tica: } M.A. = \frac{\sum Xf}{\sum f} = \frac{900}{200} = 4,50 \text{ cm} \\ \text{m\u00e9dia quadr\u00e1tica: } M.Q. = \sqrt{\frac{\sum X^2f}{\sum f}} = \sqrt{23,32} = 4,8291 \text{ cm} \\ \text{m\u00e9dia c\u00fabica: } M.C. = \sqrt[3]{\frac{\sum X^3f}{\sum f}} = \sqrt[3]{132,42} = 5,0970 \text{ cm} \\ \text{m\u00e9dia biquadr\u00e1tica: } M.B. = \sqrt[4]{\frac{\sum X^4f}{\sum f}} = \sqrt[4]{805,96} = 5,3282 \text{ cm} \end{array} \right.$$

Voltando aos Momentos Centrados:

$$n_1 = 0$$

$$n_2 = \frac{\sum (X - \bar{X})^2 f}{\sum f} = (M.Q.)^2 - (M.A.)^2 = 3,07 \text{ cm}^2$$

$$n_3 = \frac{\sum (X - \bar{X})^3 f}{\sum f} = (M.C.)^3 - 3(M.A.) (M.Q.)^2 + 2(M.A.)^3 = -0,15 \text{ cm}^3$$

$$n_4 = \frac{\sum (X - \bar{X})^4 f}{\sum f} = (M.B.)^4 - 4(M.A.) (M.C.)^3 + 6(M.A.)^2 (M.Q.)^2 - 3(M.A.)^4 = 25,5925 \text{ cm}^4$$

e os \u00cdndices Caracter\u00edsticos:

POSI\u00c7\u00c3O: $\bar{X} = 4,50 \text{ cm}$

DISPERS\u00c3O: $\sigma = \sqrt{n_2} = 1,752 \text{ cm}$

ASSIMETRIA: $\sqrt{\beta_1} = \frac{n_3}{\sigma^3} = -0,028 \text{ (abstrato)}$

ACHATAMENTO: $B_2 = \frac{n_4}{\sigma^4} = 2,715 \text{ (abstrato)}$

EXERC\u00cdCIO N.º 2

O \u00cdndice de assimetria obtido na distribui\u00e7\u00e3o de freq\u00fc\u00eancia do EXERC\u00cdCIO anterior \u00e9 praticamente nulo. Isto quer dizer que a curva representativa ser\u00e1 sim\u00e9trica. Entretanto, um exame das freq\u00fc\u00eancias da tabela dada n\u00e3o nos d\u00e1 essa impress\u00e3o. Como se explica?

Resposta: A simetria ocorre em torno da m\u00e9dia aritm\u00e9tica ($\bar{X} = 4,50 \text{ cm}$) e n\u00e3o em refer\u00eancia aos pontos centrais das classes. Se

fizéssemos um reagrupamento das 200 observações tendo a média aritmética como ponto médio da classe central (ou ponto inicial ou final entre as classes do meio), as frequências seriam praticamente simétricas. Veja-se o EXERCÍCIO seguinte.

EXERCÍCIO N.º 3

Com os índices característicos da distribuição dada no EXERCÍCIO n.º 1, reconstruí-la em classes de 1 cm.

Solução: Como os valores de $\sqrt{\beta_1} = 0$ e $\beta_2 = 3$ são os índices da Curva Normal, podemos adotá-la para curva ajustante.

Admitiremos, também, que possa haver uma classe abaixo de zero, hipótese que será anulada se o caso real indicar.

Como a média aritmética vai cair exatamente no meio da classe “4,0 a 5,0 cm”, bastará fazer o cálculo para um lado da curva, que será repetido simetricamente para o outro lado.

$$h = 1,0 \text{ cm}$$

$$\bar{X} = 4,50 \text{ cm}$$

$$\sigma = 1,75 \text{ cm}$$

I	II	III	IV	V	VI
	$x = X - \bar{X}$ Inicial Final	x/σ	Área Lida na Tabela	Área na Classe	Frequência Absoluta Teórica
Abaixo de 0 .	— ∞	— ∞	0,50 000	0,00 508	1,016
0 † 1.....	— 4,50	— 2,57	0,49 492	0,01 767	3,534
1 † 2.....	— 3,50	— 2,00	0,47 725	0,05 361	10,722
2 † 3. . .	— 2,50	— 1,43	0,42 364	0,11 853	23,706
3 † 4.....	— 1,50	— 0,86	0,30 511	0,19 102	38,204
4 † \bar{X}	— 0,50	— 0,29	0,11 409	0,22 818	45,636
\bar{X} † 5.....	+ 0,50	+ 0,29	0,11 409		
5 † 6.....	+ 1,50	+ 0,86	0,30 511	0,19 102	38,204
6 † 7. . .	+ 2,50	+ 1,43	0,42 364	0,11 853	23,706
7 † 8. . .	+ 3,50	+ 2,00	0,47 725	0,05 361	10,722
8 † 9.....	+ 4,50	+ 2,57	0,49 492	0,01 767	3,534
9 † 10....	+ 5,50	+ 3,14	0,49 916	0,00 424	0,848
10 e mais.....	+ ∞	+ ∞	0,50 000	0,00 084	0,168
—	—	—	—	1,00 000	200,000

Esclarecimento sobre a tabela de cálculos: As colunas I e II são compreensíveis por elas mesmas; foi incluída uma classe antes da primeira e outra além da última para recolher os resíduos que, teóricamente, pudessem existir. A classe que contém a média foi separada em duas partes, pois estas se localizam uma de cada lado da curva. No caso deste Exercício, especialíssimo, a média está bem no meio da classe: por isso, a 2a. metade da col. II será simétrica à 1.ª metade. A coluna III também pode ser calculada por diferenças (até a média) ou por acréscimos constantes e iguais a h/σ , no caso presente 0,571; o sinal — ou + não afeta os cálculos, sendo a curva simétrica. A col. IV é formada pela leitura na tabela de áreas da Curva Normal, como se vê em “Exercícios de Estatística”, do A., notando-se que a parte positiva (tem uma classe a mais) repete os dados da parte negativa pelo fato da média estar no meio da classe. A col. V, que o estudante deve gravar bem, será a PROBABILIDADE ou a FREQUÊNCIA RELATIVA TEÓRICA da classe, é obtida por subtração da área da col. IV e o valor seguinte na mesma col. IV; na classe da média, as duas leituras se somam, pois estão em lados diferentes da curva; daí por diante, a diferença é entre o valor da linha e o da linha anterior; a soma da col. V tem que resultar exatamente 1. A col. VI se forma pela multiplicação por N (no caso, 200) dos valores de col. V.

O problema está terminado. Se quisermos fazer o reagrupamento das frequências teóricas para serem comparadas com as frequências observadas, encontramos:

X cm	f OBSERVADAS	f TEÓRICAS
Abaixo de 0	—	1,0
0 † 2..	15	14,3
2 † 4.....	60	61,9
4 † 6.....	88	83,8
6 † 8.....	34	34,4
8 † 10..	3	4,4
10 e mais....	—	0,2
	200	200,0

Nota: No final deste estudo aparece uma tabela da Curva Normal que em função da *Área* dá a abscissa e a ordenada da C.N.

EXERCÍCIO N.º 4

Quais seriam os índices característicos de uma distribuição que tivesse as freqüências obtidas em ordem inversa à do EXERCÍCIO N.º 1?

Sugestão: Procure o leitor ver que a média aritmética 4,50, que distava 0,50 do ponto inicial da classe "4 - 6", passará a 0,50 aquém do ponto final: será 5,50 cm. A dispersão não se alterará. A assimetria trocará de sinal. E a curtose ... ficará a cargo do leitor descobrir.

X cm	f
0 - 2... ..	3
2 - 4. ..	34
4 - 6. . .	88
6 - 8.... . .	60
8 - 10. . . .	15
—	200

EXERCÍCIO N.º 5

Se somarmos uma constante aos valores de X, que acontece com as medidas de grau 1, 2, 3 e 4 e com os índices característicos de uma distribuição de freqüência?

Sugestão: O leitor verá que a média aritmética ficará acrescida dessa constante \underline{k} ; que as outras médias se deformarão de modo diverso: por exemplo, a média quadrática passará de:

$$(M. Q)^2 = \frac{\sum X^2 f}{\sum f}$$

para (sendo $M'Q$ a nova média quadrática):

$$(M' Q)^2 = \frac{\sum (X + k)^2 f}{\sum f} = \frac{\sum X^2 f}{\sum f} + 2k \frac{\sum X f}{\sum f} + k^2$$

isto é, o acréscimo (no quadrado da média) será: $k(2\bar{X} + k)$ sendo \bar{X} a média aritmética primitiva. Pode-se notar que o parêntese acima é a soma da média aritmética primitiva com a média aritmética nova. Quanto aos índices característicos de Dispersão, Assimetria e Achatamento, apesar de serem função das quatro médias citadas (v. EXERCÍCIO n.º 1), não devem sofrer alteração, pois são momentos de 2.ª, 3.ª e 4.ª ordem construídos em torno da média aritmética.

Para verificação, tomamos a distribuição dada no EXERCÍCIO n.º 1 e demos a X um acréscimo $k = 3$ cm.

X	X MÉDIO	f
3 5. . .	4	15
5 7. . .	6	60
7 9. . .	8	88
9 11. . .	10	34
11 13. . .	12	3
—	—	200

$$M.A)^1 = 7,50 \quad \eta_1 = 0$$

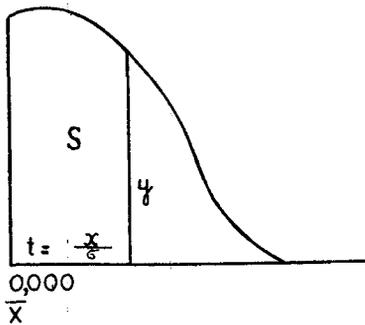
$$(M.Q)^2 = 59,32 \quad \eta_2 = 3,07$$

$$(M.C)^3 = 490,80 \quad \eta_3 = -0,15$$

$$(M.B)^4 = 4\,221,28 \quad \eta_4 = 25,5925$$

* * *

Ordenadas e abscissas da Curva Normal em função da área que subentendem



A partir da média aritmética \bar{X} , para cada área S , dada em milésimos de 0,000 até 0,500, são lidos os afastamentos (abscissas)

$$t = \frac{x}{\sigma}$$

e as ordenadas:

$$y = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}$$

O leitor encontrará, também, as potências de e na Tabela n.º 2 do Livro “Tabelas Numéricas”, do Autor.

ÁREAS ESPECIAIS

ÁREA	ABSCISSA x/σ	ORDENADA y
0,000 .	0,0000 0000	0,3989 4228
0,250. .	0,6744 8975	0,3177 7657
0,450. .	1,6448 5363	0,1031 3564
0,475 . .	1,9599 6398	0,0584 4507
0,490 . .	2,3263 4787	0,0266 5214
0,495 . . .	2,5758 2930	0,0144 5974

ÁREA 0,	ABS- CISSA x/σ	ORDE- NADA y
000	0,0000	0,3989 4
001	0025	3989 4
002	0050	3989 4
003	0075	3989 3
004	0100	3989 2
005	0125	3989 1
006	0150	3989 0
007	0175	3988 8
008	0201	3988 6
009	0226	3988 4
010	0251	3988 2
011	0,0276	0,3987 6
012	0301	3987 6
013	0326	3987 3
014	0351	3987 0
015	0376	3986 6
016	0401	3986 2
017	0426	3985 8
018	0451	3985 4
019	0476	3984 9
020	0502	3984 4
021	0,0527	0,3984
022	0552	3983
023	0577	3983
024	0602	3982
025	0627	3982
026	0652	3981
027	0677	3980
028	0702	3980
029	0728	3979
030	0753	3978
031	0,0778	0,3977
032	0803	3977
033	0828	3976
034	0853	3975
035	0878	3974
036	0904	3973
037	0929	3972
038	0954	3971
039	0979	3970
040	1004	3969
041	0,1030	0,3968
042	1055	3967
043	1080	3966
044	1105	3965
045	1130	3964
046	1156	3963
047	1181	3962
048	1206	3961
049	1231	3959
050	1257	3958

ÁREA 0,	ABS- CISSA x/σ	ORDE- NADA y
050	0,1257	0,3958
051	1282	3957
052	1307	3955
053	1332	3954
054	1358	3953
055	1383	3951
056	1408	3950
057	1434	3949
058	1459	3947
059	1484	3946
060	1510	3944
061	0,1535	0,3943
062	1560	3941
063	1586	3940
064	1611	3938
065	1637	3936
066	1662	3935
067	1687	3933
068	1713	3931
069	1738	3930
070	1764	3928
071	0,1789	0,3926
072	1815	3924
073	1840	3922
074	1866	3921
075	1891	3919
076	1917	3917
077	1942	3915
078	1968	3913
079	1993	3911
080	2019	3909
081	0,2045	0,3907
082	2070	3905
083	2096	3903
084	2121	3901
085	2147	3899
086	2173	3896
087	2198	3894
088	2224	3892
089	2250	3890
090	2275	3887
091	0,2301	0,3885
092	2327	3883
093	2353	3881
094	2378	3878
095	2404	3876
096	2430	3873
097	2456	3871
098	2482	3868
099	2508	3866
100	2533	3863

ÁREA 0,	ABS- CISSA x/σ	ORDE- NADA y
100	0,2533	0,3863
101	2559	3861
102	2585	3858
103	2611	3856
104	2637	3853
105	2663	3850
106	2689	3848
107	2715	3845
108	2741	3842
109	2767	3840
110	2793	3837
111	0,2819	0,3834
112	2845	3831
113	2871	3828
114	2898	3825
115	2924	3823
116	2950	3820
117	2976	3817
118	3002	3814
119	3029	3811
120	3055	3808
121	0,3081	0,3804
122	3107	3801
123	3134	3798
124	3160	3795
125	3186	3792
126	3213	3789
127	3239	3786
128	3266	3782
129	3292	3779
130	3319	3776
131	0,3345	0,3772
132	3372	3769
133	3398	3766
134	3425	3762
135	3451	3759
136	3478	3755
137	3505	3752
138	3531	3748
139	3558	3745
140	3585	3741
141	0,3611	0,3738
142	3638	3734
143	3665	3730
144	3692	3727
145	3719	3723
146	3745	3719
147	3772	3715
148	3799	3712
149	3826	3708
150	3853	3704

ÁREA	ABS- CISSA	ORDE- NADA
0,	x/σ	a y
150	0,3853	0,3704
151	3880	3700
152	3907	3696
153	3934	3692
154	3961	3688
155	3989	3684
156	4016	3680
157	4043	3676
158	4070	3672
159	4097	3668
160	4125	3664
161	0,4152	0,3660
162	4179	3656
163	4207	3652
164	4234	3647
165	4261	3643
166	4289	3639
167	4316	3635
168	4344	3630
169	4372	3626
170	4399	3621
171	0,4427	0,3617
172	4454	3613
173	4482	3608
174	4510	3604
175	4538	3599
176	4565	3595
177	4593	3590
178	4621	3585
179	4649	3581
180	4677	3576
181	0,4705	0,3571
182	4733	3567
183	4761	3562
184	4789	3557
185	4817	3552
186	4845	3548
187	4874	3543
188	4902	3538
189	4930	3533
190	4959	3528
191	0,4987	0,3523
192	5015	3518
193	5044	3513
194	5072	3508
195	5101	3503
196	5129	3498
197	5158	3493
198	5187	3487
199	5215	3482
200	5244	3477

ÁREA	ABS- CISSA	ORDE- NADA
0,	x/σ	y
200	0,5244	0,3477
201	5273	3472
202	5302	3466
203	5330	3461
204	5359	3456
205	5388	3450
206	5417	3445
207	5446	3440
208	5476	3434
209	5505	3429
210	5534	3423
211	0,5563	0,3417
212	5592	3412
213	5622	3406
214	5651	3401
215	5681	3395
216	5710	3389
217	5740	3384
218	5769	3378
219	5799	3372
220	5828	3366
221	0,5858	0,3360
222	5888	3355
223	5918	3349
224	5948	3343
225	5978	3337
226	6008	3331
227	6038	3325
228	6068	3319
229	6098	3313
230	6128	3306
231	0,6158	0,3300
232	6189	3294
233	6219	3288
234	6250	3282
235	6280	3275
236	6311	3269
237	6341	3263
238	6372	3256
239	6403	3250
240	6433	3244
241	0,6464	0,3237
242	6495	3231
243	6526	3224
244	6557	3218
245	6588	3211
246	6620	3204
247	6651	3198
248	6682	3191
249	6713	3184
250	6745	3178

ÁREA	ABS- CISSA	ORDE- NADA
0,	x/σ	y
250	0,6745	0,3178
251	6776	3171
252	6808	3164
253	6840	3157
254	6871	3151
255	6903	3144
256	6935	3137
257	6967	3130
258	6999	3123
259	7031	3116
260	7063	3109
261	0,7095	0,3102
262	7128	3095
263	7160	3087
264	7192	3080
265	7225	3073
266	7257	3066
267	7290	3058
268	7323	3051
269	7356	3044
270	7388	3036
271	0,7421	0,3029
272	7454	3022
273	7488	3014
274	7521	3007
275	7554	2999
276	7588	2992
277	7621	2984
278	7655	2976
279	7688	2969
280	7722	2961
281	0,7756	0,2953
282	7790	2945
283	7824	2938
284	7858	2930
285	7892	2922
286	7926	2914
287	7961	2906
288	7995	2898
289	8030	2890
290	8064	2882
291	0,8099	0,2874
292	8134	2866
293	8169	2858
294	8204	2849
295	8239	2841
296	8274	2833
297	8310	2825
298	8345	2816
299	8381	2808
300	8416	2800

ÁREA 0,	ABS- CISSA x/σ	ORDE- NADA y
300	0,8416	0,2800
301	8452	2791
302	8488	2783
303	8524	2774
304	8560	2766
305	8596	2757
306	8633	2748
307	8669	2740
308	8705	2731
309	8742	2722
310	8779	2714
311	0,8816	0,2705
312	8853	2696
313	8890	2687
314	8927	2678
315	8965	2669
316	9002	2660
317	9040	2651
318	9078	2642
319	9116	2633
320	9154	2624
321	0,9192	0,2615
322	9230	2606
323	9269	2596
324	9307	2587
325	9346	2578
326	9385	2568
327	9424	2559
328	9463	2550
329	9502	2540
330	9542	2531
331	0,9581	0,2521
332	9621	2511
333	9661	2502
334	9701	2492
335	9741	2482
336	9782	2473
337	9822	2463
338	9863	2453
339	9904	2443
340	9945	2433
341	0,9986	0,2423
342	1,0027	2413
343	0069	2403
344	0110	2393
345	0152	2383
346	0194	2373
347	0237	2362
348	0279	2352
349	0322	2342
350	0364	2332

ÁREA 0,	ABS- CISSA x/σ	ORDE- NADA y
350	1,0364	0,2332
351	0407	2321
352	0450	2311
353	0494	2300
354	0537	2290
355	0581	2279
356	0625	2269
357	0669	2258
358	0714	2247
359	0758	2237
360	0803	2226
361	1,0848	0,2215
362	0893	2204
363	0939	2193
364	0985	2182
365	1031	2171
366	1077	2160
367	1123	2149
368	1170	2138
369	1217	2127
370	1264	2115
371	1,1311	0,2104
372	1359	2093
373	1407	2081
374	1455	2070
375	1503	2059
376	1552	2047
377	1601	2035
378	1650	2024
379	1700	2012
380	1750	2000
381	1,1800	0,1989
382	1850	1977
383	1901	1965
384	1952	1953
385	2004	1941
386	2055	1929
387	2107	1917
388	2160	1905
389	2212	1893
390	2265	1880
391	1,2319	0,1868
392	2372	1856
393	2426	1843
394	2481	1831
395	2536	1818
396	2591	1806
397	2646	1793
398	2702	1781
399	2759	1768
400	2816	1755

ÁREA 0,	ABS- CISSA x/σ	ORDE- NADA y
400	1,2816	0,1755
401	2873	1742
402	2930	1729
403	2988	1716
404	3047	1703
405	3106	1690
406	3165	1677
407	3225	1664
408	3285	1651
409	3346	1637
410	3408	1624
411	1,3469	0,1610
412	3532	1597
413	3595	1583
414	3658	1570
415	3722	1556
416	3787	1542
417	3852	1529
418	3917	1515
419	3984	1501
420	4051	1487
421	1,4118	0,1473
422	4187	1458
423	4255	1444
424	4325	1430
425	4395	1416
426	4466	1401
427	4538	1387
428	4611	1372
429	4684	1357
430	4758	1343
431	1,4833	0,1328
432	4909	1313
433	4985	1298
434	5063	1283
435	5141	1268
436	5220	1253
437	5301	1237
438	5382	1222
439	5464	1207
440	5548	1191
441	1,5632	0,1176
442	5718	1160
443	5805	1144
444	5893	1128
445	5982	1112
446	6072	1096
447	6164	1080
448	6258	1064
449	6352	1048
450	6449	1031

ÁREA 0,	ABS- CISSA x/σ	ORDE- NADA y
450	1,6449	0,1031
451	6546	1015
452	6646	0998
453	6747	0982
454	6849	0965
455	6954	0948
456	7060	0931
457	7169	0914
458	1,7279	0,0897
459	7392	0879
460	7507	0862
461	7624	0844
462	7744	0826
463	7866	0809
464	7991	0791
465	8119	0773
466	8250	0755

ÁREA 0,	ABS- CISSA x/σ	ORDE- NADA y
466	1,8250	0,0755
467	8384	0736
468	8522	0718
469	8663	0699
470	8808	0680
471	8957	0662
472	9110	0643
473	9268	0623
474	1,9431	0,0604
475	9600	0584
476	9774	0565
477	9954	0545
478	2,0141	0525
479	0335	0505
480	0537	0484
481	0749	0464
482	0969	0443

ÁREA 0,	ABS- CISSA x/σ	ORDE- NADA y
482	2,0969	0,0443
483	1201	0422
484	1444	0400
485	1701	0379
486	1973	0357
487	2262	0335
488	2571	0312
489	2904	0290
490	2,3263	0,0267
491	3656	0243
492	4089	0219
493	4573	0195
494	5121	0170
495	5758	0145
496	6520	0118
497	7478	0091
498	8782	0063

ÁREA 0,	ABS- CISSA x/σ	ORDE- NADA y
4980	2,8781 6	0,0063 4
4981	8943 0	0060 5
4982	9112 4	0057 6
4983	9290 5	0054 7
4984	9478 4	0051 8
4985	9677 4	0048 8
4986	9888 8	0045 8
4987	3,0114 5	0042 8
4988	0356 7	0039 8
4989	0618 1	0036 7
4990	0902 3	0033 7
4991	3,1213 9	0,0030 6
4992	1559 1	0027 4
4993	1946 5	0024 3
4994	2388 8	0021 0
4995	2905 3	0017 8
4996	3527 9	0014 5

ÁREA 0,	ABSCISSA x/σ	ORDENADA y
4996	3,3527 9	0,0014 45
4997	4316 1	0,0011 06
4998	5400 8	0,0007 58
4999	7190 2	0,0003 96
4999 5	3,8905 92	0,0002 0607
4999 9	4,2648 91	0,0000 4478 7
4999 95	4,4171 73	0,0000 2312 4
4999 99	4,7534 24	0,0000 0494 83
4999 995	4,8916 39	0,0000 0254 09
4999 999	5,1993 38	0,0000 0053 80
4999 999 5	5,3267 24	0,0000 0027 5153
4999 999 9	5,6120 01	0,0000 0065 7804
4999 999 95	5,7307 3	0,0000 0002 9480
4999 999 99	5,9978 1	0,0000 0000 6157
5000 000 00	∞	0,0000 0000 0000

MÉTODOS DE ANÁLISE E AVALIAÇÃO*

William Brass
e
Ansley J. Coale

A necessidade de se escrever um extenso livro que no final apresente apenas, pouco mais que simples estimativas da fecundidade e mortalidade para diversas regiões da África Tropical é devida a natureza da informação estatística disponível sobre as populações africanas. A fonte convencional de informações sobre nascimentos e óbitos — o registro contínuo e completo destas ocorrências — simplesmente não é encontrado na África, e não há outra alternativa senão tentar estimar o número de ocorrências (a incidência de nascimentos por idade da mãe e de óbitos por idade e sexo do falecido) através de evidência menos direta. Existem dois tipos principais de evidências indiretas:

- 1.^a — a informação sobre a composição por idade e sexo da população obtida através de um recenseamento ou de uma pesquisa demográfica, algumas vezes suplementados por uma indicação da taxa de crescimento natural obtida através de duas enumerações da mesma população, e
- 2.^a — a informação sobre a fecundidade e a mortalidade obtida através das respostas a perguntas sobre o número de filhos tidos nascidos vivos por mulher, o número de filhos sobreviventes e

* Nota: O presente trabalho constitui o 3º Capítulo de "The Demography of Tropical Africa", de William Brass, copyright (c) 1968 da Princeton University Press, publicado sob a responsabilidade editorial do Office of Population Research e reproduzido com autorização da Princeton University Press, conforme contrato n.º 75-025. Não é permitida a reprodução sem prévia autorização da Princeton University Press.

sobre nascimentos e óbitos ocorridos no ano anterior ao recenseamento ou à pesquisa.

Mesmo quando, através desse segundo tipo de evidência se dispõe dos dados desejados, isto é, quando os levantamentos fornecem nascimentos e óbitos por idade ocorridos durante o ano anterior, esta informação não pode ser aceita integralmente sem uma apreciação crítica. Frequentemente, os entrevistados cometem erros em relação ao período de referência ao qual as informações devem ser referidas, e entre as populações da África Tropical, em sua maior parte analfabetas, tem sido usualmente impossível obter uma informação precisa da idade. Assim, o principal esforço de pesquisa deste livro é a criação, adaptação e aplicação de métodos para obtenção das melhores estimativas possíveis de fecundidade e mortalidade através de dados indiretos e/ou inexatos. Os processos empregados diferem de população para população em virtude da natureza dos recenseamentos e pesquisas realizados nas diversas regiões. Os métodos de análise particularmente empregados em cada região estão descritos nos capítulos que fornecem os resultados detalhados da região. Entretanto, foram empregados nesse projeto três métodos principais de análise e avaliação com frequência suficiente para justificar este capítulo em separado para sua descrição.

Os três métodos largamente utilizados são:

- 1 — a avaliação da fecundidade a partir da análise combinada de dados sobre nascimentos ocorridos no ano anterior ao levantamento e sobre o total de filhos tidos nascidos vivos por mulher em idades férteis;
- 2 — a avaliação da mortalidade infantil e nas primeiras idades através de informação sobre filhos tidos nascidos vivos e sobre os filhos sobreviventes; e
- 3 — a avaliação da fecundidade, da mortalidade e das distribuições ajustadas de idade a partir de um conjunto variado de informações, através do uso de tábuas de vida modelo e populações estáveis.

Os dois primeiros desses métodos são devidos a William Brass. Os métodos de Brass descritos neste capítulo (e em maiores detalhes no capítulo sobre as regiões analisadas por Brass) foram desenvolvidos no Departamento de Pesquisa Demográfica (Office of Population Research) para aplicação específica às populações africanas, apesar de representarem o ponto culminante de um trabalho anterior¹.

¹ BRASS, William The Derivation of Fertility and Reproduction Rates from Restricted Data on Reproductive Histories *Population Studies*, 7 (2) nov. 1953

———. The Estimation of Fertility Rates from Ratios of Total to First Births *Population Studies*, 8 (1) jul. 1954

———. *The Estimation of Total Fertility Rates from Data for Primitive Communities*. World Population Conference, Roma, 1954

———. The Graduation of Fertility Distributions by Polynomial Functions. *Population Studies*, 14 (2): nov. 1960

As tábuas de vida modelo e populações estáveis construídas por Ansley J. Coale e Paul Demeny são o resultado de um trabalho preliminar de muitos anos feito por Coale, com a colaboração de Erna Harm. Uma grande quantidade de cálculos eletrônicos de tábuas modelo, realizados por Paul Demeny, foi executada para possibilitar a aplicação imediata das técnicas demográficas da tábua de vida modelo aos dados africanos. Por quase dois anos esta extensa coleção de tábuas modelo foi dedicada quase que somente a esse projeto de pesquisa. Existiam apenas duas cópias fotostáticas das tábuas. Estas tábuas (perto de cinco mil) foram agora publicadas independentemente, por isso sua discussão ao final desse capítulo não é muito detalhada².

Um comentário final deve ser feito antes de começarmos uma descrição dos métodos de Brass para analisar a fecundidade e a mortalidade. Um procedimento mantido obrigatoriamente na análise dos dados relativos às diversas regiões foi comparar, sempre que possível, as estimativas obtidas através de independentes tipos de evidência, e nossa confiança nas estimativas depende do grau de concordância entre elas. Assim, a característica mais confirmadora das estimativas derivadas dos nascimentos comunicados e do número informado de filhos tidos nascidos vivos é que elas são quase sempre consistentes com as estimativas derivadas da composição etária, tanto a nível de populações inteiras como em relação a diferentes sub-grupos.

ESTIMATIVAS DE FECUNDIDADE A PARTIR DE INFORMAÇÕES SOBRE NASCIMENTOS OCORRIDOS E FILHOS TIDOS NASCIDOS VIVOS

Em muitos recenseamentos e pesquisas na África e em outras áreas subdesenvolvidas, foram coletados dados de fecundidade de dois tipos. Os dois tipos de informações, ambos registrados segundo a idade da mãe, são os nascimentos ocorridos em um período atual e o número de filhos tidos nascidos vivos por mulher. Nesta seção discutiremos os métodos de análise de tais registros para a obtenção de estimativas plausíveis do nível de fecundidade mesmo quando ambos os conjuntos de dados estiverem sujeitos a erros.

As informações sobre nascimentos atuais são obtidas interrogando as mulheres sobre filhos nascidos vivos num curto período precedente ao recenseamento. Com o objetivo de simplificar, suporemos que o período mais usual é o de um ano. Os métodos podem ser estendidos para outros intervalos de tempo sem dificuldade. As taxas de fecundidade específicas por idade podem ser derivadas a partir da divisão do número de nascidos vivos de mães de certa idade pelo total correspondente de mulheres na população na idade considerada. As avalia-

² COALE, Ansley J. & DEMENY, Paul. *Regional Model Life Tables and Stable Populations*. Princeton University Press, 1966

ções baseadas nestas taxas nos referiremos como “atuais”, e os índices de fecundidade obtidos do número médio de filhos tidos nascidos vivos por mulher de cada idade até a data do recenseamento serão chamados de “retrospectivos”.

Os dois tipos de medidas podem ser usados para detectar e corrigir erros nos dados em virtude da relação lógica entre eles. A medida em que uma coorte de mulheres atravessa a vida, o número médio de filhos tidos nascidos vivos até cada idade exata é igual ao total acumulado das taxas de fecundidade específica por idade até aquela idade. Se é possível supor que as mulheres que morrem têm a mesma fecundidade que aquelas que sobrevivem e se a fecundidade da população for constante no tempo, as taxas específicas por idade de cada coorte serão idênticas as taxas “atuais” e esta relação se mantém para todas as idades das mulheres. Estes resultados teóricos se aplicarão aproximadamente a uma população real se as taxas de fecundidade não estiveram sujeitas a uma marcada tendência no tempo.

O número médio de filhos tidos nascidos vivos por mulher em cada idade, calculado de dados atuais, pode ser comparado com os dados retrospectivos correspondentes. Se os dois conjuntos de índices concordam em todas as idades, é forte a evidência de que eles estejam certos. Aplicações desse procedimento a registros de muitas populações africanas revelaram discrepâncias sistemáticas. Estas discrepâncias têm uma forma coerente com o que se sabe sobre erros em relatórios de levantamento de fatos vitais. Uma taxa de fecundidade baseada em respostas a uma pergunta sobre nascimentos ocorridos durante o ano precedente ao recenseamento ou pesquisa está com frequência tão claramente errada a ponto de não poder ser utilizada, na maioria das vezes, porque o nível é muito baixo, apesar de que algumas vezes aparece demasiado aumentado. Por outro lado, o número médio de filhos tidos nascidos vivos informado, também, aumenta muito lentamente com a idade da mulher, especialmente acima de 30 ou 35 anos, e declina em idades além do período reprodutivo. Há uma probabilidade muito forte de que isso se deva à omissão de uma crescente proporção de filhos tidos, à medida em que as mulheres vão envelhecendo.

Cada uma dessas formas de informação pode ser nitidamente deficiente mas a natureza dos erros é diferente. Por isso é possível desenvolver uma técnica para avaliação da taxa global de fecundidade a partir das partes desses dois conjuntos de medida que provavelmente são de maior confiança. A técnica depende das seguintes hipóteses:

- 1.^a — a mais importante fonte de erro do número informado de nascimentos ocorridos no ano anterior ao recenseamento é a imprecisão no *período de referência*. As entrevistadas podem informar fatos ocorridos, em média, nos últimos oito meses ou (com uma cultura diferente ou uma colocação diferente das perguntas) nos últimos quinze meses. Dificuldades semelhan-

tes quanto a intervalos de tempo nos quais foram feitas compras ocorreram em levantamentos retrospectivos de consumo em países desenvolvidos, como os Estados Unidos. Todavia pode ser considerado válido o mesmo erro médio na duração do período de referência para cada idade materna, particularmente em levantamentos de populações analfabetas, onde o entrevistador tenta controlar a coerência das respostas. Frequentemente ele verá as crianças cujos nascimentos foram declarados e julgará a validade das informações pela estimativa que ele próprio faz sobre a idade delas. Qualquer predisposição em seu julgamento, provavelmente não será influenciada pela idade da mãe (por exemplo, a tendência de se considerar a duração do período de amamentação igual a um ano, apesar deste ser muito mais longo em sociedades africanas). A distribuição por idade das taxas de fecundidade específicas atuais isto é, as relações entre as taxas nas diferentes idades, pode plausivamente ser aceita como aproximadamente correta, apesar do mesmo não acontecer com o nível.

2.^a — o número de filhos tidos nascidos vivos é informado com boa precisão por mulheres mais jovens. Os fatos que se indagam às mulheres mais jovens ocorreram recentemente: o número de filhos tidos normalmente não são mais que dois ou três, de modo que não surgem dificuldades de contagem de grandes números numa sociedade que não enumera os filhos vivos (uma proporção maior de filhos tidos nascidos vivos por mães jovens sobreviverá até a época do recenseamento) frequentemente estarão presentes às entrevistas, e poucos deles serão omitidos por terem crescido e deixado a família.

Com base nessas proposições foi desenvolvida uma técnica para estimar a fecundidade. Aceita-se o padrão etário das taxas de fecundidade obtido através das informações de nascimentos atuais, mas o nível da fecundidade é estimado através do número médio de filhos tidos nascidos vivos informado pelas mulheres mais jovens. Na aplicação do processo o número de filhos tidos nascidos vivos obtidos pelas taxas atuais é comparado em cada idade com o valor “retrospectivo” observado. As razões entre as medidas atuais acumuladas e retrospectivas para as mulheres jovens dão um fator que é aplicado às taxas específicas de fecundidade atuais em todas as idades para ajustar a fecundidade ao nível requerido. Também é possível derivar as taxas de fecundidade específicas por idade dos dados de filhos tidos nascidos vivos e compará-las com as correspondentes taxas atuais, mas se prefere o processo de converter a fecundidade atual ao número médio de filhos tidos nascidos vivos porque a acumulação tende a reduzir os efeitos de pequenos erros e irregularidades nas observações.

COMPARAÇÃO ENTRE PARTURIÇÃO MÉDIA (NÚMERO MÉDIO DE FILHOS TIDOS NASCIDOS VIVOS) E O VALOR MÉDIO DA FECUNDIDADE ATUAL ACUMULADA

Se fossem disponíveis dados consistentes sobre nascimentos pela idade individual da mãe, os cálculos para a execução das comparações seriam diretos. Todavia, quase sempre as tabulações são apresentadas em grupos quinqüenais de idade das mulheres, normalmente 15 - 20, 20 - 25, etc., porque os registros foram coletados dessa forma, ou porque o agrupamento foi usado para reduzir os efeitos de erros na declaração de idade, devidos a atração por certos dígitos, etc., e flutuações aleatórias de pequenos números. Também deve ser notado que a experiência atual de fecundidade corresponde a idades em torno de seis meses menores que as tabuladas. Esse deslocamento surge porque os nascimentos informados ocorreram durante o ano anterior; em média, as mães eram seis meses mais jovens ao dar a luz do que na época do recenseamento.

A taxa de fecundidade para um determinado grupo de idade é um valor médio anual. Essas taxas médias de fecundidade para os sete grupos quinqüenais que cobrem efetivamente o período reprodutivo serão designadas por f_i (onde $i = 1, 2, \dots, 7$). A multiplicação pelo número de anos do intervalo dá o número médio de nascimentos que ocorreriam entre as mulheres durante o intervalo. A acumulação dos valores f_i (multiplicados por 5) desde o grupo mais jovem fornece o número médio de filhos tidos nascidos vivos nos limites superiores dos intervalos. Assim, se os grupos etários tabulados são 15 - 20, 20 - 25 anos, etc., os f_i , acumulados dão o número médio de filhos tidos nascidos vivos em idades exatas de 20, 25 anos, etc., se tiver sido feita uma correção para o deslocamento de meio ano, mas, na situação mais comum onde isso não foi feito em idades exatas de 19 anos e meio, 24 anos e meio, etc.

As informações retrospectivas, do número médio de filhos tidos nascidos vivos por mulher, para os mesmos grupos quinqüenais, serão indicados por P_1, P_2 etc. (ou P_i , onde $i = 1, 2, \dots, 7$). A letra P foi escolhida, porque os demógrafos se referem ao número de filhos que uma mulher teve como sua "parturição". Assim, P_1 é a parturição média das mulheres de quinze a vinte anos, P_2 das mulheres de vinte a vinte e cinco, etc. Estas não são medidas em idades exatas, mas médias correspondentes ao grupo etário. Com a fecundidade constante, o número médio de filhos tidos nascidos vivos deve crescer com a idade da mulher. O valor P para um determinado grupo etário estará, portanto, compreendido entre o número médio de filhos tidos nascidos vivos nas idades exatas que correspondem exatamente aos limites inferior e superior do intervalo. É portanto, necessário criar um processo para calcular, a partir dos valores de f_i , os correspondentes valores médios de fecundidade acumulada P_i para cada um dos grupos etários. Estes valores serão chamados de F_i , sendo F_1 o valor médio

estimado de fecundidade acumulada para mulheres de quinze a vinte anos, F_2 para mulheres de vinte a vinte e cinco, etc. (ou F_i onde $i = 1, 2, \dots, 7$).

Se a fecundidade específica por idade é constante em um intervalo, a fecundidade acumulada aumentará linearmente com a idade e a média para o intervalo estará a meio caminho entre os valores acumulados de f nos limites inferiores e superiores. A fecundidade média acumulada estimada seria, então, obtida somando-se duas vezes e meia a fecundidade para o intervalo à fecundidade acumulada até o limite mais baixo de idade. Se houver um deslocamento de meio ano nos grupos etários para as taxas atuais, este aspecto pode ser considerado, levando-se em conta o número de nascimentos que teriam ocorrido em meio ano. Assim, para o grupo etário da ordem de i , onde ϕ_i é o total acumulado de $5f_1 + 5f_2$, etc. até $5f_{i-1}$, a estimativa da fecundidade acumulada média no intervalo seria $\phi_i + 3,0f_i$.

O multiplicador de f_i , nesta fórmula, é válido para os grupos etários, onde as taxas de fecundidade específicas variam pouco com a idade, mas apresentam erros muito maiores nos extremos do período reprodutivo, onde a fecundidade está mudando rapidamente. A imprecisão nas idades superiores não é importante para nosso propósito, já que f_i é muito menor do que o componente ϕ_i , que se conhece exatamente. A estimativa para o primeiro grupo etário estará quase sempre grandemente impregnada de erro, e freqüentemente há uma apreciável predisposição a erro no valor obtido para o segundo grupo. Todavia, esta simples fórmula dá valores médios estimados da fecundidade acumulada que podem ser satisfatoriamente comparados com as parturições médias observadas para se obter pelo menos uma impressão aproximada dos erros.

Foi desenvolvido um processo de avaliação mais complicado, que todavia pode ser aplicado com muita facilidade. A base para o processo é uma distribuição modelo de fecundidade específica por idade em que a forma está fixada, mas a localização pode variar; isto é, a curva que descreve a forma pode se mover ao longo do eixo das idades (transformação linear, de modo que os limites superior e inferior do período reprodutivo e a média da distribuição sejam deslocados em mesma quantidade). Para localizações uniformemente intercaladas do modelo foram calculados os seguintes parâmetros:

- 1) a idade média da distribuição de fecundidade;
- 2) a razão entre f_1 e f_2 , isto é, entre a taxa de fecundidade específica no primeiro grupo etário e a do segundo;
- 3) para cada grupo etário, o fator k_i pelo qual o valor de f_i teve que ser multiplicado a fim de que a fórmula $\phi_i + k_i f_i$ desse exatamente F_i :
 - a) quando não tivesse havido nenhum deslocamento de idades nas taxas atuais de fecundidade informadas;

b) quando houvesse um deslocamento de meio ano nas taxas atuais.

Não se introduziu modificações nos cálculos para corrigir os possíveis efeitos da variação no número de mulheres em cada idade, no grupo etário, porque se pode demonstrar que estas alterações seriam muito pequenas. A forma matemática do modelo e dos métodos de cálculo está descrita no Apêndice A deste estudo. Também ali se encontram os valores calculados de k_i que se aplicam quando não há nenhum deslocamento de idades das taxas atuais. Os valores para o caso mais comum de um deslocamento de meio ano estão apresentados na Tabela 3.1.

TABELA 3.1

FATORES MULTIPLICADORES PARA O CÁLCULO DO VALOR MÉDIO, POR GRUPOS ETÁRIOS QUINQUÊNAIS, DA FECUNDIDADE ACUMULADA (F_i) DE ACORDO COM A FÓRMULA $F_i = \phi_i + k_i f_i$ (QUANDO f_i SE REFERE AOS GRUPOS DE IDADES 14,5 A 19,5; 19,5 A 24,5 ETC.)

LIMITES EXATOS DO INTERVALO ETÁRIO	FATORES							
15 — 20	1,120	1,310	1,615	1,950	2,305	2,640	2,925	3,170
20 — 25	2,555	2,690	2,780	2,840	2,890	2,925	2,960	2,985
25 — 30	2,925	2,960	2,985	3,010	3,035	3,055	3,075	3,095
30 — 35	3,055	3,075	3,095	3,120	3,140	3,165	3,190	3,215
35 — 40	3,165	3,190	3,215	3,245	3,285	3,325	3,375	3,435
40 — 45	3,325	3,375	3,435	3,510	3,610	3,740	3,915	4,150
45 — 50	3,640	3,895	4,150	4,395	4,630	4,840	4,985	5,000
f_1/f_2	0,036	0,113	0,213	0,330	0,460	0,605	0,764	0,939
\bar{m} (em anos)	31,7	30,7	29,7	28,7	27,7	26,7	25,7	24,7

Os fatores k , na tabela, estão próximos a 3,0 para os terceiro e quarto grupos etários, em todas as possíveis localizações da distribuição modelo, e a divergência para o segundo grupo etário, também, é pequena quando a vida reprodutiva começa cedo. Estes resultados são coerentes com os comentários sobre a precisão da aproximação linear para a determinação de F a partir das taxas de fecundidade atuais. O procedimento para o uso dos valores da tábua é o seguinte:

— a partir das taxas de fecundidade específicas por idade, derivadas de nascimentos informados no ano anterior ao levantamento, calculam-se a idade média (\bar{m}) da distribuição, a razão f_1/f_2 , e os

valores de ϕ_i , fecundidade acumulada até o limite inferior do grupo etário de ordem i ($\phi_i = 5[f_1 + f_2 + \dots + f_{i-1}]$). Calculam-se fatores k pela interpolação linear entre as colunas da Tabela 3.1. A interpolação é guiada pela f_1/f_2 observada, para os três primeiros grupos etários e pela \bar{m} observada para os grupos restantes. Finalmente, obtém-se F_i pela expressão $\phi_i + k_i f_i$.

Com efeito, através desse procedimento, a distribuição da fecundidade específica, por idade, observada é ajustada para os primeiros grupos etários pelo modelo, com a mesma razão f_1/f_2 , e os demais grupos pelo modelo, com a mesma idade média. Os fatores k determinados pela tabela relativa a distribuições modelo de fecundidade são tomados como as estimativas para fatores referentes aos dados observados. A seleção da fecundidade modelo através de uma razão f_1/f_2 , assegura que a distribuição observada e a distribuição modelo se ajustam bem nas idades mais jovens, e a igualdade entre as idades médias leva a um resultado similar sobre a amplitude média do período reprodutivo. Pelas razões discutidas anteriormente, a precisão nas idades mais elevadas é menos importante. A mudança de um método de ajuste para outro pode ser feita facilmente entre o terceiro e quarto grupos etários porque ali as taxas de fecundidade específicas por idade são praticamente constantes, e as variações em f_1/f_2 em \bar{m} têm pouco efeito sobre os fatores k .

A tabela foi construída supondo-se que as informações são apresentadas segundo os grupos etários quinquenais mais usados, 15 - 20, 20 - 25, etc., e que foram informados os nascimentos do ano anterior ao recenseamento. Porém é fácil modificar o método para outras condições. O único índice na tabela que depende da idade, tomada como o limite inferior do período reprodutivo é a média. Se este for reduzido em um ano para todas as colunas, a tabela se relacionará a registros de grupos etários de 14-19 anos, 19-24 anos, etc. Assim os fatores podem ser usados para qualquer conjunto de dados referentes aos sete grupos quinquenais, se as médias forem trocadas apropriadamente.

Se o período de registro para os nascimentos atuais não é de um ano, mas algum outro intervalo, de menor amplitude, o conjunto apropriado de fatores k pode ser calculado por interpolação da Tabela 3.1 e da Tabela 3.A.2, do Apêndice que se aplica a um deslocamento igual a zero. As razões f_1/f_2 não sofrem alterações, mas o deslocamento deve ser subtraído das idades médias na Tabela 3.A.2 do Apêndice para atribuição dos valores corretos.

Na técnica descrita, os valores de f acumulados para os limites inferiores dos grupos etários são aceitos como corretos e apenas se estima o número médio adicional de nascimentos para mulheres que estão no grupo etário. Métodos para derivar o número médio de filhos tidos nascidos vivos a partir das taxas de fecundidade específicas por idade para grupos quinquenais e vice-versa, nos quais se ajusta toda a distribuição das observações através de modelos de fecundidade foram

estudados por Brass³. Esses processos mais complicados são particularmente apropriados para a análise de observações oriundas de pequenos levantamentos nos quais os erros de amostragem são maiores.

A precisão do procedimento para estimar F_1

A precisão dos valores estimados de F_1 pode ser testada aplicando-se o método de estimação para dados confiáveis, consistindo de taxas de fecundidade específicas por idades individuais e por intervalos quinquenais de idade. Para reunir dados desse tipo, o número médio de filhos tidos nascidos vivos por mulher, segundo grupos etários quinquenais, foi calculado a partir das taxas de fecundidade específicas por idades individuais de quatro populações. Informações pormenorizadas sobre os dados encontram-se no trabalho de Brass a que nos referimos anteriormente⁴.

Os valores exatos construídos a partir de taxas específicas por idades individuais estão comparados na Tabela 3.2 com aqueles estimados pelo processo do fator k . Admitindo-se que os valores construídos, correspondam à parturição média P_i que poderia ser informada num levantamento preciso, e os valores estimados de F_i , calculados por nosso processo, podem ser comparados, para cada grupo etário, com estes P_i construídos.

TABELA 3.2

COMPARAÇÃO ENTRE A FECUNDIDADE MÉDIA ACUMULADA ESTIMADA (F) E O NÚMERO MÉDIO DE FILHOS TIDOS NASCIDOS VIVOS, POR MULHER, CALCULADO (P)

LIMITES EXATOS DO INTERVALO ETÁRIO	UCRÂNIA 1926/27			ESLOVÊNIA 1948/52			INGLATERRA E PAÍS DE GALES 1951			AUSTRÁLIA 1932/34		
	F	P	P/F	F	P	P/F	F	P	P/F	F	P	P/F
15 — 20	0,040	0,038	0,095	0,019	0,024	1,26	0,018	0,024	1,33	0,029	0,033	1,14
20 — 25	0,747	0,778	1,04	0,398	0,399	1,00	0,385	0,389	1,01	0,353	0,352	1,00
25 — 30	2,045	2,057	1,01	1,157	1,166	1,01	1,067	1,087	1,02	0,925	0,929	1,00
30 — 35	3,280	3,277	1,00	1,916	1,919	1,00	1,643	1,650	1,00	1,481	1,479	1,00
35 — 40	4,244	4,238	1,00	2,491	2,488	1,00	1,980	1,981	1,00	1,876	1,872	1,00
40 — 45	4,845	4,835	1,00	2,822	2,826	1,00	2,122	2,123	1,00	2,086	2,086	1,00
45 — 50	5,104	5,090	1,00	2,920	2,920	1,00	2,151	2,151	1,00	2,144	2,144	1,00
50 e +	5,135	5,135	1,00	2,920	2,926	1,00	2,152	2,152	1,00	2,147	2,147	1,00

Todas as razões P/F , exceto as do primeiro grupo etário, estão próximas a 1,00, o valor da razão P/F correspondente a concordância perfeita. A razão P/F para o primeiro grupo é muito sensível à forma exata da distribuição de fecundidade específica nos primeiros anos do período reprodutivo. Poder-se-ia suspeitar que se utilizássemos popu-

^{3, 4} BRASS, William. The Graduation of Fertility Distributions by Polynomial Functions, op. cit. p. 46

lações onde a vida reprodutiva começa cedo, como na maioria das comunidades africanas, a razão P/F para o grupo etário mais jovem seria mais consistente. Entretanto, não há registros satisfatórios de fecundidade específica por idades individuais, em populações africanas que permitam o exame dessa hipótese. Ao se derivar um fator de correção das taxas de fecundidade atuais informadas, pouco peso pode ser dado à razão P/F para o grupo etário mais jovem. Há outros motivos que dificultam uma estimativa correta da razão P/F para o primeiro grupo etário: erro de amostragem das medições para o primeiro grupo, em função do pequeno número de nascimentos, pode ser alto, e, também, este grupo é particularmente vulnerável aos efeitos da declaração inexata da idade.

O nível da fecundidade é melhor avaliado através do número médio de filhos tidos nascidos vivos informado pelo grupo mais jovem de mulheres, dado que as informações têm aceitável precisão, quase isentas do efeito das falhas de memória. Como regra, portanto, a razão P/F para o grupo etário de 20-25 anos será usada para corrigir o nível de taxas de fecundidade atuais informadas, a menos que haja evidência de que a razão para o grupo 20-25 esteja distorcida ou seja inconsistente com a tendência dos valores obtidos para os grupos de idade posteriores. A precisão do método não é tal que se justifique a correção de pequenas percentagens; correções maiores são necessárias em muitas das análises dos dados censitários. Conjuntos de razões P/F dos dados de fecundidade, para algumas das comunidades africanas estudadas, estão apresentados na Tabela 3.3, para ilustração.

TABELA 3.3

EXEMPLOS DE RAZÕES P/F COMPUTADAS A PARTIR DE DADOS CENSITÁRIOS

LIMITES EXATOS DOS INTERVALOS ETÁRIOS	TERRITÓRIOS AFRICANOS			
	Daomé	Guiné	Fouta — Toio (Vale do Senegal)	Buganda (Uganda)
15 — 20	0,89	0,78	0,87	1,36
20 — 25	0,93	0,80	1,04	1,13
25 — 30	0,89	0,75	0,95	0,89
30 — 35	0,89	0,75	0,89	0,79
35 — 40	0,84	0,72	0,86	0,74
40 — 45	0,86	0,74	0,86	0,71
45 — 50	0,84	0,76	0,86	0,72
50 — 60	0,83	0,76	0,84	—

Cálculo ilustrativo da fecundidade a partir de dados hipotéticos

Antes de considerarmos as características e os problemas do método em aplicações a dados africanos provenientes de recenseamentos, que podem ser muito deficientes, serão realizados cálculos para um conjunto hipotético de observações. O exemplo para verificação da técnica foi composto a partir dos dados de fecundidade da Ucrânia, período 1926-27, já usados na Tabela 3.2, incluindo-se as seguintes alterações:

- a) a fecundidade específica por idade foi ajustada para o que teria sido a taxa observada se o registro fosse de nascimentos ocorridos durante o ano anterior a um recenseamento, e se 25 por cento de nascimentos "atuais" não fossem informados para todas as idades das mães;
- b) uma proporção das crianças, variando linearmente de zero, no grupo de 15-20 anos, até 24 por cento, para o grupo 45-50 anos, foi omitida dos valores retrospectivos referentes aos filhos tidos nascidos vivos.

A Tabela 3.4 mostra como são feitos os cálculos. As taxas de fecundidade para grupos etários quinquenais, calculadas a partir dos nascimentos informados no ano anterior ao recenseamento da população hipotética, são multiplicadas por cinco e acumuladas para dar a coluna ϕ . O número médio de filhos tidos nascidos vivos por mulher, registrado para esta população, está na coluna encabeçada por P_i ; f_1/f_2 e \bar{m} , calculados a partir da tabela de fecundidade, encontram-se abaixo da tabela.

TABELA 3.4

EXEMPLO ILUSTRATIVO DO MÉTODO QUE UTILIZA A RAZÃO P/F PARA O CÁLCULO DO NÍVEL DE FECUNDIDADE

LIMITES EXATOS DOS INTERVALOS ETÁRIOS	i	f_i^*	ϕ_i	k_i	$\phi_i + k_i f_i = P_i$	P_i	P_i/F_i	f_i AJUSTADO
15 — 20	1	0,021		1,345	0,028	0,038	1,36	0,028
20 — 25	2	0,170	0,105	2,695	0,563	0,747	1,33	0,226
25 — 30	3	0,195	0,955	2,865	1,533	1,892	1,23	0,260
30 — 35	4	0,172	1,930	3,085	2,459	2,834	1,17	0,228
35 — 40	5	0,124	2,788	3,200	3,184	3,560	1,12	0,164
40 — 45	6	0,067	3,406	3,405	3,634	3,868	1,06	0,089
45 — 50	7	0,022	3,741	4,020	3,829	3,868	1,01	0,029
Fecundidade total		3,851	3,851	—	3,851	3,868	1,00	5,122

$f_1/f_2 = 0,124$; $\bar{m} = 30,22$

* Para intervalos etários meio ano menores que os apresentados.

A seguir, consulta-se a Tabela 3.1. A razão observada $f_1/f_2 = 0,124$ está entre a segunda e a terceira coluna de fatores, e corresponde a uma proporção $0,011/0,100 = 0,11$ do intervalo entre as colunas, a par-

tir da primeira destas. A interpolação linear entre os fatores k , usando-se esta proporção dá os valores para os três primeiros grupos etários da Tabela 3.4. Similarmente, a média 30,22 corresponde a 48% do intervalo entre a segunda e a terceira coluna. Os valores k do quarto ao sétimo grupo etário são obtidos por interpolação guiadas por \bar{m} . Os valores de F_i são calculados a partir da expressão $\phi_i + k_i f_i$ quando se substituem os fatores k estimados.

As razões P/F diminuem uniformemente com a idade, apresentando o valor de 1,33 para o grupo etário de 20-25. O nível de fecundidade indicado pelas informações retrospectivas de filhos tidos nascidos vivos nesse grupo etário é, portanto, 33% maior do que o indicado pelas taxas atuais. A multiplicação dos valores de f registrados e da fecundidade total por 1,33 leva às medidas corrigidas que se encontram na última coluna.

A estimativa da fecundidade total é 5,122, comparada com o nível verdadeiro de 5,135 na população hipotética. A precisão obtida pelo método neste exemplo é bastante melhor do que a que deveríamos esperar. As omissões supostas de 4% na informação de filhos tidos nascidos vivos dada pelas mulheres com 20-25 anos de idade foram suficientemente compensadas por erros na estimativa de k_2 . Digno de nota, é que a taxa de fecundidade total derivada é mais ou menos um terço maior do que os valores obtidos a partir das taxas atuais observadas, como do número médio de filhos tidos nascidos vivos por mulheres de fecundidade completa. A quase igualdade entre esses dois valores, que ocorre porque a proporção de nascimentos não informados no ano anterior foi mais ou menos igual às omissões de filhos tidos por mulheres mais velhas não é evidência de exatidão. Uma coincidência, em erro desse tipo, ocorreu em alguns dados censitários referentes a populações africanas.

Já que a fecundidade total estimada está próxima a da população hipotética, as medidas ajustadas de f também estão em estreita concordância com as consideradas. Para comparações com outras populações, é uma perda de tempo ter medidas de f , e as correspondentes taxas específicas, para grupos etários que estejam deslocados seis meses do sistema padrão. Um processo elaborado para se fazer essa transformação não se justifica porque, nos registros analisados, as taxas específicas por idade estão erradas em virtude dos erros de amostragem, de declaração de idade, etc. As correções simples são feitas supondo-se que nos limites a taxa de fecundidade específica é igual à média dos valores nos grupos etários adjacentes. Se δf_i é a quantidade que deve ser adicionada a f_i para os deslocamentos de meio ano, os ajustes são:

$$\delta f_1 = (f_1 + f_2)/20$$

$$\delta f_6 = (f_6 - f_4)/20$$

$$\delta f_2 = (f_3 - f_1)/20$$

$$\delta f_7 = (f_7 - f_5)/20$$

$$\delta f_3 = (f_4 - f_2)/20$$

$$\delta f_7 = (-f_6 - f_7)/20$$

$$\delta f_4 = (f_5 - f_3)/20$$

Estas correções foram feitas aos valores de f estimados para a população hipotética da Tabela 3.4. Os valores resultantes estão comparados no Quadro abaixo com as taxas de fecundidade específicas por idade observadas para os grupos etários padrões, calculadas a partir das taxas específicas originais para idades individuais.

GRUPOS DE IDADE	TAXAS DE FECUNDIDADE	
	Observada	Estimada
15 — 20	0,043	0,041
20 — 25	0,238	0,238
25 — 30	0,259	0,259
30 — 35	0,223	0,223
35 — 40	0,157	0,157
40 — 45	0,082	0,082
45 — 50	0,024	0,023
Fecundidade total	5,135	5,122

Efeitos dos erros nos dados ou nas hipóteses

O processo para estimar a fecundidade dá bons resultados se as hipóteses sobre as relações entre taxas por coorte e taxas atuais e as suposições sobre a natureza dos erros de informação se mantiverem. Também é “vigoroso”, isto é, insensível aos tipos de variação provenientes das hipóteses, que devem ser esperados nas informações relativas a comunidades subdesenvolvidas. O fator de correção para o nível de fecundidade é derivado do número médio de filhos tidos nascidos vivos por mulheres de 20-25 anos de idade. Neste grupo etário, morreu, apenas, uma pequena proporção das mulheres que entraram no período reprodutivo, e são pequenos os possíveis efeitos de uma fecundidade diferencial das mulheres falecidas. Já que a grande maioria dos nascimentos para mulheres de 20-25 anos de idade terá ocorrido poucos anos antes do recenseamento, a estimativa do nível de fecundidade se aplicará a uma experiência muito recente, mesmo quando houver uma tendência nas taxas. As razões P/F para as mulheres mais velhas podem estar distorcidas pela fecundidade diferencial das sobreviventes e as tendências no tempo das taxas de natalidade, assim como por omissões de filhos tidos por problemas referentes a falhas de memória. Apesar de se complicar dessa maneira a interpretação precisa do padrão das razões P/F , a estimativa do nível de fecundidade está pouco afetada.

Há alguns erros de informação que não se levam em conta na análise. Nenhuma correção é feita para quaisquer nascimentos que sejam omitidos igualmente por mulheres de todas as idades, tanto em registros atuais como retrospectivos. Esses nascimentos podem ser dos filhos que morrem muito cedo, nos primeiros dias ou semanas de vida.

Também é possível que as omissões ou erros relativos ao período de referência na informação de nascimentos atuais não sejam completamente independentes da idade da mãe; uma plausível sugestão é que a correção necessária deveria aumentar com a idade. Essas duas categorias de erro levariam a subestimativas de fecundidade. Algumas operações de controle quanto a essas duas possibilidades podem ser feitas pelo exame dos dados censitários sobre a mortalidade e as distribuições etárias, como será explicado mais adiante.

Uma das principais dificuldades em recenseamentos ou pesquisa em populações africanas é a determinação da idade. O problema não é simplesmente de imprecisão, preferência por dígitos etc., por parte, do entrevistado, mas comumente de ignorância completa. Frequentemente o entrevistador tem que fazer uma estimativa pela aparência, lembranças de fatos ocorridos, *status* na família, etc. O exame das distribuições etárias obtidas em recenseamentos africanos revela distorções grandes e, como está demonstrado no Capítulo 2 (da publicação original), sistemáticas. Portanto, devemos considerar cuidadosamente a sensibilidade do método para estimar a fecundidade, diante das alterações na declaração da idade. De fato, uma grande vantagem para aplicação do método é que as determinações da idade, incluindo os erros, são as mesmas para os dois conjuntos de observações comparados, a saber, o atual e o retrospectivo. Se os registros de fecundidade viessem de duas fontes distintas, por exemplo, registro civil e recenseamento, com diferentes imprecisões na idade das pessoas, as conclusões a serem tiradas das comparações seriam muito mais duvidosas.

Uma tendenciosidade constante em todas as idades femininas afetaria apenas a estimação da fecundidade total (apesar de que estariam erradas as taxas de fecundidade específica por idade) porque o fator de correção para o nível seria calculado baseado em um grupo etário ligeiramente diferente. Por outro lado o processo, que depende da relação entre observações atuais e retrospectivas, não é modificado ao se classificar os grupos etários que deveriam ter sido (digamos) 13-18, 18-23, etc., como 15-20, 20-25, etc. Consideremos, agora, o que aconteceria se combinássemos os dados referentes às duas comunidades com aproximadamente igual distribuição etária, porém com tendenciosidades constantes diferentes nos registros. Os fatores de correção para o nível de fecundidade estariam mais ou menos certos para ambas as comunidades, separadamente, e também para as observações combinadas. Expandindo essa idéia para muitas comunidades com diferentes tendenciosidades constantes podemos ver que certos tipos de alterações na idade tem pequeno efeito sobre a estimativa da fecundidade total. De modo geral, estes são erros cuja distribuição é a mesma a cada idade do período reprodutivo. Não é necessário que os erros sejam independentes da fecundidade, mas novamente a relação deve ser a mesma em cada idade do período; por exemplo, a distribuição dos erros de declaração por parte das mulheres que têm uma percentagem fixa de fecundidade acima da média deve ser independente da idade. Podemos ver isso con-

siderando a fusão de dados relativos a população com diferentes níveis de fecundidade e tendenciosidades por idades.

Para ilustrar essas idéias foram feitos cálculos para um exemplo simples, novamente utilizando-se como base as taxas de fecundidade específicas por idades individuais para a Ucrânia, período 1926-27. Foi aplicada uma distribuição retangular de erros de declaração de idade com uma tendenciosidade crescente de um ano e uma amplitude de dez anos; isto é, considerou-se que as mulheres cujas idades foram informadas como x estavam distribuídas uniformemente de seis anos abaixo a quatro anos acima da idade informada. A Tabela 3.5 mostra a fecundidade específica atual por idades (com deslocamento de meio ano), o número médio construído de filhos tidos nascidos vivos por mulher, e o cálculo das razões P/F para as informações censitárias hipotéticas com esses erros na idade.

TABELA 3.5

ILUSTRAÇÃO DO EFEITO DOS ERROS NA DECLARAÇÃO DA IDADE SOBRE AS RAZÕES P/F

LIMITES EXATOS DO INTERVALO ETÁRIO	i	f_i^*	ϕ_i	k_i	$\phi_i + kf_i = F_i$	P_i	P_i/F_i
15 — 20 . . .	1	0,045	—	1,780	0,080	0,095	1,19
20 — 25 . . .	2	0,166	0,225	2,810	0,693	0,672	0,97
25 — 30 . . .	3	0,249	1,057	2,995	1,802	1,795	1,00
30 — 35 . . .	4	0,232	2,300	3,065	3,011	3,012	1,00
35 — 40 . . .	5	0,176	3,460	3,180	4,021	4,016	1,00
40 — 45 . . .	6	0,106	4,342	3,350	4,696	4,688	1,00
45 — 50 . . .	7	0,053	4,871	3,775	5,070	5,022	0,99
Fecundidade total		5,135	5,135	—	5,135	5,123	1,00

$$f_i/f_i = 0,270; \bar{m} = 31,17$$

* Para intervalos etários meio ano menores que os apresentados.

As razões estão próximas de 1,00 exceto para o grupo mais jovem. Quando idades informadas se distribuem em torno do valor correto, a distribuição específica de fecundidade resultante é “mais ampla”, isto é, tem uma amplitude e variância maior do que a verdadeira. Isso influencia a precisão dos fatores dos quais se derivam o número de filhos tidos nascidos vivos por mulher a partir das atuais taxas de fecundidade específicas por idade. O fato de que a imprecisão introduzida é pequena, mesmo onde a extensão do erro foi grande, está ilustrado pelo exemplo.

Apesar das razões P/F serem insensíveis a algumas formas de erros nas declarações na idade, a informação extremamente errada nas idades pertencentes ao período reprodutivo pode dar origem a uma se-

qüência irregular de razões. Em alguns recenseamentos e pesquisas africanas, o erro extenso e sistemático na declaração de idade se evidencia nas distribuições etárias seriamente distorcidas — não apenas nas distribuições por idades individuais, mas também naquelas distribuídas em intervalos quinquenais. Uma forma característica da distorção nas distribuições femininas consiste nos valores insuficientemente informados nos grupos etários próximos ao começo e ao final do período reprodutivo e valores excessivamente informados em grupos próximos às idades centrais de reprodução. O efeito líquido usual é um aumento da proporção de mulheres no intervalo 15 a 49 anos, de modo que, mesmo quando os nascimentos são estimados corretamente, a taxa geral de fecundidade, a fecundidade total e as taxas de fecundidade específicas por idade ficam subestimadas pela inclusão no denominador, de mulheres que estão fora das idades férteis.

No Apêndice B, em anexo, há uma descrição de uma experiência dos efeitos, sobre as razões P/F causados por erros hipotéticos na declaração de idade que poderiam dar origem a distribuição etária muito distorcida no Congo. Para se explicar a distribuição etária informada, grande proporção das mulheres em certos grupos etários deve ter sido informada em intervalos adjacentes, e não importa qual seja a hipótese levantada sobre a fecundidade atual e a parturição das mulheres informadas nos grupos etários errados, é inevitável um efeito substancial sobre a seqüência das razões P/F .

A distorção que se evidencia em outras distribuições etárias africanas é freqüentemente menor que a do Congo (pelo menos com relação às idades reprodutivas) e as seqüências P/F em outras áreas podem não ter sido grandemente afetadas por erros de declaração de idade.

Até aqui consideramos primariamente como os erros na declaração de idade podem modificar as razões P/F . Como se notou acima, também podem levar à imprecisão na fecundidade total estimada mesmo quando o número de nascimentos registrados for correto. (Isso acontece principalmente quando um número excessivo de mulheres são informadas como estando no período reprodutivo. As idades informadas de mulheres em muitos recenseamentos africanos, mostram uma tendência à alteração para mais nos anos próximos ao começo do período reprodutivo e para menos no final, com uma aglomeração característica em direção ao centro). Se mulheres que na realidade estiverem acima e abaixo do período reprodutivo forem incluídas erradamente, as taxas de natalidade específicas por idade computadas e a taxa de fecundidade total apresentam-se menores do que deveriam ser. A inclusão errada de adolescentes que não tiveram filhos tende a diminuir tanto as taxas específicas atuais calculadas como número médio retrospectivo de filhos tidos nascidos vivos por mulheres mais jovens, e portanto, as razões P/F dessa região ficariam menos afetadas. Em idades mais velhas ocorre o contrário, já que as estimativas F_i através das taxas atuais serão reduzidas, mas o número médio retrospectivo de crian-

ças será aumentado pela inclusão de mulheres já fora do período reprodutivo.

Quando o número de mulheres de 15 a 49 anos estiver obviamente aumentado por qualquer informação errada, a razão P/F para o grupo etário de 20-24 deveria ser usada para corrigir o *número total de nascimentos* informados para o ano anterior, em vez das taxas de fecundidade específicas por idade ou a soma delas, a taxa total de fecundidade. O número de nascimentos femininos pode ser tomado como a metade do número de todos os nascimentos, já que a evidência existente quanto à razão de masculinidade, em recém-nascidos na África indica um valor próximo a 100. Deste modo, a taxa de natalidade feminina (nascimentos femininos/população feminina) pode ser estimada, e a taxa bruta de reprodução pode ser calculada aproximadamente supondo-se que a população seja estável, e aceitando-se a estimativa da taxa de natalidade feminina e da idade média da distribuição das taxas de fecundidade atuais informadas. Tal estimativa da TBR é facilmente obtida através de interpolação em populações estáveis modelo tabuladas⁵.

Outras poderosas operações de controle sobre a precisão das informações de fecundidade podem ser feitas se os dados atuais e retrospectivos forem tabulados segundo a ordem dos nascimentos. As tabelas necessárias são as de distribuição em grupos etários segundo o número de filhos tidos nascidos vivos de todas as mulheres e o daquelas que tiveram um filho no ano anterior. Assim, como, sob condições constantes, as fecundidades específicas por idade atuais acumuladas fornecem o número médio de filhos tidos nascidos vivos, a soma das taxas específicas por idade para os primeiros, segundos, terceiros, etc., nascimentos para cada idade dão a proporção de mulheres que tenham tido pelo menos um filho, dois, três etc., até esse instante, respectivamente. Os métodos descritos para comparação das observações atuais e retrospectivas de nascimentos, sem considerar a ordem, podem ser aplicados aos nascimentos segundo a ordem para se extraírem maiores informações sobre os padrões de erro. Para se obter os resultados mais precisos, os fatores para o cálculo da fecundidade acumulada média para grupos etários quinquenais, a partir das taxas atuais, teriam que ser modificados para se levar em conta as diferentes formas das distribuições da fecundidade específica, por idade, de nascimentos de ordens individuais.

Provavelmente, as comparações mais esclarecedoras são aquelas feitas entre as taxas acumuladas de primeiros nascimentos e a proporção de mulheres que são mães. A informação dos primeiros nascimentos em um ano estará afetada por erros de escala cronológica, mas é razoável esperar que, por causa da importância de um primeiro nascimento e da juventude da maioria das mães, as omissões e os diferenciais por idade no período de referência terão pouca influência. Além disso, a classificação de uma mulher como mãe estará menos sujeita a erro que a determinação do número de filhos que ela teve. Por estas

⁵ COALE & DEMENY, op cit, p 47

razões, e porque a proporção de mulheres que se tornaram mães depois da idade de trinta anos é pequena na maioria das populações africanas, a razão entre as mulheres fecundas e as taxas acumuladas de primeiro nascimento nas idades de 30-34 anos deveria ser um útil índice do erro na escala cronológica. O denominador dessa razão é, com efeito, a soma das taxas específicas por idade atuais para os primeiros nascimentos em todas as idades, que chamaremos F_c .

Infelizmente as tabulações necessárias (particularmente de nascimentos atuais) estão disponíveis para poucas populações, apesar de poderem ser preparadas sempre que o formulário de recenseamento registre a idade da mulher, o número de filhos tidos nascidos vivos, e a ocorrência ou não ocorrência de um nascimento no ano anterior para cada mulher. Apenas para um dos recenseamentos analisados há dados sobre a ordem de nascimentos atuais e mesmo assim não se acham classificados pela idade da mãe. Todavia F_c ainda pode ser estimado com pequeno erro se for possível calcular a idade média \bar{m}_c da tabela de fecundidade de primeiros nascimentos. A estimativa é feita dividindo-se os primeiros nascimentos pelo número médio de mulheres por idades individuais num intervalo centrado em \bar{m}_c e efetivamente cobrindo a amplitude na qual ocorrem esses nascimentos. A razão deste procedimento fornecer uma boa aproximação pode ser vista considerando-se uma população com um número constante de mulheres em cada idade do intervalo. Os primeiros nascimentos para as mães de cada idade são então divididos, por este número constante, esse número médio representa as taxas específicas por idade; a divisão dos primeiros nascimentos totais pela constante dá a soma das taxas. Em geral o número de mulheres por ano diminuiu com a idade, mas a média em um intervalo centrado em \bar{m}_c pode ser usada como divisor porque para taxas fixas, os nascimentos extras, devidos ao grande número de mulheres mais jovens são aproximadamente compensados pelos nascimentos que faltam, em virtude da menor quantidade de mulheres mais velhas. Um exame mais matemático desse resultado é dado por Brass⁶.

Mais detalhes sobre a aplicação do processo e a estimativa de \bar{m}_c são dados no estudo da República da Guiné.

ESTIMATIVAS DE MORTALIDADE INFANTIL E NAS PRIMEIRAS IDADES A PARTIR DE INFORMAÇÕES SOBRE A SOBREVIVÊNCIA DOS FILHOS TIDOS NASCIDOS VIVOS

Os registros de mortalidade nas comunidades africanas são ainda mais escassos e duvidosos que os de fecundidade. Os óbitos são registrados apenas em alguns países e regiões, e, na maioria destes, de for-

⁶ BRASS, William. The Estimation of Fertility Rates from Ratios of Total to First Births, op. cit., p. 46

ma incompleta. Os dados são melhores para algumas cidades, mas os habitantes destas não são representativos da população geral, quanto a distribuição por sexo e idade, condições sociais e econômicas, acesso a serviços médicos, etc. Foram feitos importantes estudos da mortalidade de grupos especiais, por exemplo, da mortalidade dos filhos tidos por mulheres que compareciam a uma clínica dos trabalhadores em mineração. Em todos estes levantamentos foram encontradas elevadas taxas de mortalidade, mas os níveis e padrões por sexo e idade foram bastante variados. Não é possível, portanto, generalizar esta evidência, com qualquer pretensão de exatidão, para proporcionar uma base que sirva para verificar as observações sobre uma comunidade africana determinada ainda que isso deva ser considerado em uma avaliação.

Nesta seção nos preocuparemos com o processamento das informações sobre a incidência de óbitos a partir dos dados obtidos em um único recenseamento ou levantamento. Em particular desenvolveu-se um método para estimar a mortalidade nas primeiras idades a partir de informações dadas pelas mães sobre o número de filhos que morreram.

Podem ser calculadas taxas de mortalidade específicas por idade, para todas as idades (incluindo a infância) a partir dos óbitos ocorridos no ano anterior, em cada intervalo de idade e segundo a população por idade, como está informada no recenseamento ou pesquisa. Essas taxas de mortalidade são taxas “atuais” de mortalidade, enquanto que as taxas de mortalidade na infância derivadas do número informado de crianças falecidas entre os filhos tidos nascidos vivos são taxas “acumuladas” (acumuladas desde o nascimento das crianças). As taxas acumuladas estão livres de erros quanto ao “período de referência” — erros que afetam a validade das taxas de mortalidade calculadas para o período imediatamente anterior ao recenseamento. Superficialmente, parece que estas taxas acumuladas desde a infância poderiam ser comparadas com as taxas atuais para a infância, a fim de que se pudesse derivar um fato de correlação para as taxas atuais de mortalidade em todas as idades, de maneira análoga ao processo usado na estimação da fecundidade. Todavia, a suposição crucial de que os erros na informação da fecundidade atual não são marcadamente seletivos, segundo a idade da mãe, não é transferível aos erros na informação da mortalidade atual. A precisão da mortalidade informada, está afetada por uma combinação de omissões e erros de período de referência, e muito provavelmente, dependerá da importância da pessoa falecida aos olhos da comunidade em geral e do informante em particular, por conseguinte, variará com a idade e o sexo. Nas estimativas de fecundidade descritas anteriormente, supôs-se que um fator de correção aplicável à fecundidade atual, de mulheres de 20 a 25 anos de idade poderia ser aplicado à fecundidade informada de todas as mulheres, mas não há garantias para se supor que, um fator de correção aplicável à mortalidade atual informada de crianças poderia ser aplicado à mortalidade informada de adultos.

Apesar dos erros na informação da mortalidade atual parecerem ser bastante distintos em diferentes idades, as próprias taxas de mortalidade em diferentes grupos etários estão fortemente inter-relacionadas em populações com dados fidedignos, e por inferência, também em populações africanas. Por isso, se deposita muita confiança em tábuas modelo de mortalidade que expressam padrões típicos de mortalidade por idades em diferentes níveis de mortalidade. Após terem sido examinados todos os dados e evidências disponíveis, as estimativas da mortalidade mais convincentes são, freqüentemente, as taxas das primeiras idades derivadas do método descrito nas páginas seguintes. Apesar das estimativas serem muito pouco precisas, parecem estabelecer um valor de mortalidade nas primeiras idades aceitável como limite inferior válido da verdadeira taxa, e isto proporciona um parâmetro muito útil — por exemplo na estimativa da fecundidade a partir dos dados da distribuição etária.

A estimativa da mortalidade nas primeiras idades

Na estimativa da mortalidade a partir de informações sobre o número de filhos tidos nascidos vivos que haviam falecido antes do recenseamento, se presumirá que as taxas específicas de fecundidade e de mortalidade por idades permaneceram constantes durante os intervalos de idade e período considerado; a experiência das mulheres sobreviventes também será tomada como sendo efetivamente a do total de mulheres que estiveram expostas aos riscos de nascimentos e óbitos de filhos. As conseqüências dos desvios oriundos destas suposições serão examinadas mais tarde. Durante toda essa apresentação, a descrição será feita em termos de ocorrências vitais para crianças de ambos os sexos, mas os resultados só se aplicam caso as observações para o sexo masculino e feminino forem tratadas separadamente.

Suponhamos que tenha havido um quadro imutável de mortalidade, tal que a proporção de sobreviventes até a idade a , (l_a/l_0 em termos de tábua de mortalidade) seja $p(a)$, e que a proporção que morre antes da idade a (${}_a q_0$ na notação da tábua de mortalidade) seja $q(a)$. Suponhamos que a proporção de mulheres na idade x que dão a luz a um filho é $f(x)$, num quadro de fecundidade que tenha sido constante nos anos recentes. A proporção de falecimentos entre os filhos tidos nascidos vivos depende de sua distribuição etária e de $q(a)$. Consideremos a distribuição etária dos filhos tidos nascidos vivos por mulheres de uma idade em particular, digamos, 25 anos, se nenhuma das crianças houvesse morrido. Essa distribuição por idades seria:

$$c_{25}(a) = \frac{f(25 - a)}{\int_0^{25} f(x) dx} \quad (1)$$

A proporção dessas crianças que teriam falecido:

$$d_{25} = \int_0^{25-\alpha} c_{25}(a) q(a) da \quad (2)$$

onde d_{25} é a proporção de filhos falecidos no total de filhos tidos nascidos vivos por mulheres de 25 anos de idade e α é a idade limite inferior do período reprodutivo.

Agora examinemos as duas funções que determinam d_{25} . A primeira, $c_{25}(a)$, é simplesmente a tábua de fecundidade proporcionada até a idade de 25 anos, *escrita de trás para frente*, e a segunda, $q(a)$, é a proporção acumulada dos filhos falecidos até a idade a sob as condições especificadas de mortalidade (ver Gráfico 3.1).

Em outras palavras, d_{25} é a média ponderada das proporções de óbitos ocorridos, onde o fator de ponderação é a distribuição etária, que os filhos tidos nascidos vivos, por mulheres de 25 anos teriam se todos houvessem sobrevivido. Existe uma idade (\bar{a}_{25}) entre zero e $25-\alpha$ tal que $q(\bar{a}_{25}) = d_{25}$. Dado que $q(a)$ cresce acentuadamente durante o primeiro ano de vida, e muito lentamente após a idade de 5 anos, e porque $c_{25}(a)$ atinge valores baixos em idades acima de 5 anos, \bar{a}_{25} é surpreendentemente pequena (2,5 anos no exemplo mostrado no Gráfico 3.1).

Para as funções de fecundidade e mortalidade especificadas, \bar{a}_x pode ser calculada para cada idade x , a partir do limite inferior do período reprodutivo, até a maior idade para a qual estão disponíveis dados sobre filhos tidos nascidos vivos (ainda vivos ou já falecidos). Sob condições de fecundidade e mortalidade constantes, tal exercício estabeleceria uma relação entre $q(a)$ (uma função de tábua de mortalidade referente aos filhos tidos pelas mulheres) e d_x (a proporção de óbitos entre os filhos tidos nascidos vivos por mulheres de idade x).

O processo de estimação de Brass torna possível usar este tipo de relação para estimar $q(1)$, $q(2)$, $q(3)$, $q(5)$, $q(10)$. . etc. (até $q(35)$ sob circunstâncias ótimas) a partir de D_1 D_2 D_3 . . . D_{10} onde D_i é a proporção de óbitos ocorridos entre os filhos tidos nascidos vivos por mulher no intervalo quinquenal de idade, sendo o primeiro intervalo 15-20.

O processo depende da igualdade aproximada entre $q(1)$ e D_1 , $q(2)$ e D_2 , . . . , $q(35)$ e D_{10} , quando as tábuas de fecundidade e mortalidade forem tábuas modelo com estruturas etárias aproximadamente iguais às encontradas em populações africanas. Estas igualdades aproximadas são afetadas mais fortemente por variações que ocorrem no padrão da fecundidade por idade, do que por variações que ocorram no padrão de mortalidade por idade. Brass calculou um conjunto de multiplicadores (ver Tabela 3.6) através dos quais os valores de D_i podem ser convertidos em estimativas de $q(a)$ — os multiplicadores variam para diferentes funções de fecundidade, mas não variam para diferentes “formas” de mortalidade. A idéia salientada, pelos multiplicadores, pode ser ilustrada por um exemplo. Uma tábua de fecundidade

FATORES QUE DETERMINAM d_{25} , A PROPORÇÃO DE ÓBITOS ENTRE FILHOS TIDOS NASCIDOS VIVOS POR MULHERES DE 25 ANOS EXATOS DE IDADE: $c_{25}(a)$, A DISTRIBUIÇÃO ETÁRIA DOS FILHOS TIDOS NASCIDOS VIVOS POR MULHERES DE 25 ANOS, E $q(a)$, A PROPORÇÃO DE ÓBITOS NA IDADE a SOB RISCOS DE MORTALIDADE PREVALECENTES

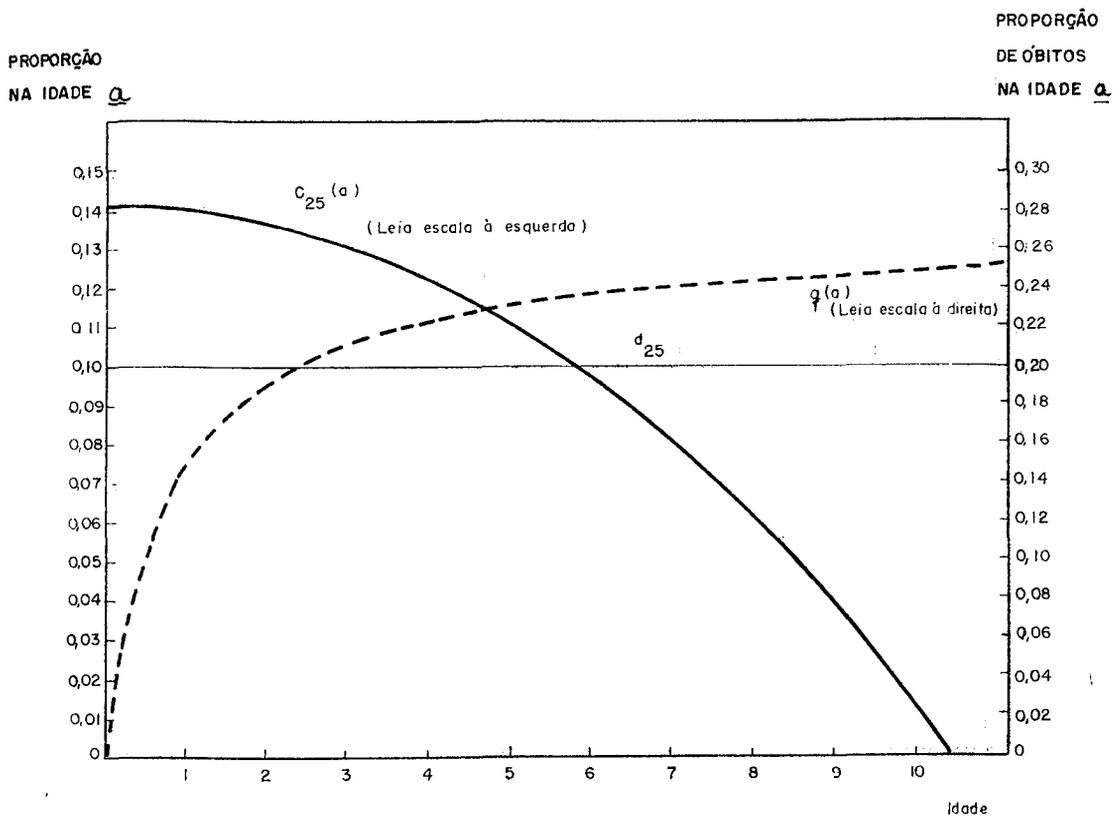


Gráfico 3.1

pode ser escolhida de modo que combinada com um padrão de mortalidade por idades, seja D_2 idêntico a $q(2)$. Considere agora uma população que tenha um período reprodutivo iniciando mais cedo. Em média os filhos tidos nascidos vivos por mulheres de 20-25 anos são mais velhos que na população onde D_2 é igual a $q(2)$. Na população com início da reprodução mais cedo, D_2 então é igual a $q(2+)$, sendo nesta população $q(2)$ menor do que D_2 . A tabela de multiplicadores de Brass dá valores que vão de mais ou menos 0,94 (reprodução iniciada muito cedo) a 1,19 (reprodução iniciada muito tardiamente) para se converter D_2 em uma estimativa de $q(2)$. Há três indicadores de reprodução iniciada cedo *versus* tardia na Tabela 3.6:

- a razão entre a parturição média de mulheres de 15-20 anos e a parturição média de mulheres de 20-25 anos (D_1/D_2);
- a idade média \bar{m} da tábua de fecundidade;
- a idade mediana, \bar{m}' da tábua de fecundidade.

TABELA 3.6

FATORES MULTIPLICADORES PARA SE ESTIMAR A PROPORÇÃO DE NASCIMENTOS DE CRIANÇAS VIVAS QUE MORREM COM IDADE a , $q(a)$, A PARTIR DA PROPORÇÃO FALECIDA ENTRE OS FILHOS TIDOS NASCIDOS VIVOS POR MULHERES DE 15-20, 20-25 ANOS, ETC.

ESPECIFICAÇÃO	IDADE	FATORES MULTIPLICADORES							
15 — 20... ..	q(1)	0,859	0,890	0,928	0,977	1,041	1,129	1,254	1,425
20 — 25	q(2)	0,938	0,938	0,983	1,010	1,043	1,082	1,129	1,188
25 — 30	q(3)	0,948	0,962	0,978	0,994	1,012	1,033	1,055	1,081
30 — 35	q(5)	0,961	0,975	0,988	1,002	1,016	1,031	1,046	1,063
35 — 40	q(10)	0,966	0,982	0,996	1,011	1,026	1,040	1,054	1,069
40 — 45	q(15)	0,938	0,955	0,971	0,988	1,004	1,021	1,037	1,052
45 — 50	q(20)	0,937	0,953	0,969	0,986	1,003	1,021	1,039	1,057
50 — 55	q(25)	0,949	0,966	0,983	1,001	1,019	1,036	1,054	1,072
55 — 60.	q(30)	0,951	0,968	0,985	1,002	1,020	1,039	1,058	1,076
60 — 65	q(35)	0,949	0,965	0,982	0,999	1,016	1,034	1,052	1,070
A. Guia para seleção do multiplicador									
P_1/P_2		0,387	0,330	0,268	0,205	0,143	0,090	0,045	0,014
\bar{m}		24,7	25,7	26,7	27,7	28,7	29,7	30,7	31,7
\bar{m}'		24,2	25,2	26,2	27,2	28,2	29,2	30,2	31,2
B. Intervalos etários decenais referentes a população feminina									
15 — 25	q(2)	0,982	1,000	1,021	1,045	1,072	1,105	1,144	1,198
25 — 35	q(5)	0,990	1,004	1,018	1,033	1,048	1,064	1,081	1,099
35 — 45	q(15)	0,977	0,993	1,009	1,024	1,040	1,056	1,071	1,086
45 — 55	q(25)	0,990	1,008	1,025	1,043	1,062	1,080	1,099	1,118
55 — 65	q(35)	0,990	1,007	1,025	1,043	1,061	1,080	1,099	1,119

O primeiro índice é uma medida do início da fecundidade e da rapidez com que cresce com a idade, e é portanto um guia aos multipli-

cadros necessários para se converter D_1 em $q(1)$, D_2 em $q(2)$, e D_3 em $q(3)$. Os outros dois índices são medidas da idade, em torno da qual está centrada a reprodução.

Em idades muito acima da média da função de fecundidade, a proporção de sobreviventes é praticamente a mesma, como se todos os nascimentos ocorressem na idade média de reprodução. Daí, \bar{m} ou \bar{m}' são os parâmetros da tábua de fecundidade indicados para seleção do multiplicador a ser usado desde D_4 a D_{10} ⁷.

Como foram desenvolvidos os multiplicadores para estimar $q(a)$

Examinaremos a relação entre D_2 e $q(2)$ para ilustrar o método pelo qual foram obtidos os multiplicadores da Tabela 3.6. Os multiplicadores se baseiam nas mesmas funções modelo de fecundidade usadas anteriormente:

$$f(x) = k(x - s) \quad (s + 33 - x)^2 \quad (3)$$

$$(\text{para } s \leq x \leq s + 33).$$

s é a idade mais jovem do período reprodutivo, e k é um fator "escala" que determina o número total de filhos tidos nascidos vivos ao final do período reprodutivo (na idade $s + 33$). k não afeta a relação entre D_2 e $q(2)$. Na ausência de mortalidade, a distribuição etária dos filhos tidos nascidos vivos de mulheres de 20 a 25 anos seria ⁸:

$$C_2(a) = \frac{\int_{20}^{25} f(x - a) dx}{\int_0^{25-s} \int_{20}^{25} f(x - a) dx da} \quad (4)$$

onde $C_i(a)$ é a distribuição etária (na ausência de mortalidade) dos filhos tidos nascidos vivos de mulheres no grupo etário de ordem i , $C_1(a)$ refere-se aos filhos tidos por mulheres de 15-20 anos, $C_2(a)$ às mulheres de 20-25 anos, etc. O Gráfico 3.2 mostra $C_2, C_3 \dots C_7$ para $s = 14,5$ anos. A proporção de óbitos ocorridos entre os filhos tidos nascidos vivos de mulheres de 20-25 anos é ⁹:

$$D_2 = \int_0^{25-s} C_2(a) q(a) da \quad (5)$$

É um processo trabalhoso, mas não é intrinsecamente difícil calcular $C_2(a)$ a partir da equação (4), para um valor dado de s , e a seguir

⁷ Não há razão forte para se preferir \bar{m} ou \bar{m}' . Se a tábua de fecundidade for conhecida, \bar{m} é facilmente calculado; de outro modo, \bar{m} é estimada como sendo a idade x na qual $P_x = 1/2 P_{s0}$

⁸ Comparar com a equação (1)

⁹ Comparar com a equação (2)

calcular D_s , a partir dessa distribuição etária, e $q(a)$ a partir de uma tábua modelo de mortalidade. Quando $s = 14,5$ e $q(a)$ é tomado de uma tábua “padrão” de mortalidade¹⁰, se encontra D_s quase igual a $q(2)$ — na verdade igual a $0,990 q(2)$. É preciso portanto um multiplicador de $1,010$ para se converter D_s em $q(2)$. P_1 e P_2 (parturição média para mulheres de 15-20 e 20-25 anos) podem ser obtidos integrando-se $f(x)$ na equação (3). Quando s for $14,5$, P_1/P_2 é $0,205$. Note que $0,205$ é um dos valores de entrada na Tabela 3.6. O cálculo dos outros valores de entrada foi feito por processo paralelo. Os fatores de correção tabulados são válidos para uma tábua “padrão” de mortalidade, e para tábuas de fecundidade pela equação (3) para diferentes valores de s , idade inicial do período reprodutivo. Sua utilidade para a aplicação prática depende destas considerações:

- 1) se (com a mortalidade “padrão”) os multiplicadores apropriados para as tábuas de fecundidade definidas pela equação (3), são também apropriados para as tábuas de fecundidade reais, com os mesmos valores de P_2/P_1 , \bar{m} , ou \bar{m}' ;
- 2) se os fatores de correção apropriados para uma tábua “padrão” de mortalidade são válidos para outros padrões de mortalidade por idade.

São tranquilizadores os extensos cálculos numéricos quanto a ambas as considerações. A insensibilidade dos multiplicadores quanto à forma detalhada da função de fecundidade é indicada pela pequena amplitude dos fatores de correção mesmo com valores de P_1/P_2 , \bar{m} ou \bar{m}' bastante diferentes. Por exemplo, um valor de \bar{m} (ou de \bar{m}') dois anos maior, ou um valor de P_1/P_2 54% maior, está associado a um multiplicador para D_s apenas 6%, e para D_i apenas 2,5% maior, apesar do multiplicador de D_i ser 11% maior. Tábuas de fecundidade observadas para idades individuais, quando combinadas com uma tábua padrão de mortalidade através da equação (5), dão os valores calculados de D_i , usualmente diferindo menos de 2% dos valores calculados com uma função de fecundidade definida pela equação (3), que tem o mesmo valor de P_1/P_2 . Ocorrem exceções — diferenças maiores — em idades mais jovens na relação de D_i com $q(1)$ em diferentes tábuas de fecundidade, para um dado valor de P_1/P_2 . Como mostra a equação (5), a relação entre D_i e $q(a)$ depende de como a estrutura etária dos filhos tidos nascidos vivos afeta a proporção de óbitos ocorridos. Uma dada razão de P_1/P_2 deixa pequeno campo de variação na divisão dos filhos tidos nascidos vivos de mulheres de 20 a 25 ou 25 a 30 em crianças mais jovens e mais velhas; mas mesmo para um dado valor de P_1/P_2 , pode haver distribuições etárias significativamente diferentes de filhos tidos nasci-

¹⁰ A tábua “padrão” de mortalidade foi obtida tirando-se a média dos valores de $q(a)$ de um certo número de tábuas de mortalidade alta. É quase idêntica à tábua modelo da família Oeste.

DISTRIBUIÇÕES ETÁRIAS DOS FILHOS TIDOS NASCIDOS VIVOS
PARA MÃES DE 20-25 ANOS [$C_2(a)$] A 45-50 ANOS [$C_7(a)$]

PROPORÇÃO NA IDADE a .

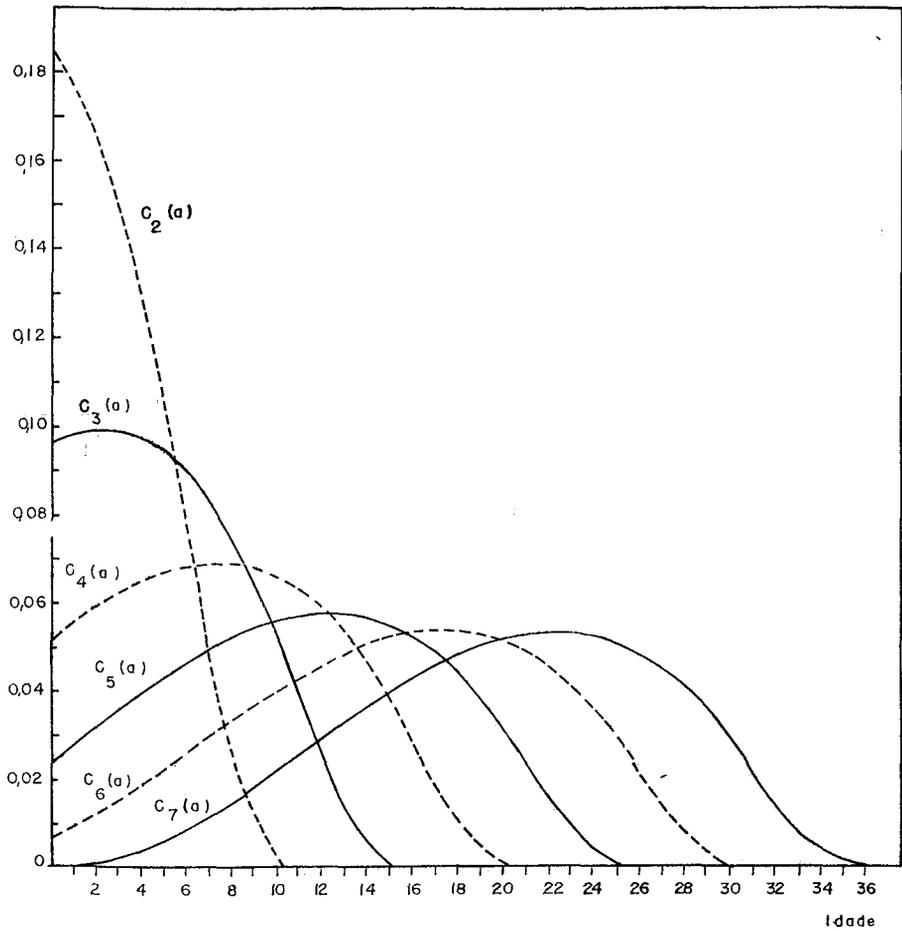


Gráfico 3.2

dos vivos de mulheres de 15-20 anos. Assim, a estimativa de $q(1)$, baseadas em D_i , estão sujeitas a grandes erros ¹¹.

Para ilustrar o efeito da forma detalhada da função de fecundidade, foram calculados valores de D_i a partir da tábua de fecundidade específica por idade (por idades individuais) para a Inglaterra e o País de Gales, em 1951, conjuntamente com a tábua "padrão" de mortalidade. Então foram computados os fatores de multiplicação, que converteriam os valores de D_i em $q(1)$, $q(2)$. . . , $q(20)$, dividindo-se $q(1)$ pelo D_1 calculado, $q(2)$ pelo D_2 calculado, e finalmente $q(20)$ pelo D_{20} calculado. Note que a tábua de fecundidade não está em boa concordância com o modelo, já que como é típico da fecundidade em países altamente modernizados, a variância é muito menor que no modelo. Os fatores de multiplicação exatos calculados para a tábua de fecundidade estão comparados aqui com os fatores tirados da Tabela 3.6 para as tábuas modelo com a mesma idade média (28,16 anos) e a mesma P_1/P_2 (0,062) da tábua para a Inglaterra e o País de Gales.

Comparação dos Fatores de Multiplicação Exatos e Estimados

As comparações a seguir mostram, nitidamente, as características das aproximações introduzidas pela aplicação do modelo de fecundidade.

GRUPOS ETÁRIOS DAS MÃES (Em anos)	FATORES DE MULTIPLICAÇÕES		
	EXATOS	ESTIMADOS: \bar{m}	P_1/P_2
15 — 20	1,119	1,006	1,207
20 — 25.	1,121	1,025	1,111
25 — 30.	1,037	1,002	1,047
30 — 35	1,017	1,008	1,040
35 — 40	1,021	1,018	1,049
40 — 45	0,998	0,995	1,031
45 — 50	0,996	0,994	1,032

Em idades acima de 30 anos, as estimativas a partir do índice de posição \bar{m} são melhores que as obtidas a partir da medida P_1/P_2 e estão muito próximas aos fatores exatos. Para os grupos etários 20-25 e 25-30 anos, os fatores obtidos através de P_1/P_2 observados dão resultados mais satisfatórios. A regra proposta (o uso do índice P_1/P_2 para os três primeiros grupos e \bar{m} para os subseqüentes) leva a um erro nos fatores estimados de menos de 1%, exceto para as mães muito jovens. Para o primeiro grupo etário de mulheres, nenhuma das medidas de localiza-

¹¹ Há outras razões para não se confiar na informação sobre a proporção de óbitos ocorridos entre os filhos tidos nascidos vivos de menores de 20 anos. Em levantamento por amostragem, ou recenseamento de uma pequena população, o número total de crianças mortas para esse grupo etário pode ser pequeno e sujeito a grandes flutuações de amostragem. Em segundo lugar, em algumas populações (se não em todas) a mortalidade infantil entre os filhos tidos nascidos vivos de mulheres muito jovens não é representativa da mortalidade infantil geral.

ção dá estimativas que estejam muito próximas ao valor exato. O erro, a partir do uso de P_1/P_2 (por volta de 8%), não levaria a um $q(1)$ grosseiramente impreciso. As razões para a impropriedade do modelo nos anos iniciais do período reprodutivo já foram discutidas com relação à estimativa do número médio de filhos tidos nascidos vivos por mulher a partir das taxas atuais de fecundidade. As limitações são impostas pela natureza das evidências e das observações; não poderiam ser ultrapassadas simplesmente pela modificação das técnicas de cálculo.

Na derivação dos fatores de multiplicação foi usada uma tábua padrão fixa de mortalidade. É fácil ver que, exatamente, os mesmos resultados teriam sido obtidos de outra tábua de mortalidade, na qual, a probabilidade de morrer, antes de qualquer idade, fosse uma constante C vezes a medida padrão correspondente. Todas as proporções de filhos falecidos por idade da mãe, e, conseqüentemente, os valores de D , seriam multiplicados por C ; as razões entre as probabilidades q e as proporções D , que dão os fatores de multiplicação, não seriam alteradas.

Os multiplicadores apropriados para uma tábua "padrão" de mortalidade, $q_s(a)$ também são apropriados para qualquer tábua de mortalidade onde a proporção de óbitos em cada idade segue o mesmo comportamento que o da tábua padrão; isto é, onde $q(a) = V \cdot q_s(a)$, sendo V qualquer constante. Para as tábuas de mortalidade que se ajustam as tábuas modelo Oeste e com \bar{e}_0 entre 30 e 50 anos, a proporção que sobrevive até as idades jovens adultas é praticamente um multiplicador constante dos valores do padrão, como mostra a Tabela 3.7.

TABELA 3.7

$q(a)$ EM QUATRO TÁBUAS DE MORTALIDADE MODELO OESTE COMO MÚLTIPLO DE $q_s(a)$

IDADE a	$e_0^o = 20,0$	$e_0^o = 30,0$	$e_0^o = 40,0$	$e_0^o = 50,0$
1... ..	2,44	1,71	1,19	0,79
5..... ..	2,31	1,68	1,19	0,79
10... .. .	2,27	1,68	1,20	0,80
15	2,24	1,68	1,21	0,81
20... .. .	2,17	1,64	1,19	0,81

Outras tábuas modelo não concordam tão bem com o padrão nos modelos de mortalidade. Entre as tábuas modelo de Coale-Demeny, as menos concordantes são as tábuas modelo Sul (ver Tabela 3.8). Para uma tábua típica de mortalidade como padrão Sul (Espanha, 1930), o fator de correção para estimar $q(2)$ tem apenas 2% de erro quando P_1/P_2 é igual a 0,205 e o erro para $q(3)$, $q(5)$ e idades maiores é ainda menor. Todavia, se a fecundidade tiver um padrão extremamente jovem, o erro na estimativa de $q(2)$ é maior, da mesma forma que o erro nas estimativas de $q(1)$, quando as estruturas de mortalidade se "desviam" do padrão.

TABELA 3.8

$q(a)$ EM QUATRO TÁBUAS MODELO DE MORTALIDADE SUL COMO UM MÚLTIPLO DE $q_s(a)$

IDADE a	$e_s^o = 20$	$e_s^o = 30$	$e_s^o = 40$	$e_s^o = 50$
1..	2,06	1,53	1,15	0,86
5	2,44	1,83	1,38	0,95
10	2,41	1,83	1,35	0,95
15	2,36	1,80	1,34	0,94
20	2,25	1,74	1,30	0,91

Há indicações (não inteiramente conclusivas) de que o padrão de mortalidade nas primeiras idades e na infância, da família de tábuas de mortalidade modelo Norte se ajustam melhor à experiência africana que os padrões da família Oeste. A relevante característica que se distingue na mortalidade nos modelos Norte, é que a mortalidade infantil é relativamente baixa e a mortalidade acima de um ano relativamente alta. Esta característica combina com o que muitos observadores (e alguns estudos especiais) informaram sobre populações africanas. Portanto, construiu-se uma tábua de mortalidade "típica africana"¹² em que a infância se baseia na família Norte, e calculou-se um conjunto alternativo de fatores para a conversão de D_i em $q(a)$, com base nesse modelo de mortalidade algo diferente. Esses multiplicadores alternativos estão comparados na Tabela 3.9 com os multiplicadores baseados numa tábua padrão de mortalidade que se ajusta ao modelo Oeste.

Apesar do padrão de diferenças entre os dois conjuntos de fatores ser complexo, as características principais são claras. Os maiores desvios ocorrem quando os fatores estão mais afastados do valor 1,00. Os maiores erros são encontrados nos grupos etários referentes a mães mais jovens, e tendem a decrescer para as mulheres mais velhas. Exceto para o primeiro grupo etário e, em algumas ocasiões, para as posições extremas da distribuição de fecundidade, a diferença máxima nos fatores é de mais ou menos 3,5%. Além disso, a direção da diferença não é a mesma em todos os grupos etários.

Estas diferenças indicam que na presença de incerteza quanto à forma do padrão de mortalidade não se pode esperar estimativas precisas. Por outro lado, sabe-se que existem outras fontes de erro e tendenciosidade, e a imprecisão resultante do padrão de mortalidade por idades não parece importante¹³.

¹² Ver págs 79/80

¹³ Os fatores multiplicadores baseados no modelo Norte foram desenvolvidos e comparados com os baseados no modelo Oeste apenas depois de ter sido completada a análise demográfica informada em outros Capítulos. Por ser o verdadeiro padrão de mortalidade assunto de conjecturas e por serem pequenas as diferenças para todas as idades exceto as jovens e para todos os padrões extremos de fecundidade, os cálculos baseados na vida "padrão" original foram mantidos

TABELA 3.9

COMPARAÇÕES ENTRE OS FATORES MULTIPLICADORES
DERIVADOS DOS MODELOS DE MORTALIDADE PADRÃO (S)
E AFRICANO (A)

INTERVALO ETÁRIO	S	A	S	A	S	A	S	A
15 — 20 .	0,890	0,831	0,977	0,936	1,129	1,124	1,425	1,530
20 — 25	0,959	0,916	1,010	0,983	1,082	1,076	1,188	1,212
25 — 30	0,962	0,917	0,994	0,960	1,033	1,013	1,081	1,081
30 — 35	0,975	0,937	1,002	0,972	1,031	1,010	1,063	1,055
35 — 40	0,982	1,001	1,011	1,038	1,040	1,075	1,069	1,115
40 — 45 .	0,955	0,962	0,988	0,999	1,021	1,037	1,052	1,076
45 — 50	0,953	0,954	0,986	0,989	1,021	1,026	1,057	1,067
\bar{m}	25,7		27,7		29,7		31,7	

O que foi dito até agora sobre o método de Brass para estimar a mortalidade pode ser sumarizado em três afirmações:

- 1) as proporções de óbitos ocorridos entre os filhos tidos nascidos vivos por mulheres de 15-20, 20-25, 25-30, . . . , 60-65 anos são aproximadamente iguais a $q(1)$, $q(2)$, $q(3)$, $q(5)$, $q(10)$, . . . , $q(35)$, desde que o padrão de fecundidade tenha uma idade média em torno de 28 anos;
- 2) a Tabela 3.6 torna possível estimar $q(1)$, $q(2)$, etc. a partir de D_1 , D_2 , . . . , etc., para padrões de fecundidade com a reprodução iniciada cedo ou tardiamente, assim como para uma idade média de 28 anos;
- 3) as variações de padrão entre as estruturas de fecundidade com a mesma P_1/P_2 (ou a mesma \bar{m}) e as variações em padrões de mortalidade por idade causam erros na estimativa de $q(2)$, $q(3)$, $q(5)$, etc., mas apenas de pequeno valor relativo; todavia, aquelas variações podem causar erros maiores nas estimativas de $q(1)$.

Foi suposto implicitamente nesta discussão que as estimativas são feitas a partir de dados precisos sobre uma população com taxas constantes de fecundidade e de mortalidade. Consideraremos agora de que maneira as estimativas de $q(1)$, $q(2)$, . . . são afetadas por mudanças nas taxas de fecundidade e de mortalidade, pelas imprecisões encontradas em recenseamentos ou levantamentos africanos, e pelas forças seletivas que tendem a tornar as taxas informadas de sobrevivência não representativas.

Fecundidade variável e as estimativas de Brass sobre mortalidade infantil e nas primeiras idades

As estimativas de Brass de mortalidade nas primeiras idades são afetadas pelo *nível* de fecundidade. Por isso a imprecisão introduzida pela fecundidade variável surge das diferenças no padrão por idades entre a fecundidade de uma dada *coorte* (digamos, a das mulheres que agora têm 25-30 anos) e uma dada amostra representativa (cross section) da população. Imagine que a fecundidade em todas as idades venha declinando. A distribuição por idades estimada para os filhos tidos nascidos vivos por mulheres de 25-30 anos é baseada na suposição de que quando essas mulheres estavam com 15-20 e 20-25 anos, tinham a mesma fecundidade que as mulheres que hoje têm essas idades. Mas quando a fecundidade caiu, as mulheres que agora têm 25-30 anos tiveram fecundidade *mais alta* quando tinham 15-20 ou 20-25 anos que a informada agora pelas mulheres mais jovens. Então a distribuição por idade dos filhos tidos nascidos vivos contém mais crianças de idades mais elevadas do que, implicitamente, se supõe ao selecionar o multiplicador para estimativa de $q(3)$. O valor de P_1/P_2 para a *coorte* é maior do que na amostra representativa (cross section) coberta pelo recenseamento. Então o multiplicador selecionado da Tabela 3.6 é muito grande, e a estimativa de $q(3)$ também é grande demais. Mas não é provável que o efeito seja importante, porque com poucas exceções, quaisquer tendências de fecundidade experimentadas na África são graduais, e porque os fatores da Tabela 3.6 não são sensíveis a pequenas diferenças em P_1/P_2 .

Taxas de mortalidade variáveis e as estimativas de Brass sobre a mortalidade nas primeiras idades

A conversão de D_i em $q(a)$ é derivada a partir do uso de valores de uma tábua de mortalidade, que expressa os riscos de mortalidade a que estão expostas as crianças menores. Todavia, $q(a)$ só pode ser identificado com a tábua de mortalidade prevalecente na época do levantamento se a mortalidade tiver sido constante durante os anos anteriores ao levantamento. Na verdade, o $q(a)$ na equação (5) é a proporção de óbitos por idade a entre as pessoas nascidas a anos antes da pesquisa ou recenseamento; e expressa mais a mortalidade de *coorte* que a mortalidade do período do levantamento. Quando a mortalidade está variando, a maior variação se dá usualmente na mortalidade infantil. Portanto a mortalidade infantil experimentada por crianças de 9 ou 10 anos pode ter sido bem diferente da experimentada pelas crianças que agora têm menos de um ano. O Gráfico 3.3 mostra um $q(a)$ que expressa a mortalidade acumulada experimentada para cada uma das coortes nascidas nos últimos quinze anos, sob a suposição de que a esperança de vida ao nascer cresceu de 38,5, para 52,4 no intervalo de

PROPORÇÃO DE ÓBITOS POR IDADE a , DE ACORDO COM A TÁBUA ATUAL DE MORTALIDADE, $q(a)$ (MODELO OESTE, $e_0^o = 52,5$), E DE ACORDO COM OS RISCOS EXPERIMENTADOS POR COORTE QUANDO e_0^o TEM SUBIDO ANUALMENTE EM 0,92 ANOS, $q'(a)$

PROPORÇÃO
DE ÓBITOS
NA IDADE a

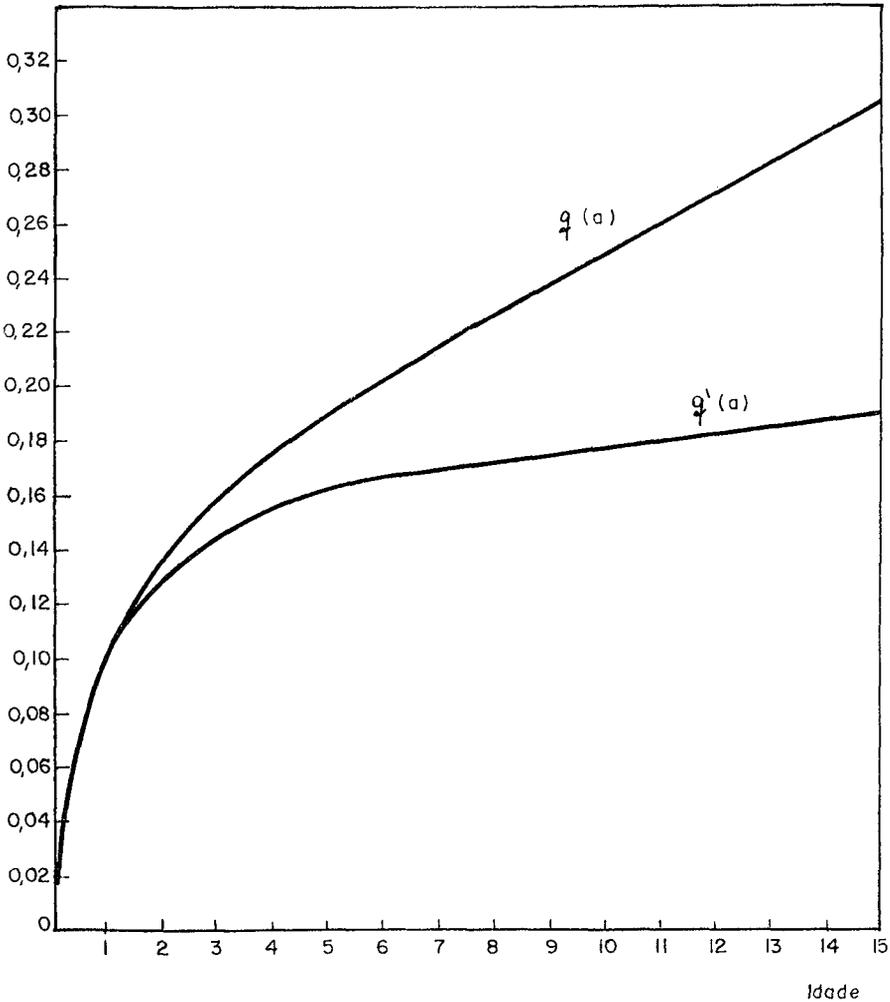


Gráfico 3.3

quinze anos. Para comparação, também está apresentada a curva representando a mortalidade [$q'(a)$] na tábua de mortalidade *atual*. Sob tais circunstâncias, D_3 é mais ou menos 1,25 vezes $q'(3)$ — e mais ou menos 1,15 vezes $q(3)$ — se a idade média das tábuas de fecundidade for 27,7 anos. O multiplicador na Tabela 3.6 é 0,994; o multiplicador apropriado para se obter o valor *atual* de $q(3)$ a partir de D_3 seria, aproximadamente apenas 0,80. Quando as taxas de mortalidade caem rapidamente, a sucessão de estimativas $q(1)$, $q(2)$, $q(3)$, $q(5)$. . . tende a aumentar mais rapidamente com a idade que $q'(a)$, que expressa a mortalidade atual, ou que o $q(a)$, que expressa os óbitos acumulados de crianças com idade a no momento do levantamento. Apesar disso as estimativas de $q(2)$ e $q(3)$ seriam representativas da mortalidade média para um curto período (menos que uma década) anterior ao recenseamento ou pesquisa.

As estimativas de Brass sobre a mortalidade e a tendência à declaração incompleta do número de filhos tidos nascidos vivos

Anteriormente, nesse capítulo (na discussão da estimação da fecundidade) notou-se uma tendência nas mulheres mais velhas (ou mulheres com muitos filhos) de declarar incompletamente o número de filhos que tiveram. Parece plausível que as mulheres tenderiam a omitir uma proporção maior de filhos que tenham morrido do que daqueles que tenham sobrevivido. Na Índia, o Levantamento Nacional por Amostragem encontrou a mortalidade infantil informada a partir das crianças nascidas no passado, remoto, menor que a partir das crianças nascidas nos anos imediatamente anteriores ao levantamento¹⁴, apesar da mortalidade ter indubitavelmente declinado ao invés de aumentado nos anos em questão. Não há garantia segura para a afirmativa de que as tendenciosidades encontradas nos levantamentos indianos sejam reproduzidas na África, mas a experiência na Índia reforça o ceticismo natural quanto à exatidão da proporção de óbitos ocorridos comunicados pelas mulheres mais velhas. A forma de tendenciosidade a se esperar deste possível efeito é um $q(a)$ estimado que aumenta muito lentamente com a idade ou, se o efeito foi suficientemente grande, diminui com a idade. Esta tendenciosidade, é claro, seria menor nas estimativas de $q(1)$, $q(2)$ e $q(3)$.

Outro tipo de omissão que se espera em dados de levantamento sobre sobrevivência infantil é a omissão das crianças que morreram pouco depois do nascimento. As pessoas entrevistadas podem não compreender a necessidade de informar o nascimento de uma criança que não tenha sobrevivido tempo bastante para ter um lugar na família. A tendenciosidade causada por tais omissões presumivelmente reduziria em aproximadamente igual quantidade, as estimativas de $q(a)$ em todas

¹⁴ DAS GUPTA, Agitet ali *Couple Fertility India*. Departamento de Assuntos Econômicos, 1955. p. 59-61. (The National Sample Survey, 7)

as idades, mas a tendenciosidade não seria revelada por qualquer distorção visível na forma de $q(a)$.

As estimativas de Brass sobre mortalidade e a declaração errada da idade

A distribuição etária registrada na maioria dos recenseamentos ou pesquisas africanas, é claramente inexata, de modo que, uma definição, cuidadosamente formulada de D_2 deveria dizer as “proporções de óbitos ocorridos entre os filhos tidos nascidos vivos informados por mulheres cuja *idade registrada* é 20-25 anos”. As idades verdadeiras de tais mulheres incluem algumas que têm realmente 20-25 anos e muitas que não têm. Uma discussão geral dos efeitos da informação errada da idade não pode ser tentada aqui porque os efeitos dependem, de maneira complexa, das relações que possam existir entre a parturição da mulher e a sobrevivência de seus filhos menores, por um lado, e a informação errada da idade por outro. D_2 é, (repetimos) realmente, uma afirmação das proporções de óbitos ocorridos entre os filhos tidos nascidos vivos por mulheres cuja parturição média é a das mulheres *que informam* ter 20-25 anos. A distribuição etária dos filhos tidos nascidos vivos é avaliada por P_1/P_2 . Mas se, por exemplo, adolescentes (de 13 a 19 anos) com parturição igual a 2 fossem sistematicamente informadas como tendo mais de 20 anos, o valor registrado de P_1/P_2 seria muito pequeno, e o multiplicador aplicado a D_2 para estimar $q(2)$ tenderia a ser muito grande. Outras distorções poderiam resultar se as estimativas da idade das mulheres cujo filho (ou filhos) tivesse morrido fossem sistematicamente diferentes das estimativas para as mulheres cujos filhos tivessem sobrevivido.

O fato, de que os índices por idades usados para se calcular os fatores de correção, da Tabela 3.6 são baseados em fecundidade informada ou parturição informada por idades significa que os erros na declaração da idade que afetam D_1 , D_2 , etc., são os mesmos que afetam os fatores de correção estimados. Se os erros na declaração da idade não tivessem nenhuma relação sistemática com parturição, fecundidade ou proporção de filhos sobreviventes, a informação imprecisa da idade não teria em consequência um efeito maior sobre as estimativas de $q(a)$.

As estimativas de Brass sobre mortalidade e as forças seletivas que afetam as proporções de óbitos ocorridos entre os filhos tidos nascidos vivos

As proporções de óbitos ocorridos entre os filhos tidos nascidos vivos por mulheres em um dado grupo etário — digamos, de 30 a 35 anos — não são necessariamente representativos das proporções de óbitos ocor-

ridos entre *todas* as crianças nascidas na mesma época, em que os filhos dessas mulheres. Pode haver, e de fato parece que provavelmente há, uma mortalidade mais baixa entre os filhos tidos por mulheres que tenham elas próprias sobrevivido, que entre os filhos tidos por mulheres que morreram a seguir. Muitos fatores tendem a causar tal correlação: contágio dentro das famílias, ambientes de mortalidade mais alta que afetam tanto a mãe como a criança, e a influência direta da própria morte da mãe que torna mais provável a mortalidade de seus filhos. Esta associação implica que as estimativas de $q(a)$ baseadas nas proporções de óbitos ocorridos entre sua prole, informadas por mulheres vivas, apresentam uma tendenciosidade para menos. A possível tendenciosidade é menos importante para mulheres mais jovens (abaixo de 30 anos) por causa do efeito acumulado relativamente pequeno da mortalidade de mães nas idades mais jovens.

Outra fonte de tendenciosidade é a associação de mortalidade infantil, por um lado, e idade e parturição da mãe, pelo outro. Em muitas populações, com informações confiáveis, a mortalidade dos primogênitos e de filhos tidos por mulheres adolescentes é substancialmente maior que a mortalidade de outras crianças. Esta diferença, se também predominante na África, tenderia a fazer com que o $q(1)$ estimado a partir de D_1 fosse alto de forma não representativa.

Tendenciosidade global nas estimativas de mortalidade pelo método de Brass

Não é possível chegar-se a uma conclusão final válida para todos os dados africanos, sobre as tendenciosidades líquidas que surgem, nas estimativas de $q(a)$, a partir das proporções de óbitos ocorridos entre os filhos tidos nascidos vivos. Informações incompletas de ocorrências passadas causariam, quase certamente, uma subestimação da mortalidade, porque a omissão de uma proporção maior da prole viva que da prole morta é pouquíssimo provável. Similarmente, as tendenciosidades na mortalidade diferencial entre os filhos de mães sobreviventes ou não sobreviventes causam uma subestimação de $q(a)$. Tanto as falhas de memória, que aumentam com a idade, como a maior mortalidade entre órfãos tendem a fazer com que $q(a)$ decresça cada vez mais por debaixo dos valores apropriados, à medida em que a aumenta, enquanto a omissão de filhos que morreram pouco após o nascimento diminui as estimativas de $q(a)$ em todas as idades. Uma tendência declinante na mortalidade tende a fazer com que as estimativas de $q(a)$ sejam muito altas se forem interpretadas como a mortalidade na época do recenseamento ou pesquisa — cada vez mais excessivamente altas à medida em que a aumenta. Todavia, quando a mortalidade tem estado declinando, as estimativas não tenderão a ser muito grandes se interpretadas como

taxas médias predominantes nos anos pré-censitários — os últimos dois ou três anos para $q(1)$, os últimos cinco ou seis para $q(2)$, os últimos oito ou nove para $q(3)$.

A informação errada da idade pode aumentar ou diminuir as estimativas de mortalidade para uma idade, considerada, talvez produzindo uma seqüência irregular de estimativas de $q(a)$, mas não levando a qualquer tendenciosidade para cima ou para baixo que de modo geral se possa predizer.

A omissão de filhos mortos em levantamentos retrospectivos tem sido notada em países não-africanos, e as informações de mortes, entre crianças de menos de um ano, relativas aos nascimentos informados, indicam taxas de mortalidade infantil inaceitavelmente baixas. Esses fatos reforçam a hipótese intrinsecamente plausível de que as entrevistadas africanas tendem a omitir uma proporção maior de crianças mortas que de crianças sobreviventes. Esta probabilidade, combinada com uma possível tendenciosidade para menos resultante de forças seletivas, leva à conclusão de que as estimativas de $q(a)$ pelos métodos de Brass usualmente deveriam ser encaradas como limites inferiores.

A seqüência exata de $q(1)$, $q(2)$, $q(3)$, $q(5)$, etc., não pode ser levada a sério como números exatos. As omissões, as declarações erradas de idade, e o efeito de tendências de fecundidade e mortalidade predominam em grande escala. $q(1)$ é um número especialmente não fidedigno, e $q(10)$, $q(15)$, etc. são baseados na lembrança de ocorrências remotas por mulheres cujas respostas provavelmente são pouco representativas, especialmente para a experiência atual de mortalidade. As estimativas $q(2)$, $q(3)$ e $q(5)$ podem, freqüentemente, ser aceitas como indicações mínimas do nível recente de mortalidade infantil e nas primeiras idades.

É surpreendente e tranquilizador quanto à utilidade do método que os níveis de mortalidade infantil e nas primeiras idades inferidos das proporções de óbitos ocorridos entre os filhos tidos nascidos vivos, seja quase sempre maior que o nível informado por mulheres, quando interrogadas sobre as ocorrências do *ano anterior*. A razão para as entrevistadas informarem sobre as crianças mortas, de forma mais completa, quando interrogadas sobre sua experiência total, que quando interrogadas sobre a experiência num período recente especificado é um enigma. Talvez uma relutância em informar sobre crianças mortas esteja mais prontamente expressa num erro de período de referência que na supressão completa da ocorrência. Qualquer que seja a explicação, as estimativas de Brass de $q(a)$ parecem dar indicações mais válidas de mortalidade de crianças muito jovens que perguntas diretas sobre a mortalidade.

Um exemplo do uso das tabelas

As proporções de óbitos ocorridos de crianças do sexo feminino, que morreram antes do levantamento, por amostragem, na Guiné Francesa em 1954-55 foram:

IDADE DA MÃE EM ANOS	PROPORÇÃO DE FILHOS MORTOS
15 — 20	0,224
20 — 25	0,299
25 — 30	0,354
30 — 35	0,379
35 — 40	0,401
40 — 45	0,429
45 — 50	0,448
50 — 55	0,478
55 — 60	0,484
60 — 65	0,505

A partir de informações retrospectivas sobre crianças nascidas, P_1 é 0,54 e P_2 é 1,75 resultando P_1/P_2 igual a 0,309. A idade média da distribuição específica de fecundidade, calculada diretamente das taxas encontradas a partir dos nascimentos do ano anterior ao levantamento é 27,32. A partir da Tabela 3.6, os fatores multiplicadores correspondentes a P_1/P_2 igual a 0,309, são encontrados, interpolando-se entre a segunda e a terceira coluna de fatores, e os correspondentes à média de 27,32 por interpolação entre a terceira e a quarta coluna de fatores, com os seguintes resultados:

GRUPOS ETÁRIOS DAS MÃES	FATORES	
	A partir de P_1/P_2	A partir de \bar{m}
15 — 20	0,903	0,957
20 — 25	0,967	1,000
25 — 30	0,967	0,988
30 — 35	0,979	0,997
35 — 40	0,987	1,005
40 — 45	0,960	0,981
45 — 50	0,958	0,979
50 — 55	0,972	0,994
55 — 60	0,974	0,995
60 — 65	0,971	0,992

O uso dos três primeiros fatores da fileira de P_1/P_2 e dos remanescentes da fileira de \bar{m} para a multiplicação das correspondentes proporções de óbitos ocorridos dos filhos tidos dá os valores de q abaixo:

$q(1)$	$q(2)$	$q(3)$	$q(5)$	$q(10)$	$q(15)$	$q(20)$	$q(25)$	$q(30)$	$q(35)$
0,202	0,289	0,342	0,378	0,403	0,421	0,439	0,475	0,482	0,501

Os oito primeiros valores mostram um padrão normal de mortalidade com pequena irregularidade, mas os dois últimos são muito baixos para estarem coerentes com os outros.

Estimativa através de tábuas modelo de mortalidade e de populações estáveis

Em anos recentes as tábuas modelo de mortalidade e as populações estáveis têm sido usadas com crescente freqüência na estimativa das características de populações que tenham apenas escassos registros de nascimentos e óbitos, e apenas dados incompletos ou imprecisos sobre distribuições por sexo e idade. As informações sobre populações africanas são escassas, incompletas e freqüentemente inexatas; e as circunstâncias da vida africana indicam que a maioria das distribuições etárias africanas não são, provavelmente, muito diferentes de distribuições estáveis. Em conseqüência, as tábuas de mortalidade e as técnicas de população estável têm sido os principais meios de estimativa usados neste livro.

Uma tábua modelo de mortalidade é um padrão estimado de mortalidade baseado em:

- 1) modelos de mortalidade observados na experiência informada de várias populações diferentes da população em questão;
- 2) algum indício, ou aspecto registrado, sobre a mortalidade desta última.

A justificativa lógica para a estimativa por meio de tábuas modelo de mortalidade é que os padrões de mortalidade por idades só variam dentro de limites restritos. As taxas de mortalidade para um grupo etário estão bastante correlacionadas com as taxas de mortalidade para outros grupos; as tábuas modelo de mortalidade são uma expressão dessa inter-relação.

Muitos métodos alternativos poderiam ser criados para sumarizar possíveis padrões de mortalidade na forma de tábuas modelo¹⁵. Dois tipos de tábuas modelo foram usados no Departamento de Pesquisa

¹⁵ Muitos métodos diferentes já usados pelos demógrafos estão em COALE & DEMENY, op. cit., p. 47

Demográfica (Office of Population Research) na análise dos dados africanos — um tipo está aqui designado como as tábuas modelo de mortalidade de Coale-Demeny e o outro como as tábuas modelo de mortalidade de Brass. Associado com qualquer tábua de mortalidade, está um conjunto de estruturas por idade estáveis, que incorporam a tábua de mortalidade e as taxas alternativas de crescimento, determinadas pela mortalidade específica e por vários padrões possíveis de fecundidade. Cada forma de tábua modelo de mortalidade dá origem, assim, a correspondentes populações modelo estáveis.

As tábuas de mortalidade e as populações estáveis de Coale-Demeny foram usadas para estimar taxas de natalidade, taxas de mortalidade, e a composição etária corrigida do Congo, Sudão, Quênia, Tanganica, Niger, Senegal, Libéria, Gâmbia, Gana, Togo, Zâmbia, Rodésia, Nigéria e dos territórios portugueses¹⁶.

O método de estimativa foi escolher uma população estável que compartilhasse de uma característica selecionada da distribuição etária registrada, e alguma outra propriedade observada ou estimada da população, tal como a taxa intercensitária de crescimento ou as taxas de mortalidade infantil. Supôs-se, então, que a taxa de natalidade, a taxa de mortalidade, e a composição etária por grupos quinquenais da população estável serviam como aproximações de parâmetros demográficos reais. Muitas vezes, as estimativas obtidas por esse método puderam ser confrontadas com estimativas de fontes essencialmente independentes; por exemplo, a taxa de natalidade pode ser comparada com as estimativas obtidas a partir do método de Brass para o cálculo da fecundidade. Onde foram possíveis comparações desse tipo, a concordância foi satisfatória.

As tábuas modelo de mortalidade e as populações estáveis de Brass foram usadas de maneira pouco diferente. As estimativas de fecundidade de áreas analisadas por Brass foram obtidas através dos métodos descritos anteriormente nesse capítulo. Foi selecionada uma tábua modelo de mortalidade que fosse coerente com as estimativas de mortalidade nas primeiras idades obtidas através de métodos já descritos, e também coerente (onde fosse possível) com as informações comunicadas sobre os óbitos do último ano, entre a população não-infantil. Foi escolhida, então, uma população estável que tivesse os níveis estimados de fecundidade e mortalidade, e a distribuição etária estável foi comparada com a distribuição registrada. Novamente o dispositivo “tábuas-modelo-de-mortalidade-população-estável” serviu como meio de controle da consistência da distribuição etária registrada e das estimativas de fecundidade e mortalidade derivadas de outros tipos de dados, e novamente a população estável deu a base para uma distribuição etária corrigida; mas a forma da operação de controle da consistência foi diferente. As tábuas modelo de mortalidade e as populações estáveis de

¹⁶ Ver os capítulos 4, 6, e 8-10 de “The Demography of Tropical Africa”, de William Brass

Brass foram usadas desse modo do Dahomey, na Guiné, na República Central Africana, no Volta Superior, no Camerum do Norte, no Vale do Senegal, na Costa do Marfim e em Mali ¹⁷.

Os dois tipos de tábuas modelo de mortalidade

Um conjunto de tábuas modelo de mortalidade apresenta as taxas específicas de mortalidade, por idade, ou o número de sobreviventes, em cada idade, a ser esperado sob condições variáveis de mortalidade. Se se supõe que a experiência de mortalidade pode ser representada estritamente, pela variação no valor de um parâmetro, uma única "família" das tábuas modelo de mortalidade pode ser calculada. Nestas circunstâncias o número de sobreviventes em cada idade, ou as taxas de mortalidade em todas as idades, é estimado a partir de um só índice da experiência de mortalidade de uma população, tal como a esperança de vida ao nascer, ou a proporção dos sobreviventes até a idade de dois anos. As tábuas modelo de mortalidade das Nações Unidas são um exemplo. O exame das tábuas de mortalidade registradas revela casos de tábuas de mortalidade divergindo marcadamente das tábuas modelo das Nações Unidas, e está claro que, nenhuma família de tábuas com um parâmetro único produziria um bom ajuste aos diversos padrões de mortalidade observadas.

Foram usados dois métodos de acomodação do recurso tábua-modelo-de-mortalidade à diversidade de padrões de mortalidade. Um (empregado nos modelos Coale-Demeny) é localizar muitas "famílias" individuais de tábuas de mortalidade que sejam mais uniformes em padrão que a totalidade das tábuas observadas de mortalidade, e então compor muitas coleções de tábuas modelo de mortalidade, com um único parâmetro, cada uma representando um conglomerado de padrões similares de mortalidade. A outra abordagem (exemplificada pelas tábuas modelo de Brass e também por tábuas modelo de mortalidade baseadas em análise de fatores) é empregar dois ou mais parâmetros para formar tábuas modelo de mortalidade que exibam alguns dos diversos padrões por idade, encontrada em tábuas de mortalidade observadas.

As tábuas de mortalidade modelo de Coale-Demeny foram deliberadamente baseadas num grupo pré-selecionado de tábuas de mortalidade, registradas, sabidamente derivadas de dados relativamente fidedignos; isto é, as tábuas de mortalidade derivadas de dados com extremos erros de informação de idade, vultosas omissões de óbitos, ou ampla subenumeração da população base, foram abandonadas. As tábuas de mortalidade derivadas de dados qualitativamente pobres, frequentemente, têm padrões de mortalidade por idade fortemente individuais, e na opinião de Coale-Demeny não é certo se tal divergência re-

¹⁷ Ver capítulo 7 de "The Demography of Tropical Africa", de William Brass

sulta de diferenças genuínas no modelo de mortalidade ou é produzida por erros na informação da idade, no registro de óbitos, ou na enumeração da população base. Não se pode supor que as tábuas de mortalidade de *todas* as populações se aproximem de uma das quatro famílias. Três das famílias são totalmente derivadas de tábuas européias de mortalidade, e 60% das tábuas em que se baseia a quarta família são européias; é improvável que a variedade de padrões europeus de mortalidade esgote a variedade a ser encontrada no mundo.

Se existissem tábuas africanas de mortalidade precisas, elas, na melhor das hipóteses, concordariam vagamente com qualquer modelo de mortalidade na experiência do limitado número de populações que têm estatísticas vitais ou recenseamentos precisos. Não há nenhuma razão para se esperar que as tábuas de mortalidade para todas as populações africanas, se fossem conhecidas, se ajustassem bem às tábuas encontradas nas famílias de Coale-Demeny. Mas uma dessas famílias (a modelo "Oeste") está próxima ao padrão central apresentado por todas as tábuas de mortalidade de boa qualidade consideradas em conjunto; ela incorpora a experiência de mais tábuas que qualquer das outras famílias; e descobriu-se que tábuas de mortalidade de Taiwan, do Japão e de Singapura se ajustam, assim como tábuas européias, ao modelo "Oeste". Então, existe algum fundamento lógico para se aceitar a "Oeste" como uma melhor *suposição* sobre o modelo predominante de mortalidade para qualquer população na ausência de evidência em contrário. Por outro lado, em níveis de mortalidade de moderados a altos (tais como os encontrados em populações africanas) as estimativas das taxas de natalidade e de mortalidade baseadas nas tábuas modelo "Norte" são pouco diferentes daquelas obtidas a partir da família "Oeste", e a família "Norte" de tábuas de mortalidade tem uma característica-mortalidade infantil baixa com relação à mortalidade de 1 a 5 anos — que observadores, freqüentemente, têm atribuído à população africana. Então as tábuas modelo "Norte" foram, algumas vezes, preferidas em lugar das modelos "Oeste" ao se fazer as estimativas para a África.

Brass construiu um sistema de tábuas modelo de mortalidade com dois parâmetros que tem um padrão básico, expressando certas características aparentemente gerais na experiência africana de mortalidade: a baixa mortalidade infantil com relação à mortalidade de 1 a 5 anos mencionada no parágrafo anterior, e taxas de mortalidade que aumentam rapidamente dos 20 aos 50 anos de idade. Seu padrão simples de mortalidade de dois parâmetros permite o uso de indícios sobre certas características da estrutura de óbitos por idades experimentadas por uma população particular. De fato, pode-se construir uma tábua modelo de Brass para reproduzir uma proporção pré-fixada de sobreviventes para cada uma de duas idades escolhidas arbitrariamente. Assim, se existirem indicações fidedignas dos níveis relativos de mortalidade nas primeiras idades e de mortalidade adulta, o sistema de Brass de tábuas modelo de mortalidade permite a seleção de uma tábua de mortalidade que incorpore tal relação.

A abordagem Coale-Demeny aceitaria como corretas as taxas de mortalidade com bases mais sólidas — por exemplo, estimativas das proporções de sobreviventes até as idades de 2 ou 3 anos, derivadas de informações sobre a sobrevivência entre os filhos tidos nascidos vivos — e então se escolheria a tábua modelo de mortalidade correspondente da família “Norte” ou “Oeste”. Tipicamente, as taxas de mortalidade adulta, numa tábua modelo selecionada dessa maneira, diferem das taxas de mortalidade adulta informadas, para o ano anterior, nos recenseamentos ou pesquisas demográficas. A abordagem Coale-Demeny aceita que a mortalidade adulta acompanharia a mortalidade nas primeiras idades estimada na experiência de populações com dados precisos, e não dá nenhuma importância às informações dos óbitos de adultos ocorridos no último ano, que por várias razões não parecem dignos de confiança. A abordagem de Brass também aceita as estimativas relativamente convincentes de mortalidade nas primeiras idades a partir de informações sobre os filhos tidos nascidos vivos, e então procura indícios (nas taxas informadas de mortalidade adulta) sobre a maneira da mortalidade variar com a idade.

As tábuas modelo de mortalidade e as populações estáveis de Coale-Demeny — Cálculo e uso

“Regional Model Life Tables and Stable Populations”¹⁸ contém centenas de páginas de tábuas; o leitor que deseje basear suas estimativas nas tábuas modelo deve recorrer a esse volume e pode consultar a descrição e as instruções na introdução. A discussão neste capítulo, portanto, se limita a um curto enunciado de como as tábuas foram construídas, e uma descrição de seu uso ao se fazer estimativas para as populações africanas.

Cada uma das quatro famílias das tábuas modelo de mortalidade expressa o modelo das taxas de mortalidade por idade encontrado num grupo de populações caracterizado pela homogeneidade do padrão de mortalidade. A homogeneidade foi detectada quando as taxas de mortalidade em todas as tábuas de mortalidade dignas de confiança foram comparadas com um conjunto preliminar de tábuas modelo de um único parâmetro, expressando a experiência mundial média. Um grupo de tábuas escandinavas de mortalidade apresentou uma tendência consistente em direção à baixa mortalidade na velhice, e também baixa mortalidade infantil quando comparada às taxas em idades de 1 a 4 anos; outro grupo de tábuas da Europa Central tinha um padrão característico que incluía altas taxas nas primeiras idades e na velhice; e um terceiro grupo do sul da Europa apresentou altas taxas no período de 1 a 4 anos e taxas baixas nas idades adultas tardias. Um grande grupo residual, abrangendo a experiência da maior parte da Europa Oci-

¹⁸ De COALE & DEMENY, *op cit*, p 47

dental, mais os Estados Unidos, o Canadá, a Austrália, a Nova Zelândia, Taiwan, e o Japão, não apresentou desvios característicos em relação ao padrão médio mundial de mortalidade por idade. As tábuas escandinavas foram usadas como base para as tábuas de mortalidade modelo Norte; as tábuas centro-européias foram a base para o modelo Leste; o sul da Europa para o modelo Sul; e as tábuas de mortalidade restante, relativamente próximas à experiência média, foram a base para o modelo Oeste. As correlações entre as taxas de mortalidade nas diferentes idades foram muito altas (quase sempre acima de 0,90, e usualmente acima de 0,95) no interior de cada família, e em cada família maior que nas quatro famílias consideradas em conjunto.

Cada conjunto de tábuas modelo foi calculado primeiro computando-se a regressão linear entre ${}_1q_0, {}_1q_1, {}_5q_5, \dots, {}_5q_{75}$ e e_{10}^o para homens e mulheres da família de tábuas de mortalidade selecionada, e depois admitindo-se que o indicador ($e_{10}^{o'}$) assumisse valores que gerassem tábuas modelo em diversos níveis, variando de uma esperança de vida ao nascer (para mulheres) desde 20 anos até uma esperança de vida maior que a experimentada por qualquer população africana. Em esperanças de vida mais elevadas foi feita uma transformação na regressão para logaritmo das taxas de mortalidade e e_{10}^o a fim de evitar absurdas taxas de mortalidade em certas idades. Mas esta característica do conjunto de tábuas não é relevante para a maioria das estimativas desse livro.

Em cada tábua modelo de mortalidade, foram calculadas 26 populações estáveis, com taxas de crescimento variando de $-0,010$ a $0,050$, e taxas brutas de reprodução variando de $0,800$ a $6,00$. Vários parâmetros — a taxa de natalidade, a taxa de mortalidade, a taxa de crescimento, as taxas bruta e líquida de reprodução, a idade média etc. — foram calculados e impressos para cada população estável.

Uma população estável tem a estrutura etária, as taxas de crescimento e muitas outras propriedades que caracterizariam uma população fechada submetida por muitos anos a tábuas especificadas de fecundidade e de mortalidade. Daí, se a fecundidade e a mortalidade não mudaram marcadamente na história recente de uma população, e se o impacto da migração sobre a composição etária foi pouco importante, a população terá as características essenciais de uma população estável. Além disso, se o padrão de mortalidade por idade se adapta aproximadamente a uma das tábuas modelo de mortalidade, as características da população em questão estarão muito próximas das características de uma das populações estáveis modelo.

Se não tiver havido nenhuma tendência recente na fecundidade e na mortalidade, a taxa de natalidade, a taxa de mortalidade, e uma estrutura etária uniforme podem ser encontradas numa população estável modelo que presumivelmente representa mais ou menos rigorosamente a população em questão. Se o padrão Oeste ou Norte da mortalidade pode ser aceito, o problema da estimativa se reduz ao de estabelecer (por interpolação) a população modelo estável que melhor se

ajuste à população dada, entre as 624 populações estáveis tabuladas segundo o sexo em cada família. O "melhor ajuste" é determinado através da comparação de certas características registradas ou estimadas da população real com os valores tabulados das populações estáveis.

Posto que a família de populações estáveis é gerada por variações no nível da mortalidade e na taxa de crescimento natural (determinadas conjuntamente por sua vez pela mortalidade e pela fecundidade), cada família de populações estáveis é um sistema de dois parâmetros. Em geral, então, duas características de uma população real são suficientes para estabelecer uma população estável dentro de uma das famílias. Entre as características usadas para selecionar uma população estável modelo estão as seguintes: as proporções de pessoas em certos grupos etários, notadamente abaixo de 5 anos e abaixo de 15 anos; a taxa intercensitária de crescimento; a proporção que se estima sobreviver até as idades de 2, 3, e 5 anos; e a taxa de mortalidade entre os não lactentes; ou pessoas que sobreviveram ao primeiro aniversário.

A escolha dos indícios através dos quais as populações estáveis foram selecionadas foi guiada (1) pela aparente qualidade de diferentes formas de dados sobre a população e (2) pela sensibilidade das estimativas a irregularidades nas suposições subjacentes ao processo de estimação. Por exemplo, há certas tendências quase universais para a declaração errada de idade nos recenseamentos e pesquisas em populações africanas a omissão das mulheres de 10-14 anos, e crescentes proporções com o aumento de idade entre 10-14 e 25-29 ou mesmo 30-34; declarações incompletas da proporção de homens de 15-19 anos, etc. Algumas das mulheres que faltam no período de 10-14 anos aparecem, em alguns recenseamentos e pesquisas, como meninas de idade abaixo de 10 anos, e freqüentemente a proporção da população abaixo de dez anos parece maior do que seria coerente com qualquer outra evidência sobre fecundidade e mortalidade. A seleção de uma população feminina estável que apresente a proporção observada de menores de 15 anos parece ser uma base moderada para estimativas da taxa de natalidade.

No levantamento demográfico do Congo, em 1956, os recenseadores foram instruídos para verificarem a idade das crianças menores pedindo as pessoas entrevistadas evidência documentada, na forma de certidões de nascimento ou registros nos folhetos de identificação dos pais. Tal verificação foi feita para mais ou menos 80% das crianças abaixo de 5 anos, e quase sempre a idade das outras crianças foi comparada com a idade verificada de um ou mais de seus irmãos. Há, portanto, uma forte razão para se confiar especialmente na idade informada das crianças mais jovens do Congo, e populações estáveis foram selecionadas através das proporções registradas de menores de 5 anos.

O exame da sensibilidade das estimativas a suposições duvidosas ou pelo menos não verificáveis levou a certas preferências entre os pos-

síveis métodos de seleção das populações estáveis modelo. Por exemplo, quando a população estável é usada para estimar a taxa de natalidade, se dá preferência à seleção por meio de proporções nos grupos etários mais jovens (isto é, em idades menores que 20 anos, se possível), combinada com níveis estimados de mortalidade infantil e nas primeiras idades. A razão para se preferir esta combinação pode ser vista no exame da relação entre a taxa de natalidade e a proporção em alguma idade jovem na população. A taxa média de natalidade nos cinco anos anteriores ao recenseamento poderia ser determinada por uma projeção retrospectiva da população de menores de 5 anos de idade para estimar-se quantos nascimentos foram necessários para proporcionar os sobreviventes encontrados nesse grupo etário, e através da projeção de toda a população, dois anos e meio atrás, para estimar-se o denominador da taxa de natalidade.

$$b = \frac{{}_5P_0 \left(\frac{l_0}{{}_5L_0} \right)}{Pe^{-2.5}} \quad (6)$$

Na equação (6), b é estimado como a população de menores de 5 anos, multiplicada por um fator de sobrevivência retrospectivo proveniente da tábua de mortalidade, dividido pela população total multiplicado por um fator que toma em conta 2,5 anos de crescimento. A taxa de natalidade de uma população estável está relacionada à proporção dos menores de 5 anos através da seguinte relação

$$b = \frac{\int_0^5 c(a) da}{\int_0^5 e^{-1a} p(a) da} \quad (7)$$

Já que

$$\frac{{}_5P_0}{P} = \int_0^5 c(a) da$$

e

$$\frac{{}_5L_0}{l_0} e^{-2.5} = \int_0^5 e^{-1a} p(a) da$$

as equações (6) e (7) são essencialmente idênticas. Estimar a taxa de natalidade através da projeção retrospectiva de menores de cinco anos usando uma tábua modelo de mortalidade escolhida para que tenha um valor conhecido de ${}_5L_0$ é equivalente à estimação através de uma população modelo estável baseada nesta tábua de mortalidade. Mas o único uso feito da hipótese de estabilidade deste fato está na determinação da taxa média de crescimento da população durante os 2,5 anos anteriores — não é uma fonte importante de imprecisão na estimação da taxa de natalidade. Além disso apenas valores ligeiramente

variáveis de ${}_5L_0$ são coerentes com um dado valor de l_2 ou l_3 . Assim, a taxa de natalidade numa população estável modelo identificada pela proporção da população de menores de 5 anos e a proporção de crianças sobreviventes à idade de 2 ou 3 anos é essencialmente a mesma que em qualquer população (seja estável ou não) com a mesma proporção de crianças menores e a mesma taxa de sobrevivência infantil (esteja a mortalidade em idades adultas adaptada à configuração modelo ou não). Este argumento favorece o argumento de que se utilize, tanto quanto possível as proporções em idades mais jovens e as taxas de sobrevivência nas primeiras idades ao fazermos as estimativas da taxa de natalidade; mas quando o método de Brass para estimar a mortalidade não puder ser usado, a mortalidade infantil e nas primeiras idades pode ficar mais difícil de ser determinada ou estimada do que a mortalidade em outras idades e em virtude da informação errada na declaração de idade e algumas vezes da omissão diferencial, a proporção de crianças abaixo de alguma idade — 1, 5, ou inclusive 10 anos — pode ser registrada de forma um pouco confiável. Nessas ocasiões o método de população estável na estimação produz aproximações menos fidedignas da taxa de natalidade.

As tábuas modelo de mortalidade e as populações estáveis de Brass

O método de Brass de construção de tábuas modelo de mortalidade consiste em aplicar a função de sobrevivência de uma tábua de mortalidade escolhida como “padrão” a chamada transformação logito, e então considerar as tábuas de mortalidade geradas supondo que seus logitos estão linearmente relacionados ao logito da tábua padrão. Dessa maneira um conjunto de tábuas modelo de mortalidade com dois parâmetros pode ser construído.

A função logito é a seguinte:

$$\text{logit}(x) = \frac{1}{2} \log_e \frac{1-x}{x} \quad (8)$$

Brass substitui x na equação por $p(a)$ (ou l_a/l_0 em notação de tábua de mortalidade). Ele seleciona então uma tábua padrão $p_s(a)$, e constrói suas tábuas modelo de mortalidade atribuindo diferentes valores a α e β em

$$\text{logit } p(a) = \alpha + \beta \text{ logit } p_s(a) \quad (9)$$

Esta transformação implica que:

$$\frac{q(a)}{p(a)} = A \left(\frac{q_s(a)}{p_s(a)} \right)^\beta \quad (10)$$

onde $A = e^{2\alpha}$

A natureza da transformação logito linear de uma tábua padrão de mortalidade

A transformação logito pode gerar uma função de sobrevivência $p(a)$ que passa por valores arbitrariamente pré-estabelecidos em qualquer de duas idades a_1 e a_2 , pela seleção de α e β ¹⁹.

A tábua de mortalidade assim gerada tem os valores pré-estabelecidos de $p(a)$ nas idades selecionadas e compartilha da forma, num sentido geral, da tábua “padrão” de mortalidade.

O Gráfico 3.4 ilustra o efeito de diferentes valores de α e β na tábua padrão de mortalidade que Brass construiu para a África. As curvas de sobrevivência incluem:

- a) o padrão ($\alpha = 0, \beta = 1,00$);
- b) uma curva que difere do padrão por ter um β de 1,20 ao invés de 1,00;
- c) uma curva que tem o mesmo valor de $p(2)$ que o padrão, mas um β de 1,20, e
- d) uma curva que difere do padrão por ter um valor de α de 0,30 ao invés de zero.

Note que esta curva com $\alpha = 0$ e $\beta = 1,20$ intercepta o padrão numa idade onde $p_s(a) = 0,50$. Esta é uma relação geral: todas as curvas com o mesmo α e diferentes β interceptam em \bar{a} determinado por $p_s(\bar{a}) = 0,500$ porque $\text{logit } 0,5 = 0$. Portanto a variação de β produz uma sub-família de curvas de sobrevivência, fixadas na idade 0 e ω , e com um ponto duplo em \bar{a} (mais ou menos 51 anos de idade na tábua padrão). Em idades abaixo desse ponto comum ou intersecção a tábua de mortalidade baseada no β maior tem valores superiores de $p(a)$, e em idades acima deste ponto aplica-se a relação oposta. Como resultado a área sob essas duas curvas (que em cada caso é a esperança de vida ao nascer) é praticamente a mesma²⁰.

¹⁹ Dados quaisquer $p(a_1)$ e $p(a_2)$ pré-estabelecidos, obtemos duas equações:

$$\alpha + \beta \text{ logit } p_s(a_1) = \text{logit } p(a_1)$$

$$\alpha + \beta \text{ logit } p_s(a_2) = \text{logit } p(a_2)$$

Os valores conhecidos de $p_s(a_1)$ e de $p_s(a_2)$ em conjunto com valores pré-estabelecidos de $p(a_1)$ e de $p(a_2)$ proporcionam uma determinação única de α e β e portanto uma tábua de mortalidade completa

$$\text{logit } p(a) = \alpha + \beta \text{ logit } p_s(a)$$

²⁰ Para a tábua padrão de mortalidade, $e_0^o = 43,6$, e para $\beta = 1,2$, $e_0^o = 44,1$. Há um tipo de simetria entre as curvas com diferentes valores de β , e $\alpha = 0$. A diferença em p para cada idade jovem tal que $p_s(a')$ seja maior que 0,5 é correspondida por uma diferença igual (e oposta) em p para a idade mais velha a'' tal que $p_s(a'') = 1 - p_s(a')$. Assim, quando $p_s(a) = 0,50$ ocorre, perto do meio do intervalo zero a ω as tábuas de mortalidade com $\alpha = 0$ e diferentes valores de β têm aproximadamente o mesmo e_0^o .

PROPORÇÃO DE SOBREVIVENTES ATÉ A IDADE a , $p(a)$, OBTIDO
 A PARTIR DA TRANSFORMAÇÃO $p(a) = \alpha + \beta \logit p_1(a)$ PARA
 VÁRIOS VALORES DE α E β

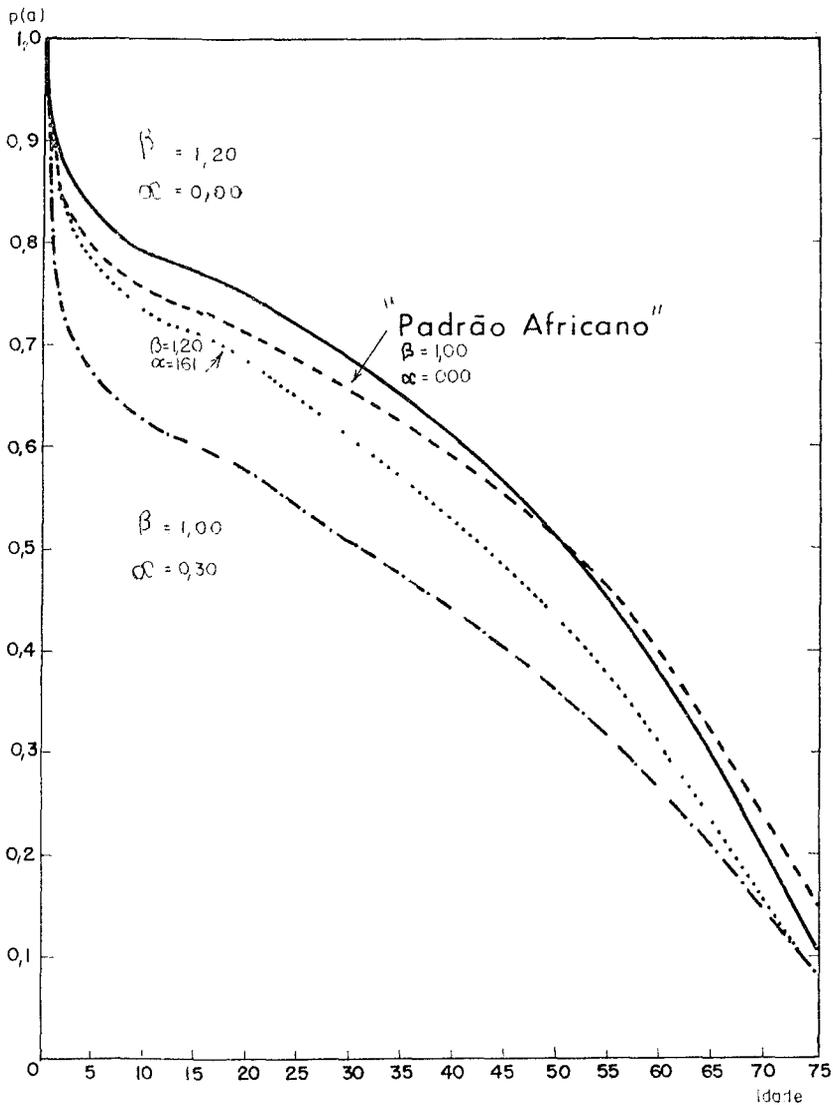


Gráfico 3.4

Todavia essa quase constância de e_0° para diferentes valores de β não é uma relação geral, mas é o resultado da seleção de uma tábua padrão de mortalidade com $p(51) = 0,50$. Se uma tábua padrão de mortalidade apresentando maior mortalidade fosse escolhida, as tábuas com o mesmo α se interceptariam numa idade mais jovem (digamos 25 anos) e valores maiores de β diminuiriam a esperança de vida ao nascer.

Note que a transformação logito permite a aceitação de uma estimativa confiável de sobrevivência nas primeiras idades — por exemplo, uma estimativa de $p(2)$ e a seleção de uma estimativa total de mortalidade adulta, através da escolha de $p(30)$ ou $p(50)$, dado o valor aceito de $p(2)$. A transformação da tábua padrão (com α e β escolhidos dessa maneira) provê então uma completa tábua modelo de mortalidade.

A importância da tábua padrão de mortalidade na transformação logito linear

A transformação logito poderá gerar tábuas de mortalidade tendo qualquer nível pré-estabelecido de mortalidade infantil — expressado, por exemplo, por um valor de $q(2)$ — e de mortalidade adulta — expressado, por exemplo, por ${}_48q_2$, ou $p(50)/p(2)$. Até que ponto essas tábuas de mortalidade se adaptam ao modelo padrão de mortalidade por idade observado em populações empíricas? As famílias de tábuas modelo de Coale-Demeny proporcionam um interessante teste. Cada família destas tábuas expressa o modelo etário padrão de mortalidade por idade, particularmente de mortalidade encontrado na experiência bem informada de um grupo de populações. Os grupos de tábuas de mortalidade que servem de fundamento a cada família foram reunidos, por causa da perceptível similaridade de padrões por idade. O Gráfico 3.5 mostra o resultado típico da tentativa de se expressar uma tábua modelo em uma das famílias

- a) como transformação logito linear de outra tábua na mesma família, e
- b) como transformação logito de uma tábua de família diferente.

Note que o logito $p(a)$ para a tábua de mortalidade modelo Sul (sexo feminino) com $e_0^\circ = 20$ tem uma relação praticamente linear com o logito $p(a)$ para a tábua modelo Sul onde $e_0^\circ = 50$ anos, mas que a relação com o logito $p(a)$ para a tábua modelo Oeste com $e_0^\circ = 50$ não é de maneira nenhuma linear. O Gráfico 3.6 apresenta o resultado da tentativa de representar a tábua modelo Sul com $e_0^\circ = 30$ anos através de uma transformação logito da tábua modelo Sul com $e_0^\circ = 50$, e da tábua modelo Oeste com $e_0^\circ = 50$. Os cálculos do mesmo tipo feitos com as outras famílias de tábuas modelo de Coale-Demeny levaram a

LOGITO $p(a)$ REPRESENTADO GRAFICAMENTE CONTRA O LOGITO $p_s(a)$, QUANDO $p(a)$ É A TÁBUA DE VIDA MODELO SUL, SEXO FEMININO, $e_0^o = 20$ ANOS, E $p_s(a)$ É ALTERNATIVAMENTE A TÁBUA MODELO OESTE E MODELO SUL $e_0^o = 50$ ANOS

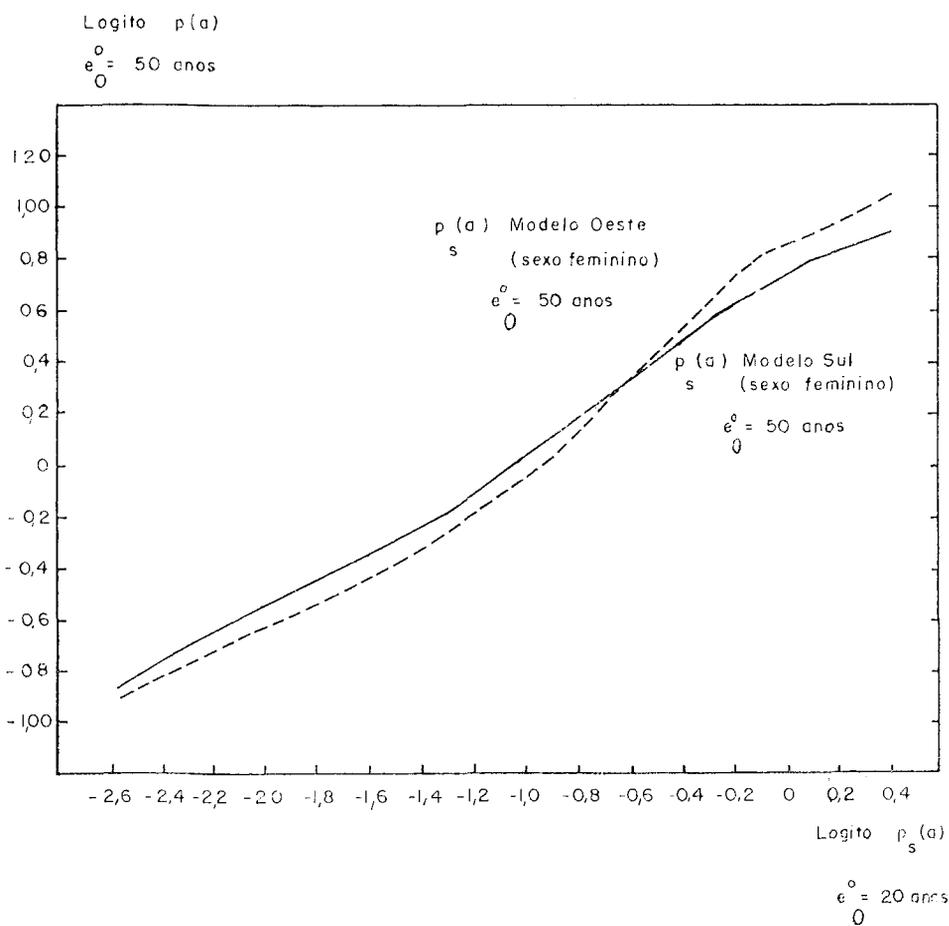


Gráfico 3.5

MODELO SUL DE MORTALIDADE SEXO FEMININO $e_0^o = 30$ ANOS,
 COMPARADA COM AS TÁBUAS DE MORTALIDADE OBTIDAS POR
 TRANSFORMAÇÃO LOGITO LINEAR DAS TÁBUAS MODELO SUL
 E OESTE, $e_0^o = 50$ ANOS

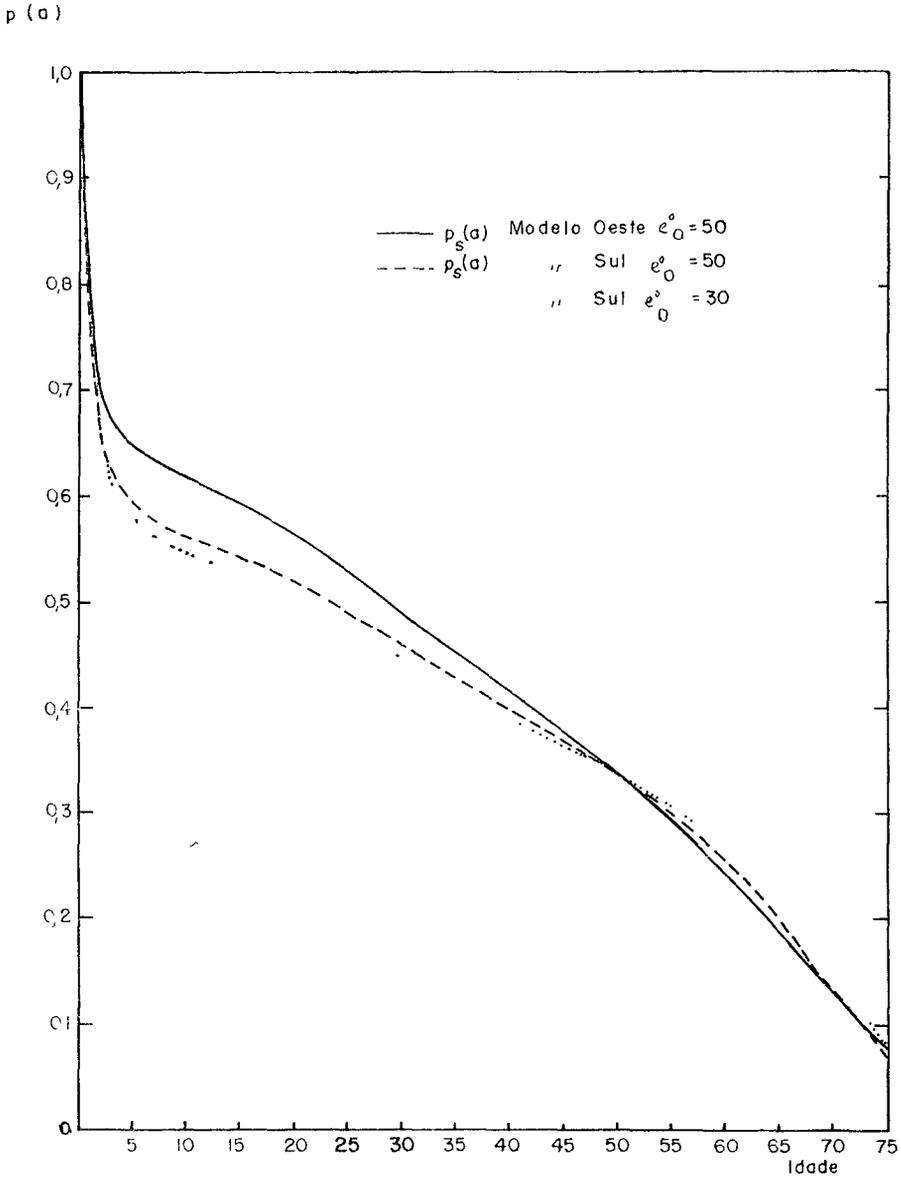


Gráfico 3.6

um resultado similar: as transformações logito lineares reproduzem as tábuas modelo de mortalidade muito aproximadamente *dentro* de uma família, mas não tão bem entre famílias. A Tabela 3.10 apresenta os valores de α e β necessários para se transformar a tábua feminina com $e_0^\circ = 50$ nas tábuas com e_0° de 20, 30, 40, 60 e 70 em cada família. Note que em todas as famílias, exceto a Sul (onde β é bastante grande em altos níveis de mortalidade), os parâmetros necessários para se mudar de um nível de mortalidade a outro são quase idênticos.

TABELA 3.10

VALORES DE α e β EM LOGITO $p(a) = \alpha + \beta \text{ LOGITO } p_s(a)$ PARA AS TÁBUAS MODELO NORTE, SUL, LESTE E OESTE DE MORTALIDADE (SEXO FEMININO VÁRIOS NÍVEIS DE e_0° , TÁBUA DE MORTALIDADE COM $e_0^\circ = 50$ QUE SERVE COMO $p_s(a)$)

ESPECIFICAÇÃO	$e_0^\circ = 20$	$e_0^\circ = 30$	$e_0^\circ = 40$	$e_0^\circ = 50$	$e_0^\circ = 60$	$e_0^\circ = 70$	
Norte.	α ..	1,008	0,612	0,295	0,000	-0,316	-0,734
	β ..	1,296	1,154	1,064	1,000	0,950	0,900
Oeste.	α ..	1,036	0,632	0,307	0,000	-0,339	-0,819
	β ..	1,306	1,161	1,068	1,000	0,944	0,893
Sul...	α ...	1,057	0,650	0,318	0,000	-0,356	-0,803
	β ...	1,533	1,318	1,155	1,000	0,810	0,599
Leste..	α ..	1,000	0,618	0,302	0,000	0,338	0,813
	β ..	1,299	1,165	1,073	1,000	0,929	0,852

Estes exemplos mostram que a seleção de uma tábua típica de mortalidade afeta a forma detalhada da curva de sobrevivência numa tábua de mortalidade ajustada à experiência estimada de mortalidade de uma população através de uma transformação logito linear. A tábua padrão de mortalidade a ser usada na África, em outras palavras, deveria incorporar teoricamente características típicas (se tais características existissem) de padrões africanos de mortalidade por idade.

Brass desenvolveu a tábua padrão africano de mortalidade dado na Tabela 3.11 para refletir características que são pelo menos aparentemente comuns na África. Até a idade de 10 ou 20 anos a curva $p(a)$ nesta tábua assemelha-se estreitamente a uma tábua modelo Norte com um e_0° de mais ou menos 46 anos; mas após a idade de 20 anos a curva "padrão" de sobrevivência cai muito mais acentuadamente que uma tábua modelo Norte com a mesma mortalidade nas primeiras idades, e na verdade mais depressa que uma tábua modelo Sul, Leste ou Oeste com a mesma mortalidade nas primeiras idades. Este aspecto

TABELA 3.11

PROPORÇÃO DE SOBREVIVENTES $p_s(a)$, E OS CORRESPONDENTES LOGITOS EM IDADES ESPECIFICADAS DA TABUA PADRÃO AFRICANA DE MORTALIDADE

IDADE EXATA EM ANOS a	$P_s(a)$	LOGITO $P_s(a)$
1 .	0,8802	0,9970
2 .	0,8335	0,8052
3 .	0,8101	0,7252
4 .	0,7964	0,6819
5 .	0,7863	0,6615
10 .	0,7502	0,5498
15 .	0,7362	0,5131
20 .	0,7130	0,4551
25 .	0,6826	0,3829
30 .	0,6525	0,3150
35 .	0,6223	0,2496
40 .	0,5898	0,1817
45 .	0,5535	0,1073
50 .	0,5106	0,0212
55 .	0,4585	-0,0832
60 .	0,3965	-0,2100
65 .	0,3210	-0,3746
70 .	0,2380	-0,5818
75 .	0,1500	-0,8673
80 .	0,0760	-1,2490
85 .	0,0310	-1,7211

combina com uma característica geral de mortalidade por idade como se informa nos levantamentos africanos para o ano pré-censitário, e é coerente com uma tendência das distribuições etárias africanas informadas de decrescer rapidamente a partir dos trinta e dos cinquenta anos. É uma questão a ser determinada se estas características comuns são uma tendenciosidade típica nos dados africanos ou uma característica típica da mortalidade africana.

Processo para uso da transformação logito da tábua padrão africana de mortalidade

O processo da tábua modelo de mortalidade de Brass é quase sempre aplicado a dados de mortalidade derivados de um levantamento demográfico — dados que incluem o número informado de filhos tidos nascidos vivos e o número de sobreviventes, classificados por idade da mãe, e o número informado de óbitos que ocorrem no ano anterior ao levantamento, classificados por idade, e comumente por sexo. Com pequeno erro (em vista da qualidade geral das informações) a população em risco pode ser considerada como a população enumerada no levantamento, e taxas específicas de mortalidade por idade podem ser calculadas para grupos etários quinquenais ou decenais. Estas taxas específicas de mortalidade por idade (valores de m) podem então ser conver-

tidas em proporções de óbitos ocorridos da tábua de mortalidade (valores de q_x através de técnicas padrões, por exemplo, o método Reed-Merrell²¹). Estas duas fontes de informação dão dois conjuntos alternativos da proporção de sobreviventes para idades acima de 5 anos ou talvez 10 anos. Os valores de $p(2)$ e $p(3)$ são as inferências mais fidedignas a partir dos dados sobre os sobreviventes entre os filhos tidos nascidos vivos, e $p(5)$ pode ser uma indicação válida dos riscos atuais de mortalidade até 5 anos de idade, se não tiver havido nenhuma tendência marcante de mortalidade e nenhuma tendência para a omissão de crianças mortas entre aquelas que se informou terem nascido nos últimos 5 ou 10 anos. Em idades maiores (10, 15 anos, etc), as inferências sobre a mortalidade atual a partir de dados sobre os filhos tidos nascidos vivos são de confiança cada vez menor.

Uma comparação dos $p(2)$, $p(3)$ e talvez $p(5)$ estimados, por um lado, a partir das informações retrospectivas de filhos tidos nascidos vivos e filhos sobreviventes, e, por outro lado, a partir de informações diretas sobre óbitos ocorridos entre as crianças no ano anterior ao levantamento, mostra, em geral, uma mortalidade estimada muito maior a partir dos dados retrospectivos. Se estivermos dispostos a supor que há uma informação incompleta da mortalidade no mesmo grau em todas as idades (por exemplo, em virtude de erro no "período de referência"), a razão de $q(3)$ a partir dos dados retrospectivos e $q(3)$ derivados dos óbitos ocorridos informados para o ano anterior ao levantamento poderia ser usada como um fator de correção para corrigir as taxas informadas de óbitos em idades acima de 5 anos. Mas, a análise revela rapidamente que (em contraste com a fecundidade informada) a estrutura etária da mortalidade informada está quase sempre claramente distorcida. Há uma nítida tendência para a informação incompleta da mortalidade entre alguns grupos por idade e sexo, mais que entre outros. Portanto não se justifica o uso de um simples ajustamento do fator de correção da mortalidade para idades acima da infância. Brass sugere que se faça duas curvas $p(a)$, cada uma começando com estimativas de $p(2)$, $p(3)$ e $p(5)$ baseadas na sobrevivência informada dos filhos tidos nascidos vivos. A primeira curva $p(a)$, simplesmente, incorpora os valores de ${}_m q_x$ obtidos da mortalidade informada, e a segunda incorpora valores baseados na mortalidade informada corrigida pelo fator de correção acima mencionado. Os logitos dessas curvas alternativas de $p(a)$ são colocados num eixo e em outro eixo o logito de $p_s(a)$ (o padrão africano). Se algum dos dois conjuntos de pontos cai aproximadamente numa linha reta, com uma inclinação entre 0,80 e 1,20, a linha pode ser tomada como base para a composição de uma tábua modelo de mortalidade. Freqüentemente uma reta intermediária entre os dois conjuntos de pontos acima de 5 anos de idade se ajusta melhor aos valores de $p(a)$ até a idade de 3 anos. Brass sugere que, quando a mortalidade adulta informada pareça quase totalmente in-

²¹ REED, Lowell J. & MERRELL, Margaret A short method for constructing an abridged life table. *American Journal of Hygiene*, 30 (2) set. 1939

consistente ou não digna de confiança, seja construída uma tábua modelo de mortalidade que se ajuste a $p(2)$ e $p(3)$ a partir dos dados retrospectivos, que tenha um valor β unitário na transformação logito da tábua padrão africana.

Assim as tábuas modelo de mortalidade de Brass são uma família de um só parâmetro — formada pela variação de α em $p(\alpha) = \alpha + p_s(\alpha)$ — sempre que houver evidência insuficiente sobre a mortalidade adulta para justificar outra escolha de β que não $\beta = 1$. O processo Coale-Demeny seria escolher uma tábua modelo Oeste ou Norte que mais se aproximasse de $p(2)$ e $p(3)$ baseados na sobrevivência entre filhos tidos nascidos vivos, e aceitar as taxas de mortalidade em todas as idades acima da infância da tábua modelo escolhida. O processo de Brass também inclui a aceitação da mortalidade infantil estimada a partir de dados retrospectivos; na ausência de informação convincente sobre a mortalidade adulta, Brass usa um conjunto de um só parâmetro de tábuas modelo derivado de seu padrão africano; mas quando as taxas de mortalidade adulta formam um modelo coerente de “logitos” com β não igual a 1,00, pode-se compor uma tábua modelo modificada dessa família.

Verificação das estimativas da mortalidade, fecundidade e estrutura etária

A estrutura etária de uma população fechada é determinada pela história recente de sua fecundidade e mortalidade. Se os padrões por idade de ambas têm sido aproximadamente constantes, a população tem a estrutura etária estável. A principal determinante das características gerais da distribuição etária — a idade média, a proporção abaixo de 15 anos, e medidas semelhantes — é a história da fecundidade, de modo que a distribuição etária de uma população de fecundidade constante assemelha-se bastante à estável, mesmo se a mortalidade não tem sido constante. A semelhança é especialmente estreita quando o comportamento da mortalidade tem sido de variação gradual e contínua. De fato, mesmo as variações de fecundidade, se tiverem ritmo lento, dão origem a uma composição etária não muito distante da população estável implícita nos quadros de fecundidade e mortalidade dominantes em qualquer momento.

Sob as circunstâncias em que vive a maioria das populações africanas, a expectativa normal é que a fecundidade tenha sido mais ou menos constante no passado recente, em níveis atuais, porque os costumes que poderiam afetar a fecundidade nas populações africanas — idade ao casar, costumes relacionados com as uniões consensuais fora do casamento, tabus sobre as relações sexuais, práticas populares de aborto e de anticoncepção — provavelmente não se alteraram na ausência de mudanças sociais ainda não evidentes na África. Uma exceção a esta generalização pode ocorrer em populações onde a fecundi-

dade tenha sido afetada por esterilidade patológica resultante de moléstias transmissíveis. A disseminação de tais doenças, e seu controle bem sucedido, podem ter causado importantes diminuições ou aumentos na fecundidade de algumas populações africanas. Na verdade, a fecundidade de algumas populações — especialmente em alguns distritos do Congo — está abaixo da reposição de acordo com os atuais níveis de mortalidade estimados, e teria estado muito abaixo da reposição sob os riscos de mortalidade, no começo desse século ou no final do século XIX. Portanto é pouco plausível que a atual fecundidade baixa de algumas populações pudesse ter persistido durante quase um século, e impossível que pudesse ter se estendido até o passado remoto, a menos que estejamos preparados para aceitar populações muitas vezes maiores do que hoje em dia há um ou dois séculos atrás. Então somos levados a suspeitar que, onde a fecundidade for muito baixa, o declínio pode ter sido recente o bastante para deixar traços na forma da atual distribuição etária.

Também é fato que uma população com uma história de mortalidade bruscamente variante tenha uma distribuição etária perceptivelmente diferente de uma com a mesma mortalidade atual e uma história sem variações, isto é, uma população estável. De acordo com nossas estimativas, muitas populações africanas têm níveis de mortalidade que produzem uma rápida taxa de crescimento — de 2 a 3 por cento ao ano. Se estas taxas tivessem se mantido por muito tempo, a população da África teria sido apenas um décimo ou menos de seu tamanho atual, há um século, o que é contra a impressão de todos os estudiosos de história africana. Ainda assim, parece provável que muitas populações tenham experimentado um declínio na mortalidade suficientemente para ter deixado uma marca — neste caso relativamente leve — sobre a atual distribuição etária.

A despeito destas circunstâncias que fariam a distribuição etária diferir da estável, pode-se esperar que a forma geral da estrutura etária real se assemelhe a estrutura por idades da população estável implicada pela atual fecundidade e atual mortalidade. Na verdade é seguramente possível se considerar que a distribuição etária estável está mais perto da distribuição etária real que a distribuição etária tabulada em um recenseamento ou levantamento demográfico de uma população africana, já que as distribuições etárias informadas se baseiam em estimativas imperfeitas da idade dos entrevistados que usualmente têm pouco conhecimento da idade ou data do nascimento.

Resulta da provável semelhança geral da verdadeira distribuição etária com a estável, que uma comparação da população estável inerente nos quadros estimados de fecundidade e mortalidade com a distribuição etária informada comprovará sua consistência aproximada. É importante notar quais aspectos da mortalidade e da fecundidade são testados através desta comparação, quando se aceita a hipótese da estabilidade da distribuição etária. Brass encontrou que as populações estáveis formadas pela combinação de várias das tábuas modelo de mor-

talidade de Brass, e várias taxas intrínsecas de crescimento têm praticamente as mesmas proporções acumuladas até as idades 5, 10, 15, 25, 35, 45, 55 e 65 anos se o produto da taxa de natalidade e da proporção de sobreviventes até a idade de 2 anos [$bp(2)$] for o mesmo — ainda quando exista um amplo intervalo de valores de b , e taxas de crescimento e de mortalidade bastante diferentes. As comparações das proporções acumuladas nas populações modelo estáveis regionais de Coale e Demeny também mostram que as distribuições etárias com o mesmo $bp(2)$ não são muito diferentes dentro da mesma família de tábuas modelo. Portanto, se uma comparação da população estável resultante da fecundidade e mortalidade estimadas com a população informada indica uma concordância satisfatória, o que se confirmou foi a consistência da distribuição etária com a taxa de natalidade estimada e o nível estimado de mortalidade nas primeiras idades. Já que vários níveis de mortalidade adulta produziram mais ou menos a mesma estrutura etária geral — por que o efeito da mortalidade adulta maior sobre $p(a)$ seria compensado por seu efeito sobre r , a taxa intrínseca de crescimento — a verificação da distribuição etária não dá maiores informações sobre a mortalidade estimada acima dos dois anos de idade. A distribuição etária informada, especialmente tendo em vista suas distorções típicas, não fornece nenhuma orientação útil na seleção de uma família de tábuas modelo de mortalidade entre as variantes regionais de Coale-Demeny, ou na determinação da mais apropriada entre as várias combinações de α e β (na transformação logito de Brass) que produzem os $p(2)$ e $p(3)$ estimados. Também a distribuição etária não proporciona uma prova satisfatória das estimativas da taxa bruta de mortalidade ou da taxa de crescimento natural.

O processo de Brass para verificação das distribuições etárias estáveis e informadas

O método de Brass da estimativa de fecundidade produz uma tábua aproximada de fecundidade por idades específicas e a combinação de seu método de estimativa da mortalidade nas primeiras idades e seu sistema de tábua modelo de mortalidade, produz um tábua aproximada de sobrevivência por idades. Essas duas podem ser combinadas por processos padrões de cálculo para se obter uma população estável²². Uma comparação das proporções em cada intervalo quinquenal de idade nessa população estável e nas populações enumeradas nunca dariam na África, a impressão de consistência uniforme, por serem as distribuições etárias registradas de formas irregulares. Uma possibilidade consiste em suavizar a distribuição registrada através de alguma fórmula de ajustamento; mas o forte ajuste necessário seria provavelmente insatisfatório nos extremos da distribuição etária, ou envolveria suposi-

²² GLASS, D. V. *Population, Policies and Movements in Europe*. Oxford University Press, 1940 p. 405-15

ções arbitrárias. Outro método é comparar as duas distribuições etárias após a acumulação até diferentes idades. Esse método tem a vantagem de eliminar automaticamente o efeito de todos os erros de declaração de idade exceto aqueles que deslocam a idade de um entrevistado para além do limite de referência. Assim a proporção registrada até a idade de 20 anos é afetada apenas pelo deslocamento líquido de pessoas do intervalo abaixo de 20 anos para o intervalo acima de 20 anos, ou vice-versa. Até que se adquira experiência na interpretação das comparações das duas distribuições etárias acumuladas, observando-se uma tábua ou um gráfico de proporções acumuladas, é difícil julgar qual constitui um ajuste satisfatório. Uma comparação mais fácil de ser entendida é a seguinte: a diferença nas proporções acumuladas até cada idade nas distribuições etárias informadas e estáveis é convertida numa declaração do número incremental de anos de idade necessário para igualar as proporções. Por exemplo, uma proporção informada de 40 por cento abaixo de 10 anos de idade, comparada a 38 por cento na população estável baseada na fecundidade e na mortalidade estimadas, se explicaria se as crianças até a idade 10,8 anos em lugar de 10,0 anos em média fossem informadas como menores de 10 anos. No Gráfico 3.7 estão apresentadas as diferenças de idade necessárias para igualar a população estável calculada para a República da Guiné, em 1955, com a população informada, juntamente com as diferenças de idade necessárias para se converter a estável calculada em estáveis que se ajustariam aos pontos mais afastados (superior e inferior) da distribuição acumulada observada. Se as áreas positivas e negativas da curva central se acham aproximadamente equilibradas, pode-se dizer que a distribuição etária informada é consistente com os níveis estimados de fecundidade e de mortalidade nas primeiras idades; e a distância entre as curvas superior e a inferior indica a medida em que o erro na declaração da idade produziu inconsistência em diferentes pontos da distribuição acumulada. Na interpretação do Gráfico 3.7, é essencial ter em mente que a população estável não foi escolhida para se ajustar à distribuição etária, mas foi baseada em estimativas de fecundidade e mortalidade que são completamente independentes das distribuições etárias.

A conversão de uma diferença na distribuição etária acumulada em uma diferença de idade é feita facilmente pelo seguinte processo:

- a) seja $C(a)$ a proporção acumulada na distribuição etária estável até a idade a , e $c(a)$ da proporção desde a idade a até $a + da$;
- b) então $dC(a)/da = c(a)$, e qualquer discrepância $\Delta C(a)$, entre as distribuições informadas e estável pode ser convertida numa diferença de idade através da relação aproximada

$$\Delta a = \Delta C(a) / c(a)$$

$c(a)$ é facilmente calculada a partir das fórmulas padrão da teoria de populações estáveis.

REPÚBLICA DA GUINÉ 1955. DIFERENÇAS DE IDADE REQUERIDAS PARA FAZER COM QUE A DISTRIBUIÇÃO ETÁRIA ACUMULADA ESTÁVEL CONCORDE COM A DISTRIBUIÇÃO REGISTRADA, E COM AS DISTRIBUIÇÕES ESTÁVEIS ALTERNATIVAS QUE SE AJUSTEM AOS PONTOS EXTREMOS DA DISTRIBUIÇÃO REGISTRADA

DIFERENÇA DE IDADE EM ANOS

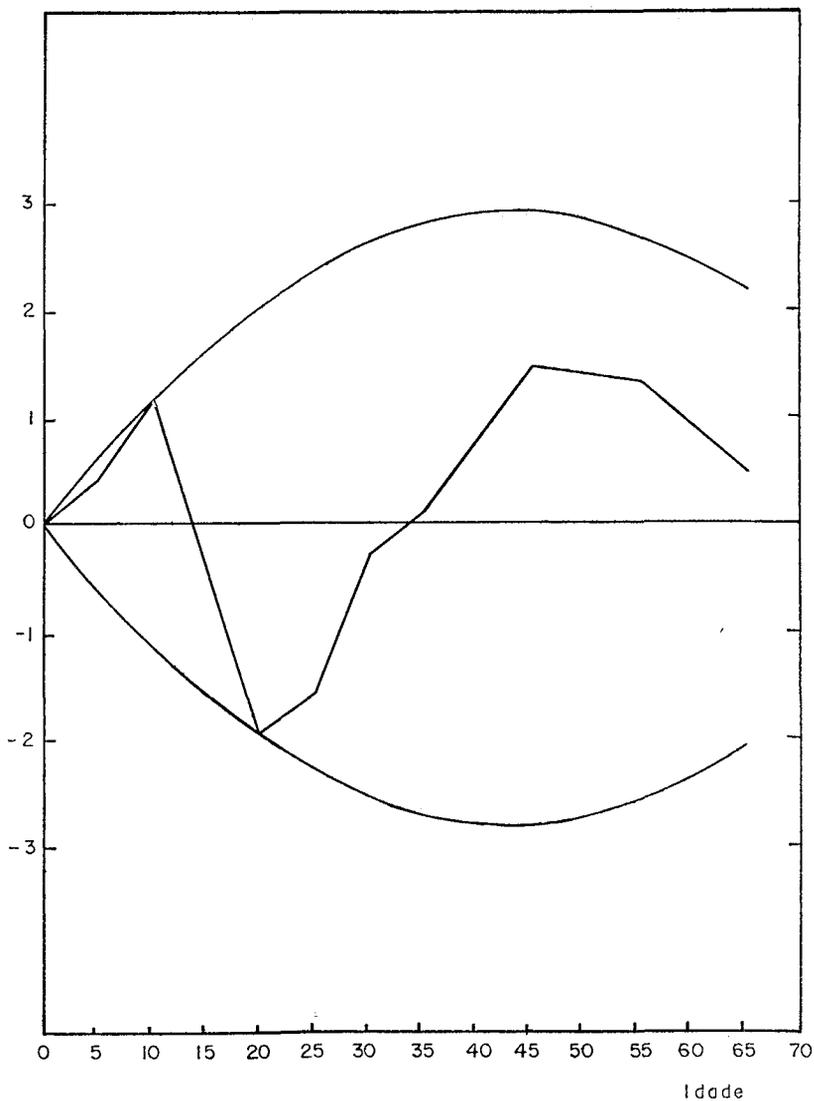


Gráfico 3.7

APÊNDICE A

NOTA SOBRE O MÉTODO BRASS DE ESTIMAÇÃO DA FECUNDIDADE

William Brass

As funções polinomiais são freqüentemente usadas para ajustar observações em virtude da facilidade com que podem ser manejadas matematicamente. Mas o uso de um polinômio geral para descrever a distribuição de fecundidade específica é incômodo porque necessita de muitos termos, e portanto são necessários muitos parâmetros para se obter um bom ajuste para as observações. O número de parâmetros desconhecidos pode ser reduzido pela imposição de restrições, e a seguinte função é satisfatória para as aplicações desenvolvidas neste estudo:

$$f(a) = C(a - s)(s + 33 - a)^2$$
$$s \leq a \leq s + 33$$

onde $f(a)$ é a taxa específica de fecundidade de mulheres com a anos de idade, s é a idade do começo do período reprodutivo, e C é uma constante que varia com o nível da fecundidade; $f(a)$ é considerada nula quando a está fora do intervalo s a $s + 33$ anos.

A função se restringe a zero em s e em $s + 33$. Entre esses limites a forma é aproximadamente a das distribuições empíricas com um aumento acentuado das taxas seguintes à idade s , um valor máximo e então uma queda gradativa até o final do período reprodutivo; a média e a moda estão mais próximas do limite inferior que do superior. A amplitude de variação de 33 anos da função foi escolhida para tornar a variância (43,6) próxima do valor médio das distribuições observadas.

Os dois parâmetros que podem ser variados quando o modelo está ajustado a dados para uma população são C e s . O valor de C , fixado pelo nível de fecundidade, não entra nas aplicações aqui consideradas, que dependem apenas da forma da curva, isto é, das relações entre suas partes. À medida em que s , o limite inferior do período reprodutivo, é aumentado, a curva se desloca de forma correspondente para cima. Particularmente, a média da função é $s + 13,2$; para s igual a 15 anos a média de 28,2 está próxima da média observada nas distribuições registradas, e uma variação de s desde 13 a 18 abrange o intervalo de variação apresentado nos dados observados para quase todas as populações. As distribuições relativas de fecundidade específica a partir do modelo, por grupos etários quinquenais, e para um conjunto de valores

de s e de correspondentes razões f_2/f_1 e \bar{m} estão ilustradas na Tabela 3.A.1. No cálculo das razões f_2/f_1 supos-se que (de acordo com a prática comum) a fecundidade das mulheres abaixo de 15 anos, apesar de apresentada separadamente na tabela, foi incluída na medição de f_1 para o grupo etário de 15-20 anos.

TABELA 3.A.1

TAXA DE FECUNDIDADE MODELO POR IDADES ESPECÍFICAS, GRUPOS QÜINQUENAIIS [$f(a) = C(a - s) (s + 33 - a)^2$; C ESCOLHIDO DE MODO QUE A FECUNDIDADE TOTAL SEJA 5,00]

LIMITES EXATOS DO INTERVALO ETÁRIO	TAXAS DE FECUNDIDADE						
10 — 15	0,020	0,005	—	—	—	—	—
15 — 20	0,179	0,148	0,111	0,075	0,044	0,020	
20 — 25	0,262	0,254	0,242	0,226	0,205	0,179	
25 — 30	0,249	0,257	0,263	0,266	0,265	0,262	
30 — 35	0,179	0,196	0,211	0,225	0,238	0,249	
35 — 40	0,090	0,108	0,126	0,144	0,162	0,179	
40 — 45	0,021	0,031	0,044	0,058	0,074	0,090	
45 — 50	—	0,001	0,003	0,006	0,012	0,021	
f_1/f_2^*	0,764	0,605	0,460	0,330	0,213	0,113	
\bar{m} (anos)	26,2	27,2	28,2	29,2	30,2	31,2	

* Computadas a partir da fecundidade específica para mais decimais do que se mostra na Tabela

A fecundidade acumulada até a idade exata a , $F(a)$, é obtida como a integral de $f(x)$ de s até a . O valor f para qualquer grupo etário (digamos c até d) é $F(d)$. A integração de $F(x)$ de c a d e a divisão por $(d - c)$ dá o número médio correspondente de filhos tidos nascidos vivos por mulher no grupo. As integrações e cálculos são facilmente feitos em virtude da forma polinomial simples de $f(x)$. Dessa maneira, as medidas f e F foram computadas para os grupos quinquenais padrão para uma série de posições de $f(x)$, com s e \bar{m} espaçados em intervalos de um ano, cobrindo a amplitude de variação para as distribuições observadas. A partir dessas medidas foram obtidos os fatores multiplicadores k necessários para se calcular o número médio de filhos tidos nascidos vivos, a partir da equação $Fi = \phi_i + k_i f_i$. Os fatores estão apresentados na Tabela 3.A.2.

TABELA 3.A.2

FATORES MULTIPLICADORES PARA ESTIMATIVA DO VALOR MÉDIO DA FECUNDIDADE ACUMULADA (F_i) POR GRUPOS QUINQUÊNAIS DE ACORDO COM A FÓRMULA $F_i = \phi_i + k_i f_i$ (ONDE f_i SE APLICA A IDADES 15-20, 20-25 ETC.)

IDADE EXATA DAS MULHERES	FATORES								
15 — 20 .	0,335	0,690	1,030	1,390	1,760	2,130	2,460	2,745	
20 — 25 .	2,025	2,170	2,265	2,330	2,380	2,420	2,455	2,485	
25 — 30 .	2,420	2,455	2,485	2,510	2,535	2,560	2,580	2,605	
30 — 35 .	2,560	2,580	2,605	2,625	2,650	2,675	2,700	2,730	
35 — 40 .	2,675	2,700	2,730	2,760	2,800	2,845	2,895	2,960	
40 — 45 .	2,845	2,895	2,960	3,040	3,145	3,285	3,470	3,720	
45 — 50	3,195	3,455	3,720	3,980	4,240	4,495	4,750	5,000	
f_1/f_2	0,036	0,113	0,213	0,330	0,460	0,605	0,764	0,939	
\bar{m} (anos)	32,2	31,2	30,2	29,2	28,2	27,2	26,2	25,2	

A Tabela 3.A.2 fornece índices para um conjunto de distribuições modelo de fecundidade de um parâmetro especificadas alternadamente pelos valores de \bar{m} ou de f_1/f_2 . Para estas distribuições a multiplicação de um f_i (taxa de fecundidade específica por idade), para um intervalo de idades, pelo apropriado fator tabulado, dá exatamente a média *adicional* de filhos tidos nascidos vivos desde o começo do intervalo das mulheres destas idades na população. A soma deste valor e das medidas de f acumuladas desde o começo do intervalo dá o número médio de filhos tidos nascidos vivos por mulher para o grupo etário. Os fatores multiplicadores, em geral, não são exatos para as distribuições de fecundidade observadas, mas darão uma boa aproximação dos valores corretos. Já que a exatidão da estimativa para cada grupo etário depende apenas da concordância com o modelo para aquele intervalo, diferentes distribuições do conjunto poderiam ser ajustadas para cada grupo de mulheres. Descobriu-se que na prática se obtém bons resultados com o uso da razão f_1/f_2 observada para a determinação dos fatores multiplicadores para os três grupos etários mais jovens e da média \bar{m} das taxas específicas de fecundidade registradas para determinar os restantes.

Os fatores multiplicadores para se derivar o número médio de filhos tidos nascidos vivos por mulher a partir das taxas específicas de fecundidade por grupos etários com um deslocamento de meio ano para o passado foram calculados pelo mesmo processo com as mudanças apropriadas nos limites de integração. Estes fatores, que devem ser usados quando as taxas específicas de fecundidade atuais são calculadas a partir dos nascimentos ocorridos no ano anterior, estão apresentados na Tabela 3.1 do texto.

APÊNDICE B

NOTA SOBRE O EFEITO DO ERRO NA DECLARAÇÃO DA IDADE

Etienne van de Walle

O objetivo da presente nota é apresentar empiricamente alguns possíveis efeitos de erros na declaração da idade do tipo que se encontram nas pesquisas demográficas em populações africanas, a saber, aqueles que se baseiam em informações quanto à fecundidade “atual” e “retrospectiva”; número de nascimentos ocorridos durante um período de referência — usualmente 12 meses — precedente à pesquisa; e número de filhos tidos nascidos vivos e filhos sobreviventes. Estarão inerentes à nossa discussão certas suposições simplificadoras. O possível efeito da omissão sobre a distribuição etária feminina não será tomado em conta. Será suposto que a distribuição etária real pode ser descrita apropriadamente por uma distribuição etária estável que possa ser ajustada à população registrada, e aquela diferença entre a população estável e a registrada se deve exclusivamente às transferências de mulheres de uma classe etária para a seguinte, como resultado de erro na estimativa das idades. Fazendo-se uma suposição sobre a natureza das transferências, o número de mulheres transferidas e a direção das transferências podem ser estimados através da comparação da distribuição etária estável com a registrada. As mulheres transferidas têm uma certa fecundidade e uma certa parturição. Serão investigadas as implicações de muitas hipóteses sobre a fecundidade e a parturição dessas mulheres.

Com propósitos ilustrativos, selecionamos a população do Congo (Leopoldville). A Tabela 3.B.1 apresenta a distribuição etária da população estável estimada. A última foi obtida selecionando-se um nível apropriado de mortalidade e supondo-se que os grupos etários de idade 0-9 anos estão informados com exatidão. Apenas a população feminina é relevante para a nossa discussão.

EFEITO DA MÁ DECLARAÇÃO DA IDADE SOBRE A FECUNDIDADE REGISTRADA

Durante o desenvolvimento deste trabalho, será considerado que a natureza das transferências é tal que as mulheres são mudadas apenas de uma classe etária para outra contígua, seja para frente ou para trás.

As meninas que faltarem na classe etária 10-14 serão encontradas na classe 15-19, e etc. Pode haver transferência nos dois sentidos, mas estamos interessados apenas no resultado líquido. Sejam 1, 2 e 3 os três grupos etários envolvidos na má declaração da idade do grupo etário 2. (Algumas mulheres se transferem do grupo etário 1 para o grupo etário 2, e algumas por sua vez do grupo etário 2 para o 3 — ou possivelmente do grupo etário 2 para o 1 ou do grupo etário 3 para o 2).

Sejam f_1 , f_2 e f_3 a fecundidade média (número de filhos tidos nascidos vivos por mulher ocorridos durante o ano anterior) dos grupos etários reais; n_1 , n_2 e n_3 , o número de mulheres em cada grupo etário; n_{12} e n_{23} , o número líquido transferido do grupo 1 para o 2, e do 2 para o 3; e f_{12} e f_{23} , a fecundidade média das mulheres transferidas. (No caso de uma transferência para o grupo mais jovem, o sinal de n_{12} e de n_{23} será negativo).

TABELA 3.B.1

DISTRIBUIÇÃO ETÁRIA DA POPULAÇÃO REGISTRADA E ESTÁVEL DO CONGO, POR MIL, PARA AS MULHERES

IDADE	REGISTRADA	ESTÁVEL	DIFERENÇA ACUMULADA (Estável-Registrada)
0 — 9.	296	296	0
10 — 14.....	82	114	—32
15 — 19....	72	100	—60
20 — 24.	83	87	—64
25 — 29	92	76	—48
30 — 34	106	65	— 7
35 — 44	136	104	+25
45 +	133	158	0
TOTAL... . . .	1 000	1 000	

Então a fecundidade f'_2 do grupo etário registrado é:

$$f'_2 = \frac{n_2 f_2 + n_{12} f_{12} - n_{23} f_{23}}{n_2 + n_{12} - n_{23}}$$

ou

$$f'_2 = f_2 \cdot \frac{n_2}{n_2 + n_{12} - n_{23}} + \frac{n_{12} f_{12} - n_{23} f_{23}}{n_2 + n_{12} - n_{23}}$$

Assim a fecundidade registrada depende tanto do número de mulheres transferidas de um grupo etário para outro como da fecundidade das mulheres transferidas.

f_2 e f'_2 seriam iguais sempre que

$$n_{12} (f_{12} - f_2) = n_{23} (f_{23} - f_2)$$

Esta condição pode ser preenchida em três casos:

- a) As mulheres transferidas de, e removidas para, um dado grupo etário, têm as mesmas taxas específicas de fecundidade por idade (média) que aquelas do grupo etário real:

$$f_{12} = f_{23} = f_2$$

- b) os números de mulheres transferidas para, e removidas de, um grupo etário são iguais, e ambos os grupos têm as mesmas taxas específicas de fecundidade por idade:

$$n_{12} = n_{23} \text{ e } f_{12} = f_{23}$$

- c) o número de mulheres transferidas para e removido de um grupo etário está em razão inversa às diferenças absolutas entre a fecundidade específica por idade desses grupos e a do grupo etário original:

$$\frac{n_{12}}{n_{23}} = \frac{f_{23} - f_2}{f_{12} - f_2}$$

Todavia, nenhum desses casos ocorrerá provavelmente de forma consistente e na presença de erros na declaração da idade. Além disso, mesmo se for cumprida uma das condições para qualquer grupo 2 de idades dado de modo que $f'_2 = f_2$, f'_1 e f'_3 ainda estariam afetados pelas transferências. Em outras palavras, sempre que houver erros na declaração de idade, as séries de taxas específicas de fecundidade por idade estarão distorcidas.

Prosseguiremos agora com um exemplo numérico, usando as distribuições etárias registradas e estável, apresentadas anteriormente na Tabela 3.B.1, e taxas específicas de fecundidade por idade parecidas às das ilhas Cocos Keeling¹.

Não obstante, supõe-se que a fecundidade antes dos 15 anos e depois dos 45 seja nula.

Nossas suposições, sobre a natureza da má declaração de idade nesse exemplo implicam que, das 114 meninas de 10 a 14 anos de idade, 82 foram registradas como tendo de 10 a 14 anos e 32 como tendo de 15 a 19 anos... e etc., como está apresentado nas Tabelas 3.B.2 e 3.B.3.

¹ Na verdade, consideram-se as médias das taxas para diferentes coortes. Ver SMITH, T. E. The Cocos-Keeling Islands: a Demographic Laboratory *Population Studies*, 14 (2): 109, nov. 1966

TABELA 3.B.2

DISTRIBUIÇÃO EM CLASSES DE IDADE REGISTRADAS DAS MULHERES PERTENCENTES A CLASSES DE IDADE DADAS DA POPULAÇÃO ESTÁVEL; FECUNDIDADE POR IDADES ESPECÍFICAS E NÚMERO DE NASCIMENTOS NA POPULAÇÃO ESTÁVEL

IDADES REGISTRADAS	IDADES ESTÁVEIS (Reais)								TOTAL REGISTRADO
	0-9	10-14	15-19	20-24	25-29	30-34	35-44	45 +	
0 - 9	296								296
10 - 14		82							82
15 - 19		32	40						72
20 - 24			60	23					83
25 - 29				64	28				92
30 - 34					48	58			106
35 - 44						7	104	25	136
45 +								133	133
Total Estável..	296	114	100	87	76	65	104	158	1,000
TFE	0	0	0,165	0,365	0,390	0,320	0,200	0	
Nascimentos Estável	0	0	16,50	31,76	29,64	20,80	20,80	0	119,50

TABELA 3.B.3

COMPARAÇÃO ENTRE A POPULAÇÃO REGISTRADA E A ESTÁVEL

GRUPO DE IDADE	TRANSFERIDAS PARA UM GRUPO DE IDADE INFERIOR		CORRETA-MENTE INFORMADAS NO GRUPO ETÁRIO	TRANSFERIDAS PARA UM GRUPO DE IDADE SUPERIOR		POPULAÇÃO REGISTRADA (1) + (3) + (4)	POPULAÇÃO ESTÁVEL (2) + (3) + (5)
	(1)	(2)		(4)	(5)		
0 - 4	—	—	296	—	—	296	296
10 - 14	—	—	82	—	32	82	114
15 - 19	32	—	40	—	60	72	100
20 - 24	60	—	23	—	64	83	87
25 - 29	64	—	28	—	48	92	76
30 - 34	48	—	58	—	7	106	65
35 - 44	7	—	104	25	—	136	102
45 +	—	25	133	—	—	133	158

Agora, serão feitas muitas hipóteses quanto à fecundidade das mulheres transferidas, isto é, sobre f_{12} e f_{23} . Devido à falta de conhecimento sobre a natureza do erro de declaração da idade nossas suposições serão mantidas tão simples quanto possível. Serão considerados especificamente os três casos a seguir:

- 1) As mulheres transferidas têm uma fecundidade menor que a média. Suponhamos por exemplo que sua fecundidade é apenas metade da fecundidade média original do grupo etário:

$$f_{12} = 1/2 f_1, \text{ e } f_{23} = 1/2 f_2$$

- 2) As mulheres transferidas têm a mesma fecundidade que o grupo etário do qual são transferidas:

$$f_{12} = f_1 \quad \text{e} \quad f_{23} = f_2$$

- 3) As mulheres transferidas têm uma fecundidade acima da média. Por exemplo, consideramos a fecundidade como um quarto maior do que a do grupo do qual as mulheres se transferiram:

$$f_{12} = 1,25 f_1 \quad f_{23} = 1,25 f_2$$

Quando falamos da fecundidade das mulheres transferidas, queremos dizer a fecundidade atual, ou o número médio de nascimentos ocorridos por mulher durante o ano anterior. Para cada mulher individual há apenas a alternativa de ter apenas um filho, ou não ter. Uma "fecundidade maior que a média da mulher transferida" significa que as mulheres que tiveram um filho durante o ano anterior têm uma maior probabilidade de serem transferidas quando sua idade for mal informada. Se o fato de ter tido um filho durante os últimos doze meses não teve nenhuma influência na má declaração da idade, seria lógico supor que as mulheres transferidas tinham a fecundidade média de seu grupo etário. Se as mulheres com um filho recém-nascido são mais facilmente transferidas que as outras, então sua fecundidade estará acima da média; se forem menos facilmente transferíveis, sua fecundidade estará abaixo da média do grupo. Outros fatores, além da fecundidade atual, são obviamente operantes na transferência; um fator, a parturição, será discutido mais tarde. A fecundidade não é operante na transferência das mulheres dos grupos etários de 10-14 e de 45+, mas o estado civil das mulheres pode então ser importante.

As fecundidades específicas por idade que resultam de nossas hipóteses são apresentadas na Tabela 3.B.4. Esta tabela indica que, qualquer que seja a fecundidade das mulheres transferidas, a fecundidade "registrada" altera-se seriamente em presença de uma forte má declaração da idade.

Será observado que os cálculos anteriores envolvem (a) um único modelo de redistribuição de mulheres de dez anos de idade ou mais entre os grupos etários como está indicado nas Tabelas 3.B.2 e 3.B.3, e (b) três modelos de redistribuição de filhos a serem atribuídos às mulheres de diferentes grupos etários, como está implicado por nossas três hipóteses. Já que o número total de mulheres distribuídas pelas idades férteis não muda de caso para caso, e já que o número de filhos atribuídos a essas mulheres permanece o mesmo todo o tempo, a taxa geral de fecundidade é idêntica em todos os três exemplos. (Ela difere da taxa da população estável apenas na medida em que a redistribuição em (a) cruza os limites do período reprodutivo). Todavia a distribuição por idade das mães, se modifica sob nossas hipóteses. A fecundidade máxima pode apresentar-se, não no grupo etário de idades 25-29,

TABELA 3.B.4

TAXAS DE FECUNDIDADE ESPECÍFICA POR UNIDADE DA POPULAÇÃO ESTÁVEL E DA POPULAÇÃO REGISTRADA CONFORME AS TRÊS HIPÓTESES SOBRE A FECUNDIDADE DAS MULHERES TRANSFERIDAS

IDADE	ESTÁVEL	"REGISTRADAS" SEGUNDO A HIPÓTESE		
		Fecundidade mais Baixa (1)	Fecundidade Média (2)	Fecundidade mais Alta (3)
15 — 19 . . .	0,165	0,160	0,092	0,057
20 — 24. . . .	0,365	0,302	0,220	0,180
25 — 29. . . .	0,390	0,353	0,373	0,385
30 — 34.	0,320	0,269	0,352	0,391
35 — 44.	0,200	0,161	0,169	0,174

como sucede entre os habitantes das ilhas de Cocos, mas mais cedo ou mais tarde. A idade média do período reprodutivo aparentemente mudará de 27,8 anos na população estável para 29,0 anos sob a hipótese 1, 30,1 anos sob a hipótese 2 e 30,7 anos sob a hipótese 3. A fecundidade acumulada, incluindo a fecundidade total, também é marcadamente diferente sob as três hipóteses, como está apresentado na Tabela 3.B.5 (A fecundidade acumulada F foi obtida através de interpolação linear simples:

$$F_{(17,5)} = 2,5 f_{(15-19)}$$

$$F_{(22,5)} = 5 f_{(15-19)} + 2,5 f_{(20-24)}$$

etc.).

Nesse ponto a análise produz dois resultados:

- 1) A má declaração da idade afeta a forma da curva de fecundidade de uma maneira que depende tanto do número de mulheres deslocadas de um grupo etário para o seguinte como de sua fecundidade;
- 2) a fecundidade acumulada calculada a partir das taxas específicas de fecundidade por idade registradas, na presença de erros de declaração na idade são substancialmente diferentes dos números originais.

TABELA 3.B.5

FECUNDIDADE ACUMULADA COM A FECUNDIDADE ESPECÍFICA POR IDADE ORIGINAL E CONFORME AS HIPÓTESES 1, 2 E 3

IDADE	ESTÁVEL	"REGISTRADAS" SEGUNDO A HIPÓTESE		
		Fecundidade mais Baixa (1)	Fecundidade Média (2)	Fecundidade mais Alta (3)
17,5.. . .	0,413	0,400	0,230	0,143
22,5 . . .	1,713	1,555	1,010	0,735
27,5	3,625	3,193	2,493	2,148
32,5	5,400	4,748	4,305	4,088
40	7,200	6,225	6,030	5,935
45 + ..	8,200	7,030	6,875	6,805

EFEITO DA MÁ DECLARAÇÃO DA IDADE SOBRE A PARTURIÇÃO REGISTRADA

Com a mesma notação anteriormente usada consideraremos agora para o efeito dos erros de declaração da idade sobre a parturição registrada das mulheres em nossa população.

Sejam P_1 , P_2 e P_3 as parturições médias nos grupos etários considerados; P_{12} , a parturição média das mulheres transferidas do grupo 1 para o 2; P_{23} , parturição média daquelas transferidas do grupo 2 para o 3. Então a parturição "registrada" P'_2 será:

$$P'_2 = \frac{n_2 P_2 + n_{12} P_{12} - n_{23} P_{23}}{n_2 + n_{12} - n_{23}}$$

Reportamos o leitor à discussão dos casos onde f'_2 seria igual a f_2 , por ser a fórmula para o cálculo de P'_2 essencialmente similar à usada para o cálculo de f_2 . Em virtude da forma da curva de parturição em função da idade, em comparação com a curva de fecundidade, é ainda menos provável que P'_2 fosse igual a P_2 do que f'_2 fosse igual a f_2 .

Aplicando-se a fórmula anterior ao número de mulheres transferidas nas Tabelas 3.B.2 e 3.B.3, levantamos uma série de hipóteses a respeito da parturição das mulheres transferidas, paralelas àquelas feitas anteriormente em relação à fecundidade:

A) a parturição das mulheres transferidas está abaixo da média do grupo etário. Se P_0 for a parturição do grupo etário 0 imediatamente precedente ao grupo etário 1, tomamos como exemplo:

$$P_{12} = P_1 - \left(\frac{P_1 - P_0}{2} \right) \text{ e } P_{23} = P_2 - \left(\frac{P_2 - P_1}{2} \right)$$

B) As mulheres transferidas têm a mesma parturição média dos grupos etários dos quais foram transferidas:

$$P_{12} = P_1 \text{ e } P_{23} = P_2$$

C) As mulheres transferidas têm uma parturição maior que a média. Por exemplo:

$$P_{12} = P_1 + \left(\frac{P_1 - P_0}{4} \right) \text{ e } P_{23} = P_2 + \left(\frac{P_2 - P_1}{4} \right)$$

As parturições "registradas" resultantes estão dadas na Tabela 3.B.6.

TABELA 3.B.6

PARTURIÇÃO DA POPULAÇÃO ESTÁVEL E DA POPULAÇÃO REGISTRADA SEGUNDO AS TRÊS HIPÓTESES SOBRE A PARTURIÇÃO DAS MULHERES TRANSFERIDAS

IDADE	ESTÁVEL	PARTURIÇÃO "REGISTRADA" SEGUNDO A HIPÓTESE		
		Parturição mais Baixa (A)	Parturição Média (B)	Parturição mais Alta (C)
17,5.. . . .	0,413	0,401	0,230	0,143
22,5.	1,738	1,142	0,781	0,600
27,5.	3,625	2,343	2,312	2,297
32,5	5,400	4,227	4,596	4,780
40	7,200	7,154	7,291	7,360

O EFEITO DA MÁ DECLARAÇÃO DA IDADE SOBRE AS RAZÕES PARIDADE "REGISTRADA"/FECUNDIDADE ACUMULADA (P/F)

Podemos agora combinar as três hipóteses feitas sobre a parturição das mulheres transferidas com as três hipóteses feitas sobre sua fecundidade. Isso é feito na forma de razões P/F . Elas aparecem na Tabela 3.B.7.

TABELA 3.B.7

RAZÕES P/F SEGUNDO UMA COMBINAÇÃO DE HIPÓTESES SOBRE
A FECUNDIDADE E A PARTURIÇÃO DAS MULHERES
TRANSFERIDAS

ESPECIFICAÇÃO	IDADES	RAZÃO P/F SEGUNDO A HIPÓTESE			
		Fecun- didade mais Baixa (1)	Fecun- didade Média (2)	Fecun- didade mais Alta (3)	
Combinadas com as hipó- teses	(A) Parturição mais Baixa	15 — 19	1,00	1,74	2,80
		20 — 24	0,73	1,13	1,55
		25 — 29	0,73	0,94	1,09
		30 — 34	0,89	0,98	1,03
		35 — 44	1,15	1,19	1,21
	(B) Parturição Média.	15 — 19	0,58	1,00	1,61
		20 — 24	0,50	0,77	1,06
		25 — 29	0,72	0,93	1,08
		30 — 34	0,97	1,07	1,12
		35 — 44	1,17	1,21	1,23
	(C) Parturição mais alta	15 — 19	0,36	0,62	1,00
		20 — 24	0,39	0,59	0,82
		25 — 29	0,72	0,92	1,07
		30 — 34	1,01	1,11	1,17
		35 — 44	1,18	1,22	1,24

A Tabela 3.B.7 mostra que o erro na declaração da idade normalmente deformará as razões P/F . Não resulta inconcebível que o registro da idade esteja relacionado com a parturição e/ou a fecundidade atual. Para os distritos do Congo sempre se encontra que a razão P/F nas idades 15-19 é maior que nas idades 20-24, que por sua vez ultrapassa a razão P/F nas idades 25-29. Esta seqüência pode dever-se a um padrão sistemático no erro da declaração da idade, talvez relacionando com a parturição ou fecundidade. Em nosso modelo, um padrão como o de transferência ascendente das mulheres de fecundidade média mas com uma paridez inferior a média poderia produzir a seqüência registrada. Não obstante, não está claro porque se atribuirá às mulheres de paridez inferior a média uma idade superior a que têm realmente. Talvez seja possível que as mulheres infecundas, cuja idade é mais difícil de ultimar, estejam deslocadas até uma categoria central de idades. Isto, não obstante, só obedece a conjecturas.

PESQUISA ESPECIAL DE BOVINOS

Estruturada na mais moderna técnica de levantamentos estatísticos, e efetuada por uma equipe de pesquisadores devidamente qualificados, foi realizada a 1.^a *Pesquisa Especial de Bovinos — 1972*, através da Superintendência de Estatísticas Primárias do IBGE.

A Pesquisa, realizada em 1972, baseou-se em informações referentes ao ano de 1971, e abrangeram oito Unidades da Federação (Bahia, Minas Gerais, São Paulo, Paraná, Santa Catarina, Rio Grande

do Sul, Mato Grosso e Goiás) que no seu conjunto representavam, quando do Censo Agropecuário de 1970, 84,14% do rebanho bovino existente no País.

Trata-se de trabalho pioneiro, mas de profundidade, focalizando um dos aspectos mais importantes da economia brasileira, que define perfeitamente em seus diversos levantamentos estatísticos a grandiosidade da Pesquisa e os resultados positivos da agropecuária brasileira.

1. CARACTERÍSTICAS DA PESQUISA

A Pesquisa objetivou, principalmente, o levantamento de informações sobre o número de cabeças de bovinos existente no País, a distribuição do rebanho segundo a idade e sexo, e a dos estabelecimentos quanto a finalidade principal da exploração do rebanho.

O levantamento dos dados foi feito por uma amostra nacional e

mediante a aplicação de um questionário em cada estabelecimento agropecuário selecionado. Para o estabelecimento do âmbito de investigação da Pesquisa, ou seja, a área geográfica que foi coberta no País, considerou-se a distribuição do rebanho bovino e sua representatividade em termos de concentração dos efetivos, volume da pro-

dução e importância das regiões em função do abastecimento interno e da exportação.

Assim, foram escolhidas as oito Unidades da Federação que, no seu conjunto, caracterizavam os objetivos perseguidos pela Pesquisa e que representavam uma redução no seu custo operacional pela maior concentração geográfica da área de atividades.

Dentro desse critério foram selecionados 229 municípios e ...

12.492 estabelecimentos agropecuários, assim relacionados:

<hr/>				
Rio Grande do Sul	44 municípios	—	3 213 estabelecimentos	
Santa Catarina	8	>	— 457	>
Paraná	10	>	— 965	>
São Paulo	34	>	— 1 902	>
Minas Gerais	55	>	— 2 531	>
Goiás	33	>	— 1 531	>
Mato Grosso	26	>	— 1 370	>
Bahia	19	>	— 750	>

2. METODOLOGIA

2.1 Discriminação do modelo utilizado

Foi adotado na Pesquisa Especial de Bovinos, por amostragem probabilística, a nível de produtor, como sistema de referência, o cadastro de estabelecimentos do Censo Agropecuário de 1970 segundo os procedimentos metodológicos a seguir descritos.

Dada a necessidade de redução dos custos de operação, do tempo de coleta e da precisão desejada, aplicou-se um modelo de estratificação de unidades compostas.

A população estatística foi dividida em duas subpopulações:

- a) A primeira — constituída pelas Unidades da Federação da Região Sul (RS, SC e PR) denominada de Região 1;
- b) A segunda — constituída por Unidades da Federação das Regiões Sudeste (SP e MG), Centro-Oeste

(GO e MT) e Nordeste (BA), denominada de Região 2.

Em cada subpopulação foram aplicados dois estágios de seleção da amostra:

1.º estágio — seleção de municípios;

2.º estágio — seleção de conglomerados de estabelecimentos.

Para efeito de conceituação, as unidades dos dois estágios foram assim definidas:

Município — Unidade administrativa menor que compõe a Divisão Territorial do Brasil;

Conglomerado — Unidade constituída com base no sistema de referência adotado, que contém um certo número de estabelecimentos agrupados, possuindo no seu conjunto um total de bovinos, que em cada município considerado foi, tanto quanto possível, o mesmo.

Nas duas subpopulações os municípios foram divididos em dois grupos distintos. O primeiro deles, abrangendo os principais municípios no que tange ao número de cabeças de bovinos, constituindo o grupo dos municípios auto-representativos. Os remanescentes foram estratificados, em função da variável — número total de cabeças de bovinos — selecionando-se um município em cada estrato.

Observe-se que os municípios auto-representativos participaram do segundo estágio de seleção da amostra, com probabilidade certa de inclusão.

Por este sistema, nas duas subpopulações, foram identificados 43 municípios auto-representativos e 186 municípios selecionados, em cada um dos estratos, segundo critério de equiprobabilidade, totalizando 229 municípios. Assim procedendo compôs-se o painel para o segundo estágio de seleção da amostra.

A fim de conseguir-se uma representatividade maior da amostra, identificaram-se nos municípios selecionados, os principais estabelecimentos criadores de gado bovino designados estabelecimentos auto-representativos, participando da amostra com probabilidade certa de inclusão.

O restante do município sofreu um processo de conglomeração, de onde foram selecionados, em média, 5 conglomerados por município, culminando finalmente, com a identificação de 12.492 estabelecimentos, dos quais 3.559 considerados como auto-representativos.

2.2 Erros de amostragem

2.2.1 — Método de estimação

O uso de estimativas de razão é um dos métodos mais comumente empregados, quando há disponibilidade de informações nesse sentido. Seu emprego torna-se apropriado sempre que as unidades da população possuam duas características que sejam positivamente correlacionadas. Tanto maior a correlação entre as variáveis, maior será o efeito de aprimoramento nas estimativas obtidas.

Para a 1.^a Pesquisa Especial de Bovinos — 1972, haviam em disponibilidade no Centro de Informática, dados referentes a estabelecimentos com bovinos, obtidos pelo Censo Agropecuário de 1970 (Folha de Coleta do Recenseador CA-204). Levando-se em conta estes fatores decidiu-se pela utilização de um estimador de razão de seguinte tipo:

$$x'' = \frac{x'}{y'} \cdot Y \quad \text{Onde:}$$

x' — estimativa simples do total para a variável x , com dados da pesquisa

y' — estimativa simples do total para a variável y , com dados do cadastro

Y — total conhecido para a variável Y

2.2.2 — Tendenciosidade do estimador de razão

Sabe-se estatisticamente que o estimador de razão é um estimador tendencioso, isto é, ao calcu-

lar-se a média das razões $\frac{x'}{y'}$, para todas as amostras possíveis, a média assim calculada não será igual à média verdadeira da população. Entretanto, na maioria das aplicações práticas, a tendenciosidade é tão pequena, que compensa o seu emprego pela redução no erro de amostragem. Além do mais, o estimador de razão possui a qualidade de consistência. Isto significa que utilizando-se uma amostra suficientemente grande é possível afirmar que as estimativas obtidas estarão tão próximas dos valores verdadeiros quanto se desejaria que estivessem.

2.2.3 — Variância do estimador de razão

A variância relativa do estimador empregado na Pesquisa Especial de Bovinos é dada pela seguinte fórmula:

$$v_{x'}^2 = v_x^2 + v_y^2 - 2\rho v_x \cdot v_y \quad \text{Onde:}$$

$v_{x'}^2$ — Variância relativa para o total estimado da variável x

v_y^2 — variância relativa para o total estimado da variável y

ρ — coeficiente de correlação entre as variáveis x e y

$v_{x'}$ — coeficiente de variação para o total estimado da variável x

$v_{y'}$ — coeficiente de variação para o total estimado da variável y

Destaque-se o fato de que, no cálculo das variâncias relativas para os totais estimados, em virtude da seleção de somente um município por estrato, houve necessidade da utilização do método de agrupamento por estratos¹.

Apresenta-se a seguir um quadro demonstrativo, a nível de Região, com os coeficientes de variação calculados para as principais variáveis da pesquisa.

1.ª PESQUISA ESPECIAL DE BOVINOS — 1972

VARIÁVEIS PRINCIPAIS	COEFICIENTE DE VARIAÇÃO (%)		
	Região I (RS, SC PR)	Região II (SP, MG, BA, GO, MT)	Total das Regiões I e II
N.º total de estabelecimentos	2,00	2,24	1,41
Área total dos estabelecimentos	3,87	2,83	2,45
N.º total de cabeças de bovinos	3,00	2,65	2,24
N.º total de estabelecimentos com áreas de pastagens	3,46	2,24	2,44
Área total dos estabelecimentos com áreas de pastagens	3,74	3,00	2,45
N.º total de cabeças em estabelecimentos com áreas de pastagens	3,00	2,83	2,24

3. ESTRUTURA DAS ATIVIDADES DE CAMPO

A estrutura estabelecida em seus vários níveis para a realização da coleta de dados, visou o atendimento de exigências próprias de uma pesquisa de tal natureza, em

que a especialização do pessoal utilizado, seu treinamento adequado e coordenação indicada, tornou-se imprescindível para um bom resultado das atividades de campo.

¹ Sample Survey Methods and Theory Hansen — Hurwitz — Madow — pág 400-419

Assim, estabeleceu-se a seguinte estrutura:

- a — *Coordenação Geral e Supervisão Técnica* — A cargo do Centro Brasileiro de Estatísticas Agropecuárias do IBGE, o qual facilitou aos membros componentes da Comissão Especial de Planejamento, Controle e Avaliação das Estatísticas Agropecuárias ... (CEPAGRO), ampla assistência para atuarem como observadores nos trabalhos de campo;
- b — *Coordenação Estadual* — Subordinada administrativa-mente ao Delegado de Estatística do IBGE, no Estado, e sob a orientação técnica do dirigente do Projeto. Nas Unidades da Federação em que existem Setores de Estatísticas Agropecuárias nas Delegacias de Estatística, o Chefe deste Setor, um enge-

nheiro-agrônomo, atuou como Coordenador Estadual. Nas demais Unidades, foi contratado um engenheiro-agrônomo para suprir as necessidades de coordenação técnica da pesquisa;

- c — *Supervisão a nível de Área de Coleta* — O Supervisor, diretamente subordinado ao Coordenador Estadual, foi o responsável pela pesquisa a nível de Área de Coleta estabelecida para o levantamento, compreendendo um determinado número de municípios envolvidos no trabalho;
- d — *Coleta de dados* — O Entrevistador foi o responsável pela coleta de dados propriamente dita, a nível de Setor estabelecido para o levantamento e compreendendo um certo número de estabelecimentos selecionados a serem investigados.

4. INSTRUMENTO DE COLETA

Para o registro e controle dos dados coletados e das operações de campo foram utilizados os seguintes instrumentos de coleta:

Questionário — PB.1 — Formulário destinado ao registro das informações obtidas em cada estabelecimento visitado.

Folha de Listagem do Entrevistador — PB.2 — Lista de enumeração dos estabelecimentos levantados, contendo o registro das principais informações de identifi-

cação dos questionários e das ocorrências verificadas, para fins de controle.

Folha de Controle do Supervisor — PB.3 — Formulário destinado ao resumo e controle das atividades de cada Entrevistador, por município.

Folha de Controle do Coordenador — PB.4 — Formulário destinado ao resumo e controle das atividades em cada município, por Unidade da Federação.

5. CONCEITOS BÁSICOS

Estabelecimentos — Considera-se como “estabelecimento” todo o terreno de *área contínua*, independente de tamanho e situação, formado de uma ou mais *parcelas confinantes*, subordinado a uma *única administração*, onde se processa uma exploração agropecuária.

Área contínua — Os cursos d'água, rodovias ou ferrovias que cortem as terras do estabelecimento, não caracterizam descontinuidade de sua área.

Tamanho — Diz respeito à dimensão do estabelecimento. Não importa a área que tenha o terreno, basta que preencha as outras condições para ser considerado *um estabelecimento*.

Situação — Refere-se à localização do estabelecimento, que poderá estar situado em zona urbana, suburbana ou rural.

Parcelas confinantes — São as áreas contíguas sob a mesma administração, ocupadas segundo diferentes condições legais (próprias, arrendadas, ocupadas gratuitamente).

Única administração — Caracteriza uma única administração, o fato de as terras, contínuas e confinantes, encontrarem-se sob a responsabilidade de uma só pessoa — física ou jurídica.

Exploração agropecuária — Consideradas para efeito desta pesquisa, como exploração agropecuária, apenas as seguintes atividades:

— cria, recria e engorda de bovinos, independente da finalidade; produção de leite e derivados; cria e/ou engorda de gado para produção de carne; produção de reprodutores; produção de gado para o trabalho.

Produtor — Considera-se como “produtor” a pessoa, física ou jurídica, economicamente responsável pela exploração do estabelecimento, durante o período de referência da pesquisa, seja qual for a *condição de posse das terras*.

Administrador — Considera-se “administrador”, a pessoa física diretamente responsável pela operação do estabelecimento, que não detém o poder de decisão sobre a direção econômica do mesmo, constituindo-se em um subordinado do produtor.

Sob a competência do administrador, se encontra a direção das atividades do dia a dia do estabelecimento que lhe tenham sido confiadas pelo produtor.

O administrador pode ter diversas denominações, conforme os seus encargos: administrador, gerente, diretor-técnico, capataz etc.

6. PESSOAL DA PESQUISA

Objetivando a melhor qualificação possível do pessoal envolvido na pesquisa foi elaborado um

Plano de Treinamento que contém diversos programas setoriais abrangendo os vários níveis de ser-

vidores que iriam atuar nas atividades de coleta, controle e crítica das informações.

Os Delegados de Estatística do IBGE nas Unidades onde se realizou a pesquisa, participaram de reuniões no CBEA onde lhes foram fornecidas informações detalhadas sobre:

Características básicas da pesquisa; Estrutura estabelecida para os trabalhos de campo; Sistema de recrutamento, seleção e contratação de pessoal; Organização e funcionamento dos Centros de Treinamento; Sistemas de controle dos instrumentos de coleta e das atividades de campo, etc.

Os técnicos designados para instruírem os Coordenadores Estaduais da pesquisa, receberam treinamento intensivo de 5 dias no CBEA, onde lhes foram ministrados ensinamentos sobre:

Características básicas e específicas da pesquisa; Definição das áreas de atuação, em vários níveis; Material de coleta — Relação dos estabelecimentos selecionados, questionários, manual de instruções, folhas de controle da coordenação, da supervisão e folha de listagem do entrevistador; Controle de distribuição, recepção e consistência dos instrumentos de coleta; Processamento de dados no órgão central — crítica, codificação e apuração de dados; Recrutamento, seleção, qualificação e remuneração do pessoal da coleta; Coleta de dados — mecanismo das atividades do coordenador estadual, supervisor e entrevistador; Orçamento da pesquisa e classificação das despesas, a nível de Unidade da

Federação; Descrição do programa de treinamento, a nível estadual.

Após o estabelecimento dos quadros do pessoal de coleta em cada Unidade da Federação, foram instalados 10 Centros de Treinamento, assim distribuídos:

RS	—	2
SC	—	1
PR	—	1
SP	—	1
MG	—	2
GO	—	1
MT	—	1
BA	—	1

O treinamento dos supervisores foi de 15 dias e dos entrevistadores de 12 dias, em regime intensivo e com dedicação integral em atividades escolares e de campo.

O programa de capacitação do pessoal foi desdobrado abordando os seguintes aspectos:

Características principais da pesquisa; Definição da área de atuação do supervisor e do entrevistador — estrutura das atividades de campo.

Material de coleta:

- uso e aplicação do Manual de Instruções
- conceitos básicos e específicos das variáveis levantadas
- questionário — o correto preenchimento, item a item
- relação dos estabelecimentos selecionados, por área de supervisão e por entrevistador

- folha de listagem do entrevistador
- folha de controle do supervisor
- seqüência de preenchimento dos instrumentos de coleta

Localização dos estabelecimentos selecionados:

- uso da relação de estabelecimentos selecionados
- utilização dos mapas por setor
- identificação dos estabelecimentos por outras indicações complementares
- roteiro para a cobertura da área de trabalho

A entrevista:

- a quem entrevistar
- como entrevistar

Obrigatoriedade da prestação de informações e sigilo das informações.

Durante o período de treinamento efetuaram-se diversas avaliações dos treinandos, mediante a aplicação de testes escritos e de

trabalhos de campo em área-piloto, e avaliação final.

O pessoal que participou da pesquisa em número de 586, compreendia as seguintes categorias:

Direção Geral (técnica e administrativa)	1
Supervisão Administrativa Estadual	8
Coordenação Técnica Estadual	8
Instrutores	10
Auxiliares de Coordenação	4
Auxiliares de escritório	8
Supervisores	111
Entrevistadores	392
Pessoal de Planejamento (na sede)	12
Pessoal de Operações (na sede)	32

O IBGE distribuirá, brevemente, o volume com os resultados desta 1.^a Pesquisa Especial de Bovinos — 1972, contendo um conjunto de 60 tabelas, sendo 24 por grupos de área dos estabelecimentos agropecuários, 24 por grupos de cabeças de bovinos e 12 por grupamentos especiais.

CALENDÁRIO DAS REUNIÕES INTERNACIONAIS DE ESTATÍSTICA

Divulgamos a seguir, o Calendário das Reuniões dos Organismos Internacionais de Estatística, extraído da Revista do Instituto Internacional de Estatística, volume 42, n.º 3, dezembro de 1974.

DATAS	LOCAL	REUNIÃO
1975		
27-29 janeiro	Brussels	First European Conference on Operations Research Sheraton Hotel, Place Rogier, Brussels
26-31 maio	Lofthus, Hardanger, Norway	6.ª Conferência Nórdica de Estatística Matemática <i>Informações:</i> Eivind Damsleth, Norsk Regnesentral, Forskningsveien 1B, Oslo 3, Norway
23-27 junho (provisório)	Geneva	Economic Commission for Europe *23.ª Sessão da Conferência de Estatísticos Europeus <i>Informações:</i> Director, Statistico Division, ECE, Palais des Nations, CH 1211 Geneva 10, Switzerland
21-23 julho	Kioto, Japão	The Institute of Management Sciences, in conjunction with IFORS and ORSJ 22.ª Reunião Internacional <i>Informações:</i> TIMS, 146 Westminster Street, Providence, R.I. 02903, USA
20-26 agosto	Toronto, Canadá	The Econometric Society 1975 — Congresso Mundial <i>Informações:</i> M. Nerlove, Dept. of Economics, North Western University, 629 Noyes Street, Evanston, Illinois 60201, USA

* Reuniões fechadas.

DATAS	LOCAL	REUNIÃO
25-28 agosto	Atlanta, Georgia	American Statistical Association 135. ^a Reunião Anual <i>Informações:</i> ASA, 806-15th Street, N. W., Washington D.C. 20005, USA
25-29 agosto	Atlanta, Georgia	Institute of Mathematical Statistics Reunião 149 1975 — Reunião Anual
25-30 agosto (tentativa)	Helsinki	International Association for Research in Income and Wealth Conferência Geral <i>Informações:</i> Mrs. N.D. Rugles, IARIW, Box 2020, Yale Station, New Haven, Conn. 06520, USA
1-9 setembro	Warsaw	International Statistical Institute 40. ^a Sessão <i>Informações:</i> E. Lunenberg, Director, Permanent Office, ISI, Prinses Beatrixlaan 428, The Hague-Voorburg
16-18 setembro	Venice	European Organization for Quality Control 19. ^a Conferência EOQC <i>Informações:</i> AICQ-1975 Conf. Secretariat, Piazza Diaz 2, 20123 Milan, Italy
outubro (tentativa)	Lima, Peru	Inter-American Statistical Institute XII Sessão da Comissão de Aperfeiçoamento das Estatísticas Nacionais (COINS) <i>Informações:</i> IASI, Secretaria-Geral da Organização dos Estados Americanos, Washington DC. 20006, USA

DATAS	LOCAL	REUNIÃO
3-14 novembro	Geneva	*18.^a Sessão da Comissão de População das Nações Unidas <i>Informações:</i> U.N. Population Division, New York, NY 10017, USA
1976		
15-18 junho	Copenhagen	European Organization for Quality Control 20. ^a Conferência Anual <i>Informações:</i> DFK's Sekretariat, Danmarks tekniske Hojskole, Bygning 229, 2800 Lyngby, Denmark
15-20 agosto	Boston, Mass.	Biometric Society 9. ^a Conferência Internacional <i>Informações:</i> Dr. H. Thöni, Abteilung Biometrie, Universität Hohenheim (L. H.) 7 Stuttgart 70, Postfach 106
23-26 agosto	Boston, Mass.	American Statistical Association 136. ^a Reunião Anual <i>Informações:</i> ASA, 806-15th Street, N.W., Washington DC 20005, USA

CONSELHO FEDERAL DE ESTATÍSTICA AMPLIA REGULAMENTAÇÃO PROFISSIONAL

A laboriosa classe que congrega os profissionais de estatística obtém mais uma vitória no campo de suas atividades, através da Resolução número 31, de 11 de dezembro de 1974, do Conselho Fe-

deral de Estatística, órgão este subordinado ao Ministério do Trabalho.

Para maior esclarecimento aos interessados, inclusive empregados e empregadores, transcrevemos a

seguir os considerandos da mencionada Resolução, bem como o seu texto integral:

“Dispõe sobre o Exercício das Atividades Auxiliares do campo Profissional do Estatístico e dá outras providências.

O Conselho Federal de Estatística (CONFE), no uso das atribuições que lhe confere o artigo 31 do Regulamento aprovado pelo Decreto n.º 62.497, de 1º de abril de 1968, especialmente pelo disposto nos incisos XVII, XIX e XX do referido artigo e tendo em vista a Lei n.º 4.739, de 15 de julho de 1965, que dispõe sobre o exercício da profissão de Estatístico.

Considerando que já existem, no País, cursos regulares, oficiais ou reconhecidos, formando Técnicos em Estatística de nível médio;

Considerando, por outro lado, o apreciável número de pretendentes ao registro profissional, como Estatístico, os quais, embora tenham requerido o registro, no prazo legal, não puderam comprovar o exercício profissional nos termos do artigo 43, do Regulamento, uma vez que apenas exerceram atividades auxiliares da especialidade do Estatístico;

Considerando, ainda, ser de justiça criar condições que possibilitem uma definição profissional compatível com o grau de formação conferido a uma considerável e crescente parcela da juventude, participante da mão-de-obra especializada formada pelo contingente oriundo dos Colégios de formação técnica em nível médio; e

Considerando, finalmente, que o registro, mesmo em condição

profissional de nível médio, tanto desses Técnicos em Estatística, possuidores de formação profissional intermediária, como daqueles pretendentes que não conseguiram atender ao disposto no artigo 43 do Regulamento, para registro profissional como Estatístico muito contribuirá, de um lado, para o benefício dessa Classe, abrindo-lhe campo e assegurando-lhe condições de trabalho compatíveis com a sua formação, e, de outro lado, em geral, que teriam, assim, onde recrutar mão-de-obra auxiliar especializada, no campo profissional do Estatístico, resolve:

Art. 1.º — O exercício de atividades auxiliares da especialidade do Estatístico, em todo o território nacional, observadas as condições de capacidade previstas nesta Resolução, na Lei n.º 4.739, de 15 de julho de 1965, e no seu Regulamento, aprovado pelo Decreto número 62.497, de 1.º de abril de 1968, no que couber, é livre:

I — Aos possuidores de diploma de conclusão de Curso Técnico de Estatística, em nível médio, concedido, no Brasil, por estabelecimento de ensino oficial ou oficialmente reconhecido;

II — Aos diplomados, por instituto estrangeiro de ensino médio, como Técnico em Estatística, que revalidem seus diplomas de acordo com a lei;

III — Aos que, comprovadamente, na data da publicação da presente Resolução, ocupem cargo, função ou emprego de Auxiliar de Estatístico, de Agente de Estatística ou de Agente de Coleta, em entidade pública ou privada;

IV — Aos que, comprovadamente, até a data da publicação da presente Resolução, hajam exercido, efetivamente, por período não inferior a 1 (um) ano, cargo, função ou emprego de Auxiliar de Estatístico, de Agente de Estatística ou de Agente de Coleta, em entidade pública ou privada, muito embora não mais estejam exercendo esses cargos, funções ou empregos;

V — Aos professores de disciplina de Estatística em estabelecimento de ensino médio, oficial ou oficialmente reconhecido, portadores de habilitação na forma da lei do ensino médio, bem como aos que, comprovadamente, até a data da publicação da presente Resolução, hajam exercido o magistério de disciplina de Estatística, em estabelecimento de ensino médio oficial ou oficialmente reconhecido, por período não inferior a 1 (um) ano letivo, com vínculo empregatício bem definido;

VI — Aos que, comprovadamente, na data da publicação da presente Resolução, ocupem cargo, função ou emprego de natureza semelhante à dos mencionados no inciso III deste artigo, bem como aos que, muito embora não mais estejam exercendo esses cargos, funções ou empregos os tenham exercido, efetivamente, por período não inferior a 1 (um) ano, em entidade pública ou privada e, ainda, aos que tenham realizado trabalho de natureza estritamente estatística, em entidade pública ou privada, com vínculo empregatício bem definido.

Parágrafo único — O livre exercício das atividades de que trata o presente artigo é permitido a

estrangeiros, quando compreendidos:

a) no inciso II, independentemente de revalidação do diploma, se exerciam, legitimamente, no Brasil, atividades auxiliares da especialidade do Estatístico na data da promulgação da Constituição de 1934, 16 de julho de 1934;

b) nos incisos III, IV, V e VI, desde que satisfaçam as condições neles estabelecidas.

Art. 2.º — O prazo para apresentação do requerimento para registro de pessoa física, ao Conselho Regional de Estatística (CONRE), é de 1 (um) ano, a contar da data da publicação da presente Resolução.

Parágrafo único — O prazo previsto neste artigo poderá ser prorrogado por mais 1 (um) período de 1 (um) ano, a critério do Conselho Federal de Estatística (CONFÉ).

Art. 3.º — Os diplomados como Técnico em Estatística, de nível médio, por estabelecimento de ensino oficial ou oficialmente reconhecido, ou em fase de reconhecimento, no Brasil, ficam obrigados, em obediência à legislação vigente, a providenciar o registro de pessoa física no Conselho Regional de Estatística (CONRE) de sua jurisdição, no prazo de 180 (cento e oitenta) dias contados da data da diplomação.

Parágrafo único — Aplica-se o disposto neste artigo aos professores de disciplina de Estatística que vierem a habilitar-se, na forma da lei do ensino médio, em data posterior à da publicação da presente Resolução, contando-se o respecti-

vo prazo de 180 (cento e oitenta) dias a partir da data da habilitação no Ministério da Educação e Cultura.

Art. 4.º — Satisfeitas as condições de comprovação previstas nesta Resolução, na Lei n.º 4.739, de 15 de julho de 1965, e no Regulamento aprovado pelo Decreto número 62.497, de 1.º de abril de 1968, será fornecida a cada inscrito, como documento comprobatório do registro, uma carteira de identidade profissional numerada que conterá os dados necessários e as assinaturas dos Presidentes do CONFE e do CONRE respectivos, assim como a do registrado.

§ 1.º — A emissão da Carteira de Identidade Profissional de Técnico em Estatística de Nível Médio obedecerá, no que couber, à legislação e as normas que disciplinam a emissão da Carteira de Identidade Profissional de Estatístico.

§ 2.º — Aos registrados no Conselho Federal de Estatística (CONFE), além da carteira profissional especial de que trata este artigo, poderá ser fornecido, pelo Conselho Regional de Estatística (CONRE), um cartão plastificado de identidade de Técnico em Estatística de Nível Médio, com as características previstas na Resolução n.º 17, de 21 de janeiro de 1972, do CONFE.

Art. 5.º — O exercício das atividades auxiliares da especialidade do Estatístico compreende:

a) executar cálculos estatísticos em geral;

b) participar, sob a orientação de Estatístico, de trabalhos re-

lacionados com a execução de pesquisas, levantamentos e análises de dados estatísticos;

c) integrar equipe chefiada por Estatístico encarregada de realizar estudos para elaboração padronizada de instrumentos de coleta de dados, gráficos, relatórios e pareceres no campo da Estatística;

d) auxiliar o Estatístico em tudo o que se relacionar com sua atividade profissional.

Art. 6.º — Satisfeitas as exigências da legislação específica do ensino médio e sem prejudicar o disposto no artigo 8.º da Lei número 4.739, de 15 de julho de 1965, poderão os Técnicos em Estatística de Nível Médio, exercer privativamente, o magistério das disciplinas de Estatística em estabelecimentos de ensino médio, oficiais ou oficialmente reconhecidos.

Art. 7.º — Na administração pública ou privada, o provimento ou o exercício de cargo, função ou emprego de atividade auxiliar de especialidade do Estatístico, bem como o exercício do magistério das disciplinas de Estatística, em estabelecimentos de ensino médio oficiais ou oficialmente reconhecidos, requerem, como condição essencial, que o interessado apresente a Carteira de Identidade Profissional de Técnico em Estatística de Nível Médio.

§ 1.º — A apresentação da Carteira de Identidade Profissional não exime o interessado da prestação do respectivo concurso, quando este for exigido para o provimento a que se refere este artigo.

§ 2.º — O disposto neste artigo, enquanto não houver habilita-

dos, não prejudica a situação daqueles que, à data da publicação da presente Resolução, estejam no pleno exercício de cargo privativo de Técnico em Estatística, em nível médio, ou estejam exercendo o magistério de disciplina de Estatística em estabelecimento de ensino médio, oficial ou oficialmente reconhecido, ou ainda, que, tendo sido habilitados em concurso público para Auxiliar de Estatístico ou assemelhado, ainda no prazo de sua validade aguardam provimento do cargo.

Art. 8.º — A prova de capacidade para obtenção do registro como Técnico em Estatística de Nível Médio será feita mediante a apresentação dos documentos previstos em um dos seguintes incisos:

I — diploma de conclusão de Curso de Técnico de Estatística, em nível médio, registrado, de acordo com a legislação vigente, no Ministério da Educação e Cultura;

II — a) ato original, ou cópia autenticada, de nomeação ou admissão para o exercício de cargo, função ou emprego de Auxiliar de Estatístico, Agente de Estatística, Agente de Coleta ou outros assemelhados, na administração pública;

b) recorte, original ou cópia autenticada, do órgão oficial de divulgação que publicou o ato ou, na falta deste, declaração oficial de que surtiu os efeitos, como se publicado fosse;

c) comprovante de que, na data da publicação da presente Resolução, ocupava, ou tinha exercício por período não inferior a 1 (um) ano, o cargo, a função ou o emprego referidos na alínea “a”;

III — certidão, passada pelo órgão de pessoal, do inteiro teor do ato de nomeação ou designação para o exercício de cargo, função ou emprego, contendo ainda indicação da publicação em órgão oficial de divulgação, bem como a afirmação de que, na data da publicação desta Resolução, ocupava, ou tinha exercido por período não inferior a 1 (um) ano, o cargo, a função ou o emprego para que fora nomeado ou designado.

IV — a) Carteira Profissional do Ministério do Trabalho, da qual conste, na data da publicação da presente Resolução, o registro da atividade profissional do interessado, na qualidade de Auxiliar de Estatístico, Agente de Estatística, Agente de Coleta ou equivalente, ou de que, até aquela data, houvesse permanecido no exercício dessa atividade por período não inferior a 1 (um) ano;

b) comprovante do órgão empregado confirmando os registros constantes da referida Carteira Profissional, bem como o efetivo exercício da atividade pelo interessado.

V — Carteira ou Certificado de registro no Ministério da Educação e Cultura, na condição de Professor habilitado na forma da legislação específica do ensino médio;

VI — a) ato original, individual ou coletivo, ou cópia autenticada, de nomeação, admissão ou contrato para o exercício do magistério de Estatística em estabelecimento de ensino médio;

b) recorte, original ou cópia autenticada, do órgão de divulga-

ção que publicou o ato, ou, na falta deste, declaração oficial de que surtiu os efeitos, como se publicado fosse;

c) comprovante de que, na data da publicação da presente Resolução, exercia, ou tinha exercido por período não inferior a 1 (um) ano letivo, o magistério para o qual fora nomeado ou contratado.

VI — Carteira Profissional do Ministério do Trabalho, de que conste anotação do exercício do magistério de disciplina de Estatística, na data da publicação da presente Resolução ou de que haja exercido o magistério dessa disciplina por período não inferior a 1 (um) ano letivo, acompanhada ainda de comprovantes do estabelecimento de ensino em que foi exercido o magistério.

§ 1.º — Os documentos de que trata este artigo deverão ser acompanhados, no que couber, de:

a) prova de quitação com o serviço militar;

b) título eleitoral;

c) prova de quitação com o imposto sindical;

d) prova de revalidação do respectivo diploma, de conformidade com a legislação em vigor, quando o requerente, brasileiro ou não, se tiver diplomado em curso técnico de Estatística, de nível médio, por instituto estrangeiro;

e) prova de que exercia, legitimamente, no País, atividade auxiliar de especialidade do Estatístico, na data da promulgação da Constituição de 1934, a qual desobrigará o estrangeiro da revalidação do diploma;

f) prova de permanência regular no País se estrangeiro;

g) requerimento assinado pelo interessado e dirigido ao Presidente do CONRE, solicitando o registro de que trata a presente Resolução, no qual serão consignados seu nome por extenso, a nacionalidade, a naturalidade, o estado civil, a residência, a data do nascimento, a filiação, o ano e o nome do estabelecimento em que concluiu o curso, se for o caso;

§ 2.º — O CONRE poderá exigir outros documentos esclarecedores julgados necessários à complementação da inscrição, além dos especificados neste artigo.

Art. 9.º — Aos diplomados como Técnico em Estatística, referidos no artigo 3.º desta Resolução, que ainda não possuírem o respectivo diploma registrado no Ministério da Educação e Cultura, será conferido registro provisório válido pelo prazo de 1 (um) ano.

§ 1.º — O prazo previsto nesse artigo poderá ser prorrogado, a critério do CONFE, mediante petição do interessado, até que possa ser apresentado o diploma, quando o registro será efetivado em caráter definitivo.

§ 2.º — Como comprovante do registro provisório, será expedido um Certificado, na forma estabelecida na Resolução n.º 21, de 23 de março de 1973, do CONFE, adaptado, ainda, ao disposto nos parágrafos 2.º, 3.º e 4.º, do artigo 2.º da Resolução n.º 20, de 12 de janeiro de 1973, do CONFE.

§ 3.º — Os documentos a serem apresentados, no ato do requere-

rimento, pelos candidatos referido neste artigo, serão:

a) certificado de conclusão de Curso de Técnico em Estatística, de nível médio, do qual devem constar, além de outros, os seguintes elementos:

- 1) data da diplomação
- 2) histórico escolar completo
- 3) assinaturas do Diretor do estabelecimento de ensino e do respectivo Secretário;

b) a documentação prevista no parágrafo 1.º do artigo 8.º da presente Resolução, no que couber.

Art. 10 — No cumprimento do que estabelece esta Resolução, quanto ao pagamento de taxas, emolumentos, anuidades e multas, aplicar-se-ão os dispositivos da Lei, do Regulamento, das Resoluções números, 10, 13, 20 e 22, bem como das Instruções números 1, 2, 4, 5 e 9, do CONFÉ, feitas as necessárias adaptações ao caso em espécie.

§ 1.º — Para efeito de pagamento das anuidades devidas pelos Técnicos em Estatística de Nível Médio registrados de acordo com o disposto nesta Resolução, observar-se-á o seguinte:

a) a cobrança de anuidade será efetuada a partir de 1975, inclusive;

b) a anuidade será devida:

1) a partir do ano seguinte, ao da diplomação pelos diplomados em Curso Técnico de Estatística, de nível médio, em estabelecimento de ensino oficial ou oficialmente reconhecido;

2) a partir da data do registro no Ministério da Educação e Cultura, como professores de Estatística de nível médio, pelos que o obtiverem nessa condição;

§ 2.º — As taxas, anuidades, multas e emolumentos referidos neste artigo, assim como outros tributos que vierem a ser estipulados, serão cobrados aos Técnicos em Estatística de Nível Médio na base de 50% (cinquenta por cento) dos valores ou percentuais já fixados ou que vierem a ser estabelecidos para os Estatísticos.

Art. 11 — O exercício de atividades auxiliares da especialidade de Estatístico, previstas no inciso XIX do artigo 31 do Regulamento aprovado pelo Decreto n.º 62.497, de 1.º de abril de 1968, reger-se-á pela Lei número 4.739, de 15 de julho de 1965, pelo citado Regulamento, pelo Regimento Interno do CONFÉ e pelo do CONRE, pela presente Resolução e pelos demais dispositivos que disciplinam o exercício da profissão de Estatístico.

Parágrafo único — A fiscalização do exercício de atividades auxiliares de especialidade do Estatístico compete ao CONFÉ e aos CONRE, em todo o território nacional, na forma do artigo 9.º da Lei n.º 4.739, de 15 de julho de 1965.

Art. 12 — Os Conselhos Regionais de Estatística (CONRE), tomarão providências efetivas no sentido de esclarecer os responsáveis pelos estabelecimentos de ensino médio sediados em sua jurisdição, quanto à obrigatoriedade do registro, por parte de recém-diplomados, quanto ao prazo estabelecido nesta Resolução e, ainda, quan-

to às penalidades legais a que estarão incursos aqueles que não cumprirem a legislação pertinente.

Art. 13 — O Conselho Federal de Estatística (CONFÉ) reexaminará todos os processos de pedido de registro profissional que, apreciados e julgados definitivamente, não tenham dado condição para o registro como Estatístico, com o objetivo de verificar a possibilidade de amparar os respectivos candidatos com o registro como Técnico em Estatística de Nível Médio.

§ 1.º — Para possibilitar a execução das medidas estabelecidas neste artigo, de maneira regular e uniforme, os CONRE, e suas Delegacias, nos Estados e Territórios, expedirão aviso aos interessados, no sentido de orientá-los quanto aos efeitos da presente Resolução e solicitar-lhes que, mediante petição, manifestem expressamente o interesse de obter registro profissional como Técnico em Estatística de Nível Médio.

§ 2.º — Os CONRE, antes de encaminharem os processos ao CONFÉ, verificarão os documentos deles constantes, tendo em vista o disposto no artigo 8.º desta Resolução, solicitando dos interessados a complementação dos documentos previstos.

§ 3.º — Cada processo, ao ser encaminhado ao CONFÉ, para reexame, deverá conter, em seguida a nova petição do interessado, o despacho do Presidente do CONRE,

em que se fará referência expressa de ter sido juntada documentação adicional caso de não ter esta sido necessária, nos termos dos parágrafos anteriores, deste artigo.

§ 4.º — Para o cumprimento deste artigo, os CONRE não emitirão parecer sobre a validade ou não dos documentos, incumbindo ao Conselho Federal a apreciação dos processos, em caráter de revisão especial, concedendo ou não registro aos interessados, na condição profissional de Técnico em Estatística de Nível Médio.

Art. 14 — Na execução do disposto nesta Resolução, aplicar-se-ão, no que couber, os dispositivos da Lei número 4.739, de 15 de julho de 1965, do Regulamento aprovado pelo Decreto n. 62.497, de 1.º de abril de 1968, das Resoluções e das Instruções baixadas pelo CONFÉ.

Parágrafo único — Os casos omissos serão resolvidos pelo Conselho Federal de Estatística (CONFÉ).

Art. 15. — Esta Resolução entra em vigor na data de sua publicação no Diário Oficial da União, revogadas as disposições em contrário.

Sala das Sessões, 11 de dezembro de 1974. — Anchizes do Egito Lopes Gonçalves, Presidente.

Aprovado na Sessão n.º 512 — Ordinária — de 11 de dezembro de 1974”.

SISTEMA ESTATÍSTICO NACIONAL

Os órgãos que integram o Sistema Estatístico Nacional, nos termos da Lei n.º 6.183, de 11 de dezembro de 1974, receberão orientação normativa do IBGE, sem prejuízo da subordinação administrativa a que estejam sujeitos.

Em cumprimento à referida Lei, a Fundação Instituto Brasileiro de Geografia e Estatística baixou a Instrução n.º 1, de 2 de janeiro de 1975, que a seguir transcrevemos:

“O Presidente da Fundação Instituto Brasileiro de Geografia e Estatística — IBGE —, no uso de suas atribuições,

Considerando que, nos termos do art. 2.º da Lei n.º 5.878, de 11 de maio de 1973, constitui objetivo básico do IBGE assegurar informações e estudos de natureza estatística, entre outros, necessários ao conhecimento da realidade física, econômica e social do País, visando especialmente ao planejamento econômico e social e à segurança nacional;

Considerando que, para consecução desse objetivo básico, o IBGE, consoante, o art. 3.º do aludido diploma legal, atua em diferentes áreas de competência, dentre as quais se incluem as seguintes:

I — estatísticas primárias (contínuas e censitárias);

II — estatísticas derivadas (indicadores econômicos e sociais, sistemas de contabilidade social e outros sistemas de estatísticas derivadas);

III — sistematização de dados sobre meio-ambiente e recursos naturais, com referência a sua ocorrência, distribuição e frequência;

Considerando que a atuação do IBGE se exerce mediante a produção direta de informações e a coordenação, a orientação e o desenvolvimento das atividades técnicas do Sistema Estatístico Nacional, definido pela Lei n.º 6.183, de 11 de dezembro de 1974;

Considerando que, na conformidade da citada Lei n.º 6.183, de 11 de dezembro de 1974, integram o Sistema Estatístico Nacional todos os órgãos e entidades da Administração Pública direta e indireta, de âmbito federal, estadual ou municipal e demais entidades de direito público e de natureza privada que também exerçam atividades estatísticas compreendidas nas áreas de competência definidas no art. 3.º, itens I, II e V, da Lei n.º 5.878, de 11 de maio de 1973, e para isso recebam subvenção ou auxílio dos cofres públicos;

Considerando, finalmente, o Plano Geral de Informações Estatísticas e Geográficas, aprovado pe-

lo Decreto n.º 74.084, de 20 de maio de 1974, resolve:

Art. 1.º — Os órgãos e entidades da Administração Pública direta e indireta, de âmbito federal, estadual ou municipal, e demais entidades de direito público e de natureza privada, que integram o Sistema Estatístico Nacional, definido pela Lei número 6.183, de 11 de dezembro de 1974, deverão encaminhar ao IBGE (Fundação Instituto Brasileiro de Geografia e Estatística — IBGE — Presidência — Avenida Franklin Roosevelt, 166 — ZC-39 — Caixa Postal 1753 — Rio de Janeiro, RJ), dentro do prazo de 90 (noventa) dias, a partir da publicação oficial deste ato, para fins cadastrais, as informações seguintes:

a) Denominação; constituição; subordinação ou vinculação, conforme o caso; sede e representação legal;

b) Atividades estatísticas exercidas e programadas nas áreas de competência definidas no art. 3.º, itens I, II e V, da Lei n.º 5.878, de 11 de maio de 1973;

c) Natureza e montante dos recursos financeiros, recebidos nos exercícios de 1972, 1973 e 1974, para atender as despesas decorrentes das atividades referidas na letra b, bem como os previstos para o ano de 1975.

Art. 2.º — Os órgãos e entidades a que se refere o art. 1.º deverão dar continuidade aos seus trabalhos, até que sejam expedidas normas operacionais específicas, visando a possibilitar ao IBGE exercer, na forma prevista na Lei n.º 6.183, de 11 de dezembro de 1974, a coordenação, a orientação e o desenvolvimento, em todo o território nacional, das atividades técnicas compreendidas na sua área de atuação.

Art. 3.º — O Presidente do IBGE baixará, consoante o art. 6.º, item VII, da Lei n.º 6.183, de 11 de dezembro de 1974, Instruções para coordenar as atividades do Sistema Estatístico Nacional e Normas Operacionais para orientar a execução dos trabalhos técnicos dos órgãos dela integrantes. — *Isaac Kerstenetzky*, Presidente.”

NORMAS TÉCNICAS PARA NUMERAÇÃO PROGRESSIVA DAS SEÇÕES DE DOCUMENTOS

Objetivando a uniformidade de critério quanto à numeração de documentos que não tenham sistematização própria, o IBGE passa

a adotar, para a NUMERAÇÃO PROGRESSIVA DAS SEÇÕES DE um DOCUMENTO *, as normas a seguir especificadas:

* Norma da Associação Brasileira de Normas Técnicas (ABNT), revisão da edição de 1964, transcrita de Ciência da Informação, v. 2, n.º 1, 1973, do Instituto Brasileiro de Bibliografia e Documentação (IBBD)

“1 — OBJETIVO

Esta Norma tem por objetivo descrever um sistema de numeração progressiva das partes de um documento de modo a permitir a exposição mais clara da matéria e a localização imediata de cada parte.

2 — CAMPO DE APLICAÇÃO

2.1 — Esta Norma aplica-se à redação de documentos submetidos à discussão, apreciação ou inquérito (documentos de trabalho em reunião, congressos, mesas redondas etc.: normas, especificações, regulamentos; parecer, relatórios, notas informativas etc.), assim como em documentos expositivos em geral (livros e documentos técnicos ou didáticos etc.) e as publicações estatísticas constituídas unicamente de tabelas.

2.2 — Esta Norma não é indicada para obras e documentos que tenham sistematização própria (dicionários, vocabulários etc.) ou ainda que não tenham necessidade de sistematização (romances, poesias).

2.3 — Na aplicação desta Norma não se deverá cair no exagero de subdividir demasiadamente as seções sacrificando, assim, a principal qualidade do sistema que é a concisão. Recomenda-se, para isso, não ultrapassar o máximo de seis algarismos, nem usar subdivisões além da quinária.

3 — DEFINIÇÕES

a) Seções¹: partes em que se divide o texto do documento, cada

uma delas contendo as matérias consideradas afins na exposição ordenada do assunto do texto;

b) Seções primárias: seções que resultam da primeira divisão do texto do documento (geralmente correspondente à divisão “capítulos”);

c) Seções secundárias, terciárias, quaternárias, quinárias: seções que resultam da divisão do texto de uma seção primária, secundária, terciária, quaternária, respectivamente;

d) Indicativo de uma seção: grupo numérico, constituído de acordo com as regras expostas em 4.2, que permite a localização imediata da seção a que se refere.

4 — NUMERAÇÃO E INDICATIVOS

4.1 — A presente seção primária (ou capítulo) trata da maneira de numerar as várias seções de um documento. Esta numeração progressiva não deve ser confundida com o sistema de classificação decimal, cuja estrutura é diferente.

4.2 — As seções primárias são numeradas consecutivamente, segundo a série natural dos números inteiros, a partir de 1, pela ordem de sua sucessão no documento. O indicativo de cada seção é o número que lhe foi atribuído, escrito em algarismos arábicos.

Exemplo: Os quinze capítulos em que foi dividido um documento são numerados consecutivamente de 1 a 15.

4.3 — As seções secundárias, terciárias, quaternárias, quinárias,

¹Outros termos são igualmente empregados na mesma acepção, tais como: parte, capítulo, item, tópico etc.

são numeradas consecutivamente seguindo a série natural dos números inteiros, a partir de 1, pela ordem de sua colocação na seção a que pertencem. O indicativo de uma seção secundária é constituído pelo indicativo da seção primária a que pertence, seguido do número que lhe for atribuído na sequência do assunto; com um ponto de separação repete-se o mesmo processo em relação as seções terciárias, quaternárias e quinárias.

Exemplo: Se a seção secundária 4.5 fosse dividida em 14 seções terciárias, teriam estas os indicativos: 4.5.1, 4.5.2. 4.5.9. 4.5.10. 4.5.14.

4.4 — O indicativo da seção precede imediatamente o título (ou a primeira palavra do texto, se a seção não tiver (título), e é separada dela por espaço ou hífen, preferivelmente. Na apresentação tipográfica destaca-se o indicativo em negrito (ou grifo).

4.5 — Leitura de um indicativo constituído por mais de um número faz-se lendo os números que o constituem pela ordem de sua colocação.

Exemplo: 4.5.14 lê-se “quatro cinco catorze”, o que significa que a seção em causa é a seção terciária n.º 14 da seção secundária n.º 5 da seção primária n.º 4 do documento.

4.6 — O texto de uma seção de um documento pode incluir vários parágrafos ou uma série ordenada de alíneas ou itens.

4.6.1 — Convém evitar a existência de parágrafos separados dentro de uma seção, sendo prefe-

rível considerá-los subdivisões da seção e numerá-los como tal.

4.6.2 — As alíneas incluídas numa seção caracterizam-se por meio de letras minúsculas do alfabeto latino (a, b, c, . . .), na ordem do alfabeto. A letra, seguida de um parêntese, procede imediatamente a primeira palavra do texto da alínea.

Exemplo: A seção 3 desta Norma.

4.6.2.1 — Se o número de alíneas for superior a 26, depois de esgotadas as letras duplas (a, b, c, x, y, z; aa, bb, cc, zz) ou grupos de duas letras (aa, ab, ac. az; ba, bb, bc, . . .; za, zb, zc, . . . zz).

4.6.2.2 — Quando a exposição da idéia assim o exigir a alínea pode ser subdividida em incisos com numeração romana maiúscula ou minúscula.

Exemplo: 2.3 a) 1 — 2.3. a) i

4.6.3 — Os itens incluídos numa seção caracterizam-se como subdivisões da respectiva seção.

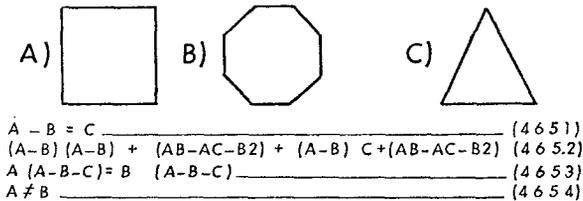
4.6.4 — Pode-se referenciar abreviadamente uma alínea ou item por meio do indicativo da seção em que está incluído, seguido imediatamente da letra ou número ordinal que o caracteriza, com a letra ou número entre parênteses.

Exemplo: A seção 3 inclui quatro alíneas que podem ser referenciadas abreviadamente por 3(a), 3(b), 3(c) e 3(d).

4.6.5 — Se o documento contiver figuras ou ilustrações, fórmulas ou tabelas, estas serão referenciadas de acordo com o próprio in-

dicativo da seção em que aparecem. Quando houver várias figuras ou ilustrações, fórmulas ou tabelas, dentro de uma mesma seção, são elas referenciadas quer como

alíneas, quer como incisos, quer ainda como subdivisões da respectiva seção, com a ressalva de se manter o mesmo critério em todo o trabalho.



4.7 — Os textos complementares, suplementares, finais ou afins, anexados ao documento, podem constituir uma seção primária em continuação, com seções secundárias correspondentes a cada um dos anexos.

no texto principal (ver 1). A sua localização deve ser assinalada no texto da redação provisória, onde deve figurar apenas o respectivo indicativo, seguido, entre parênteses, da indicação remissiva: “(ver p. . .)”.

Exemplo: Se este documento tivesse quatro anexos, teriam eles sucessivamente os indicativos 7.1 a 7.4.

5 — TÍTULOS

4.7.1 — Os anexos devem trazer, logo depois do indicativo próprio entre parênteses, o indicativo do texto a que se referem, precedido de “ver”.

5.1 — Os títulos das seções primárias, secundárias etc. são destacados gradativamente, seguindo-se as recomendações contidas no projeto em elaboração sobre apresentação de originais.

Exemplo: Se o documento anexado 7.2 se referisse à seção 4.1, seria indicado: 7.2 (ver 4 1).

5.1.1 — De um modo geral, não devem ser intituladas as subdivisões de uma seção sem título.

4.7.2 — Os documentos anexados a anexos constituem seções terciárias.

5.1.2 — O uso de título deve ser sistemático: dessa forma, deve-se evitar que, no mesmo documento, seções do mesmo nível ora tenham título, ora não.

Exemplo: Se o anexo 7.2 tivesse três anexos próprios, estes seriam indicados: 7.2.1, 7.2.2 e 7.2.3.

5.1.3 — Os títulos de todas as seções do mesmo nível de um documento devem ser escritos com o mesmo tipo e corpo.

4.7.3 — Quando um texto complementar destina-se a ser incorporado ao texto principal do documento, o seu indicativo deve corresponder à sua exata colocação

6 — PAGINAÇÃO

A fim de facilitar a organização de sumários parciais e a edição

independente das várias partes de um manual ou dos vários capítulos de um documento provisório, a numeração das páginas poderá ser feita individualmente por partes ou capítulos, a partir de 1. Neste caso, o número do capítulo deve

preceder o número de páginas, do qual deve ser separado por uma barra.

Exemplo: 1/1 a 1/17 — (entenda-se: cap. 1/p.1 a cap. 1/p.17)
3/30 a 3/58 — (entenda-se: cap. 3/p.30 a cap. 3/p.58)”.
3/30 a 3/58 — (entenda-se: cap. 3/p.30 a cap. 3/p.58)”.

RESENHA BIBLIOGRÁFICA

MARKS, Eli S.; SELTZER, William; KRÓTKI, Karol J. *Population Growth Estimation (PGE) A Handbook of vital Statistics Measurement*. New York, The Population Council.

1 — Os autores começam dando uma visão sumária dos problemas demográficos com que se depara a administração pública, de maneira cada vez mais acentuada, para fazer face aos programas de planejamento econômico-social. Entre os elementos mais importantes figuram as estatísticas de nascimentos e óbitos. Essas estatísticas são necessárias para a formulação de uma política de saúde. Além disso, elas contribuem com dados básicos, fundamentais para as projeções de população e essas, por sua vez, são essenciais ao estabelecimento dos planos de desenvolvimento econômico-social do país, permitindo conhecer a oferta

de mão-de-obra, a demanda de escolas, de recursos médicos, de novas moradias etc. Várias são as fontes através das quais se podem obter dados sobre fatos vitais, como componentes do crescimento da população: censos periódicos, censos especiais de áreas localizadas, um sistema nacional de registros dos fatos vitais*, um sistema de registro de população, esquemas de registro dos fatos vitais por amostragem, e vários tipos de amostragem domiciliares. Cada uma dessas fontes apresenta vantagens e limitações em termos da espécie e qualidade dos dados produzidos e do esforço requerido para sua aplicação com sucesso.

2 — A abordagem do "Population Growth Estimation" (de agora por diante denominada PGE) quando utilizada na medida ou na avaliação de estatísticas vitais, possui três características distintas:

* No Brasil isto é uma das funções do Registro Civil, realizada ainda com grandes deficiências embora esteja, no momento, a administração do IBGE fortemente empenhada na melhoria do sistema

a) o levantamento de registros dos eventos vitais por dois processos ou sistemas quase independentes de coleta;

b) o emparelhamento caso a caso dos registros nos dois sistemas para determinar quais os eventos que foram registrados apenas pelo primeiro, apenas pelo segundo e por ambos;

c) o preparo de uma estimativa do número total de eventos, corrigidos das omissões de ambos os sistemas, com base nas taxas de emparelhamento obtidas.

Como salientam os autores, essa definição é suficientemente ampla para incluir tanto os estudos tradicionais da cobertura do Registro Civil (por exemplo, o teste do registro de nascimentos realizado nos Estados Unidos em 1950), como os esforços realizados nas medidas demográficas básicas em estudos PGE realizados no Paquistão e na Turquia. No entanto, assinalam os autores, reconferências e reentrevistas do tipo frequentemente empregado nos censos e levantamentos em geral, para fins de controle e estimativas de cobertura, não constituem estudos do tipo PGE segundo a definição por eles adotada. Em primeiro lugar esse tipo de supervisão não é feita mediante "registro dual" no sentido em que essa expressão é empregada no livro em questão, uma vez que envolve o uso de processos de coleta que são altamente *dependentes*. Além disso, o principal objetivo da supervisão utilizada nesses levantamentos é o de limitar e corrigir os enganos do pessoal de campo mais do que proporcionar uma estimativa para um

ajustamento estatístico capaz de corrigir globalmente as omissões do processo de levantamento. Note-se que a coleção dual de registros independentes, seguida ou não de emparelhamento caso a caso não é suficiente para caracterizar uma abordagem PGE. O mesmo processo pode ser utilizado, por exemplo, quando, se deseja cruzar variáveis que figuram em um determinado arquivo com outras que se acham em outro arquivo. Sistema análogo utilizado para avaliar a qualidade das respostas dadas a um ou mais itens do questionário de um censo, levantamento por amostragem, ou sistema de registro. Assim, as certidões de óbito de um dado período posterior a um recenseamento da população podem ser emparelhadas com os registros censitários para os mesmos indivíduos, a fim de avaliar a qualidade da informação dos registros de óbitos (U.S. Nation Center for Health Statistics, 1969). De modo geral, a qualidade da informação censitária referente a determinados itens pode ser aferida mediante comparação das respostas obtidas no censo com as obtidas para os mesmos indivíduos em uma enumeração posterior, realizada para fins de controle (U.S. Bureau of the Census, 1960). Por outro lado, uma coleção dual está implícita em qualquer comparação de totais, por exemplo quando o número total dos nascimentos registrados é comparado com o número de crianças apuradas no censo da população com menos de um ano, corrigido da mortalidade e da subenumeração estimadas (Welp-ton, 1934), ou quando a população correspondente a uma coorte de nas-

cimentos é acompanhada em sucessivos recenseamentos (Barclay, 1958). Assim, a abordagem PGE implica necessariamente na existência simultânea das três características apontadas, o registro dual independente, o emparelhamento caso a caso e a estimativa do total ou das omissões.

3 — Para esclarecer bem o tipo da abordagem PGE imaginemos que se deseja estimar o total de peixes de um lago. Para isso são sorteados aleatoriamente, digamos, 100 pontos do lago e em cada um deles são pescados 100 peixes, num total de 10.000 peixes. Cada peixe, uma vez pescado, é marcado (mediante um grampo especial posto na cauda, na barbatana ou na boca) e em seguida solto no lago, no mesmo local em que foi pescado. Dias depois nova pescaria é feita nos mesmos locais da 1.^a. Seja n_1 o total de peixes da 1.^a pescaria (= 10.000 no exemplo suposto). Na segunda pescaria serão encontrados n_{12} peixes marcados, indicando a notação n_{12} que eles foram pescados em ambas as pescarias. Representamos por n_2 o total de peixes da 2.^a pescaria. Os n_{12} peixes de ambas as pescarias estão marcados ao passo que os demais não o estão. Seja finalmente n o total de peixes do lago. Esse total se subdivide assim em quatro categorias, isto é:

n_{12} : peixes pescados nas duas pescarias;

$n_1 - n_{12}$: peixes pescados só na 1.^a pescaria;

$n_2 - n_{12}$: peixes pescados só na 2.^a pescaria;

x : peixes que não foram pescados.

Assim, podemos colocar esses números em um quadro 2x2, com a seguinte disposição:

ESTIMATIVA DO NÚMERO DE PEIXES DE UM LAGO MEDIANTE DUAS PESCARIAS

1. ^a PESCARIA	2. ^a PESCARIA		TOTAL
	Pescado	Não Pescado	
Pescado ..	n_{12}	$n_1 - n_{12}$	n_1
Não Pescado.	$n_2 - n_{12}$	x	$n_2 - n_{12} + x$
TOTAL	n_2	$n_1 - n_{12} + x$	n

É claro que o total de peixes do lago será:

$$n = n_1 + n_2 - n_{12} + x$$

Assim, para estimar n é necessário estimar x , isto é, o total de peixes "omitidos" em ambas as pescarias. Se as pescarias forem independentes, pode-se escrever:

$$\frac{n_2 - n_{12}}{n_2} = \frac{x}{n_1 - n_{12} + x}$$

De onde resulta a estimativa de x

$$\hat{x} = \frac{(n_1 - n_{12})(n_2 - n_{12})}{n_{12}}$$

de modo que a estimativa do total de peixes será:

$$\hat{n} = n_1 + n_2 - n_{12} + \frac{(n_1 - n_{12})(n_2 - n_{12})}{n_{12}} = \frac{n_1 n_2}{n_{12}}$$

Para ilustrar, suponhamos que nas duas pescarias foram pescados 10.000 peixes, em cada, isto é:

$$n_1 = n_2 = 10.000$$

tendo sido encontrados, na segunda pescaria, 20 peixes marcados, ou seja:

$$n_{12} = 20$$

Obtém-se a seguinte estimativa do total:

$$\hat{n} = 19980 + \frac{9980 \times 9980}{20} = \frac{10000 \times 10000}{20} = 5000000$$

4 — De modo análogo, se considerarmos duas listas de n_1 e n_2 nomes, respectivamente, obtidas através de dois sistemas de registros independentes de determinadas ocorrências, podemos compará-las aos resultados de duas “pesquisas”, em que n_{12} corresponde ao número de ocorrências incluídas nas duas listas, $n_1 - n_{12}$ ao das que foram incluídas somente na 1.ª e $n_2 - n_{12}$ ao das ocorrências registradas somente na 2.ª lista. Assim, pode-se estimar, pelo processo anterior, o número x de ocorrências omitidas nas duas relações e, conseqüentemente, o número n , isto é, o total de ocorrências que se pretende conhecer. O aspecto do quadro será análogo ao do item anterior. Esse quadro acha-se indicado a seguir:

SISTEMA DE REGISTRO DUAL DE OCORRÊNCIAS

1.º SISTEMA	2.º SISTEMA		TOTAL
	Registradas	Omitidas	
Registradas.	n_{12}	$n_1 - n_{12}$	n_1
Omitidas	$n_2 - n_{12}$	x	$n_2 - n_{12} + x$
TOTAL	n_2	$n_1 - n_{12} + x$	n

Quando se trata de obter, além de uma estimativa de n , uma medida da eficiência dos sistemas de registro (ou de um deles) é importante conhecer o grau de cobertura, medido para cada sistema pela

relação denominada “taxa de cobertura”. Assim tem-se as seguintes taxas:

$$\text{Para o 1.º sistema } c_1 = n_{12}/n_2$$

$$\text{Para o 2.º sistema } c_2 = n_{12}/n_1$$

É claro que, conhecidas as taxas de cobertura, pode-se estimar n a partir das expressões:

$$\hat{n} = (n_1/c_1) = (n_2/c_2)$$

Segundo salientam os autores, as percentagens de cobertura do registro de crianças até 3 meses de idade, no censo americano de 1950, foi de 98%, no conjunto, atingindo 99% na zona urbana e 93% na zona rural (Tabela 2.2, pág. 25).

5 — No esboço histórico (Historical Review) que constitui a parte C do capítulo 2, os autores descrevem alguns estudos de tipo PGE realizados desde 1930 em três tipos diferentes de regiões econômico-geográficas, isto é:

a) Canadá, Estados Unidos e União Soviética;

b) Ásia (principalmente na Índia), e

c) África, América Latina e Caraíbas.

Nesse esboço histórico eles descrevem minuciosamente alguns dos estudos realizados, apresentando uma ampla informação sobre os coeficientes de cobertura e outros resultados obtidos. Em seguida fazem uma apreciação sobre outras aplicações relacionadas com esse tipo de técnica dual, detendo-se, em especial, nos estudos denominados CTR (sigla de “Capture-

Tagged-Recapture”) destinados a estimativas de populações animais. A esse sistema de “captura-marcação-recaptura” já nos referimos anteriormente como ilustração do próprio sistema dual. Os autores chamam a atenção para a possibilidade de ocorrência de um “bias” na estimativa das populações animais sempre que a natalidade, a mortalidade e/ou as correntes migratórias, no intervalo entre a captura e a recaptura, não sejam desprezíveis, e bem assim, quando a mistura dos animais marcados com os demais, durante o mesmo intervalo, não for completa ou quando as probabilidades de recaptura se alterarem em virtude do processo de marcação etc.

6 — O capítulo 3.º trata da escolha das técnicas apropriadas para o desenho eficiente de um estudo PGE, focalizando, particularmente, um balanço adequado entre qualidade e custo. A discussão é amplamente desenvolvida no capítulo 3.º, dedicando o capítulo 4.º à apresentação dos procedimentos práticos correspondente a três abordagens alternativas do problema. Salientam os autores, no capítulo 3.º que, idealmente, um procedimento de estimativa deve ser tal que conduza à maximização do valor atual líquido do estudo, definido como diferença entre o valor da informação produzida e o custo de produzi-la, isto é, o custo envolvido no processo de estimação, tais como, coleta e processamento dos dados (inclusive o emparelhamento, correção dos erros e tabulação), além dos custos gerais (planejamento, amostragem, desenho do questionário, supervisão etc.). O

valor da informação, mais difícil de definir e de medir, é amplamente discutido, considerando a possibilidade de uma otimização por assim dizer gradativa, isto é, uma vez determinado o ótimo em relação a certas variáveis, verificar a possibilidade de obter uma melhor solução mediante alteração das variáveis de modo a reduzir o custo e/ou aumentar o valor da informação. Salientam os autores que esse processo envolve um certo grau de arbitrariedade, como ocorre em todo processo complexo de otimização. Como medida da qualidade da informação, uma vez definido o seu campo, os autores recomendam o erro médio quadrático da estimativa \mathcal{E} , definido como sendo o valor cujo quadrado é a soma da variância \mathcal{G}^2 , e do quadrado do “bias” (erro sistemático), \mathcal{S}^2 , de modo que \mathcal{E} será a hipotenusa de um triângulo retângulo A, B, C cujos catetos são \mathcal{G} e \mathcal{S} (Fig. 1).

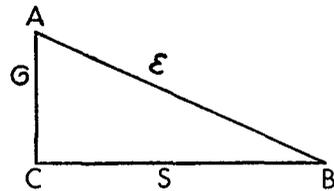


Fig. 1

O processo de otimização começa, pois, por verificar, dentre os procedimentos permissíveis com o orçamento disponível qual o que torna mínimo o valor de \mathcal{E} . Se esse mínimo ainda for considerado excessivo cabe verificar, mediante alterações do projeto, se o acréscimo do “valor adicional líquido” da informação proporcionado por essa redução do erro médio quadrático da estimativa justifica o eventual

acréscimo de despesas. Se com referência à variância é possível uma razoável precisão, algumas componentes do erro sistemático (bias) só podem ser determinadas, frequentemente, em ordem de grandeza. Seltzer e Adlaka (1969), definem três componentes do erro sistemático, isto é:

- (1) efeito devido à correlação entre respostas;
- (2) erro do emparelhamento, e
- (3) erro cobertura (outofscope error)

Analogamente, são definidas três componentes da variância:

- (1) a variância da amostragem;
- (2) a variância do erro de resposta, e
- (3) a variância do processo de emparelhamento.

Os autores fazem uma ampla análise dos métodos alternativos de medida (levantamentos domiciliares, censos, registros existentes etc.) seguida de uma exposição minuciosa sobre as componentes da variância e do erro sistemático. Em particular analisam minuciosamente as componentes mais importantes dos dois tipos de erros: os devidos à influência de fatores aleatórios (variância) e à de fatores não aleatórios (bias). Esse último tipo de erro pode resultar de deficiências no desenho da amostra, do processo de coleta e do método de estimação utilizado. Muitas vezes aceita-se um pro-

cesso de estimação com pequeno "bias" com o objetivo de conseguir uma substancial redução da variância, como ocorre, por exemplo, com a adoção de "estimativas de razão" (ratio estimates). Há certos processos de coleta que são reconhecidamente afetados por erro sistemático. Por exemplo, a estimativa dos filhos tidos por idade da mãe proporcionam frequentemente resultados subestimados da fecundidade feminina acima de 30-35 anos (United Nations, 1967, p. 31) assim como o levantamento retrospectivo único tende a ser particularmente deficiente para a obtenção de taxas de mortalidade (Seltzer, 1969). O erro sistemático pode ainda resultar de numerosos fatores durante o processo de coleta, e são provenientes do agente de coleta, da pessoa inquerida, dos conceitos e definições estabelecidos, dos registros utilizados etc. É difícil a medida exata do erro sistemático, em primeiro lugar por que se desconhece o verdadeiro valor estimado e, além disso, porque frequentemente a demarcação entre uma resposta errada por erro sistemático ou por erro aleatório é algo arbitrária. De acordo com a figura 1, resulta:

$$e_x^2 = G_x^2 + S_x^2$$

onde x indica a estimativa a que estão referidos os erros. Assim, uma estimativa será de má qualidade quando G_x^2 ou S_x^2 ou ambos forem grandes. Devem ser feitos esforços no sentido de reduzir tanto G_x^2 como S_x^2 .

7 — Cabe ter presente que a legitimidade da estimativa de n dada pela expressão (3-1) decorre

da suposição de *independência* dos dois registros. Assim, se os registros dos sistemas utilizados forem:

Sistema 1 — Registro Civil

Sistema 2 — Censo, Pesquisa domiciliar etc.

a independência exigida significa que a probabilidade de uma ocorrência (morte, nascimento etc.), ser registrada por um dos sistemas não se modifica pelo fato de que a mesma ocorrência seja ou não registrada pelo outro sistema. Isso implica naturalmente, face a um conhecido teorema do cálculo das probabilidades, em que a probabilidade p_1 de que uma ocorrência seja registrada pelo sistema 1 e a probabilidade p_2 de que a mesma ocorrência seja registrada pelo sistema 2, permite calcular a probabilidade p_{12} de que seja registrada *por ambos os sistemas* a partir da fórmula:

$$p_{12} = p_1 \cdot p_2$$

A essa altura, esclarecem os autores, é possível fazer, de uma maneira mais completa, um resumo das fontes de “bias” e de variância. Com referência ao primeiro tipo de erro, reportamo-nos à classificação de Seltzer e Adlaka.

O erro (bias) devido à correlação entre respostas resulta do fato de não ser satisfeita a equação anterior devido à não independência dos dois sistemas de registro utilizados. De fato, em consequência disso, os eventos *incluídos* no sistema 1 não têm a mesma probabilidade de serem incluídos no sistema 2, que os eventos *não incluídos* no 1.º. Quanto aos erros

de emparelhamento podemos considerar dois tipos: os *emparelhamentos* errôneos, isto é, ocorrências diferentes que foram confundidas e dadas erradamente como emparelhadas e os *não emparelhamentos* errôneos, isto é, ocorrências que deveriam ser emparelhadas e que por engano deixaram de sê-lo. Desde que não haja exata compensação, ocorrerão erros sistemáticos (bias) de maior ou menor magnitude. Com relação ao “bias” decorrente de erro de cobertura eles provêm da inclusão indevida de fatos em qualquer dos sistemas, ou em ambos, ou, ainda, da não inclusão de fatos que deveriam ser incluídos.

Com relação à variância podem ser citadas as seguintes fontes:

1) variância de amostragem cuja origem é comum a todo processo de amostragem;

2) erros de resposta de um tipo especial, que apresenta um caráter nitidamente aleatório. Mesmo que não esteja envolvida a variação inerente ao processo de amostragem, os próprios fatos ocorridos no universo variam de uma época para outra em torno de certo valor médio. Assim a taxa de mortalidade de uma região apresenta variações anuais de caráter aleatório devido ao fato de que um determinado ano constitui apenas uma única observação de um processo estocástico, isto é, de uma família $X(t)$ de variáveis aleatórias dependentes de um parâmetro contínuo, o tempo. Para um dado valor t_1 de t , isto é, em um dado instante em que o levantamento é feito, o número de ocorrências

acumuladas até então constitui uma determinação empírica da variável aleatória $X(t_1)$ sujeita, portanto, a uma variação aleatória;

3) os erros de emparelhamento não resultarão em erro sistemático se o número esperado de emparelhamento indevidos for igual ao de não emparelhamento errados. A diferença efetiva entre esses dois tipos de erro em uma particular aplicação das regras de emparelhamento é uma particular determinação da variável cuja esperança matemática é precisamente o erro sistemático (bias).

8 — A variância de amostragem pode ser reduzida mediante técnicas convencionais, tais como o aumento do tamanho da amostra, a estratificação e a redução do tamanho dos conglomerados. Essa redução da variância de amostragem é geralmente acompanhada de aumento de custo, embora não na mesma proporção. Infelizmente a redução da variância de amostragem é muitas vezes, acompanhada de um acréscimo do erro sistemático (bias) — por exemplo, quando aquela redução decorre de um aumento do tamanho da amostra conseguida às custas de uma deterioração da qualidade da coleta. Em última análise o que interessa é a redução do erro médio quadrático da estimativa que é uma soma de numerosas componentes que integram a variância e o “erro sistemático”. O problema crítico consiste pois em saber como obter uma redução de *erro total*, para uma dada despesa. Nesse particular os problemas mais sérios resultam do erro de emparelhamento e do alto custo do em-

parelhamento, muito embora, julgam os autores, as objeções nesse sentido tenham sido em geral um pouco exageradas, esquecendo-se muitas vezes, outras vantagens do sistema. Por exemplo a estimativa de \hat{n} através da expressão $n_1 n_2 / n_{12}$ apresenta a característica de uma maior constância no tempo, porque n_{12} tende a variar no mesmo sentido de n_1 ou n_2 . Para ilustrar supõem os autores que num determinado ano a cobertura da fonte 2 decresceu, digamos, de 10%, enquanto a da fonte 1 permaneceu praticamente no mesmo nível. Se a correlação entre as fontes permanecer baixa e constante, o valor esperado da taxa de emparelhamento da fonte 1 (n_{12}/n_1) que é a estimativa da cobertura da fonte 2, também aumentará de uns 10% e a esperança matemática de \hat{n} permanecerá a mesma. Ainda que a cobertura dos dois sistemas variem consideravelmente no tempo, o efeito dessa variação sobre as estimativas de variações das taxas de natalidade e mortalidade é muito menor em um sistema dual do que em um sistema simples.

9 — Em seguida os autores desenvolvem certos pormenores relacionados com os problemas decorrentes da não independência das duas fontes. Em primeiro lugar salientam, por exemplo, que quanto à natureza dos eventos e sistemas de registro, algumas ocorrências tendem a ser omitidas por ambas as fontes porque ocorrem em populações ou subgrupos de populações difíceis de enumerar e registrar, tais como domicílios de pessoas de “status” econômico muito baixo ou que vivem em regiões

muito distantes. Em segundo lugar, pode haver uma correlação devida à comunicação entre os sistemas de coleta. Um caso externo de intercomunicação ocorre, por exemplo, quando o pessoal de uma das fontes informantes recebe a lista dos eventos registrados na outra com a solicitação de completá-la com as ocorrências omitidas na lista. A correlação entre os dois sistemas tende sempre a ser positiva, embora possa haver certos tipos de influência que tendem a operar em sentido contrário. Assim o valor de n tende a ser sempre subestimado. A correlação de resposta, ρ , é dada pela expressão, indicada pelos autores:

$$\rho = \frac{p_{12} - p_1 p_2}{\sqrt{p_1 q_1} \sqrt{p_2 q_2}}$$

onde:

p_1 — é a probabilidade de o evento ser registrado pelo sistema 1;

p_2 — é a probabilidade de o evento ser registrado pelo sistema 2;

p_{12} — é a probabilidade de o evento ser registrado em ambos os sistemas;

$$q_1 = 1 - p_1; q_2 = 1 - p_2$$

O erro devido apenas à correlação de respostas será dado então pela fórmula

$$B = - \frac{\rho \sqrt{q_1 q_2}}{\rho \sqrt{q_1 q_2} + \sqrt{P_1 P_2}}$$

por onde se verifica o resultado de que, sendo ρ positivo (como ocorre geralmente) B será negativo e conseqüentemente \hat{n} constituirá

uma subestimativa de n . Esse tipo de erro pode ser reduzido mediante um processo de estratificação, estimando-se o \hat{n} relativo a cada estrato e somando-se as estimativas obtidas. Do ponto de vista prático, porém, existe uma certa dificuldade na escolha adequada da variável de estratificação. Os autores desenvolvem bastante esse aspecto apresentando a experiência obtida nos diferentes estudos disponíveis já referidos no capítulo 2. Deve-se notar, no entanto, que se a divisão da amostragem for feita em sub-grupos muito pequenos os resultados da estratificação podem vir a ser contraproducentes pela introdução de novo "bias", que pode ser predominante. O capítulo 3 que compreende quase 100 páginas trata ainda, com grande desenvolvimento, dos erros de emparelhamento (cerca de 16 páginas), do "bias" de cobertura e dos erros de amostragem, dos trabalhos de campo, da estimativa da população de base, do emprego de estimativas de razão, estimativas por sub-grupos etc., todos esses tópicos comportando grande amplitude e profundidade.

10 — No capítulo 4.º os autores desenvolvem três abordagens alternativas para pesquisa de fatos vitais:

1) sistema de registro civil combinado com uma pesquisa domiciliar periódica;

2) sistema especial de registro combinado com uma pesquisa domiciliar periódica;

3) uma pesquisa domiciliar periódica com substancial superposição nos diferentes períodos e

enumeração independente em etapas sucessivas.

O capítulo, com 100 páginas, cogita inicialmente, com grande minúcia, dos procedimentos comuns às três alternativas, tais como mapeamento, endereçamento e numeração das unidades, levantamentos domiciliares periódicos, conteúdo dos questionários, emparelhamento, tabulações etc. Em seguida examina as diferentes fases de cada uma das alternativas sugeridas, com a indicação de numerosos pormenores da execução, e, bem assim, das fórmulas matemáticas para a estimativa da cobertura, da variância das estimativas etc. O capítulo 5.º é todo dedicado ao estudo de caso descrevendo um estudo PGE em um país hipotético no qual se incorporam a experiência real e as condições de numerosos estudos realizados em vários países. Trata-se pois de um plano ideal baseado porém nas condições efetivas do mundo real. Nesse estudo são novamente passadas em revista, sob uma forma mais objetiva, todas as fases do processo, da definição do problema dos seus objetivos, da população utilizada, da seleção da alternativa a adotar, do sistema de registro do país hipotético, da amostra, do processo de emparelhamento etc.

O capítulo 6 trata dos resultados do estudo PGE e das implicações da avaliação, a fim de manter e, se necessário, implementar a precisão e a utilidade das estimativas das estatísticas vitais. Os autores fazem uma distinção muito nítida entre as funções de supervisão e avaliação, as quais, devem ser realizadas por pessoal diferente em níveis operacionais diferen-

tes. Os supervisores devem estar em estreito contato diário com a coleta de dados e com as operações de processamento, enquanto que aqueles que fazem o trabalho de avaliação devem, ao contrário, estar suficientemente isolados das operações diárias a fim de facilitar a aferição dos diferentes tipos gerais de erros, mais do que propriamente a performance de determinados agentes recenseadores, oficiais de registro ou funcionários administrativos. Os autores desenvolvem essa matéria com grande amplitude e particular minúcia, abordando todos os pontos importantes. Finalmente, o capítulo 7 dedica cerca de 80 páginas e tópicos especiais que vão desde a dedução matemática, além das que figuram no próprio texto do livro, das fórmulas das variâncias correspondentes a diferentes alternativas tais como amostragem em ambas as fontes, amostragem só na fonte 2 com todo o universo na fonte 1, efeito da correlação entre o tamanho da amostra por área e a cobertura do registro, amostras estratificadas etc. Além disso analisa matematicamente diferentes tipos de estratificação, de subamostragem. Deduz e analisa também, as fórmulas para o caso de mais de dois sistemas de coleta estabelecendo, no caso de três fontes, as fórmulas propostas por Das Gupta (7.109, pág. 404). Essa fórmula, porém não leva em conta a informação sobre a correlação fornecida pelo emparelhamento triplo. Assim, a equação de Das Gupta corresponde à máxima verossimilhança, isto é, conduz a uma estimativa que maximiza a probabilidade de obter o resultado especificamente

obtido na suposição de que as três fontes são independentes, ignorando assim qualquer relação do método de estimação com os efeitos do emparelhamento ou dos erros de cobertura. Os autores propõem um novo processo de estimativa que permitirá levar em conta a possível correlação entre as fontes, adotando uma combinação das várias células do quadro (fórmula 7.116 e 7.117, pág. 405). Essa fórmula fornece para estimativa dos eventos omitidos em todas as fontes em um exemplo dado, o valor

$$0,0148N$$

onde N é o total de eventos ocorridos no universo e como estimativa \hat{n} o valor

$$0,9823N$$

contra $0,9803N$ fornecido pela fórmula de Das Gupta. O "bias" é ligeiramente inferior ao da estimativa de máxima verossimilhança embora a variância seja ligeiramente maior. Como resultado final, o erro médio quadrático das duas estimativas será praticamente o mesmo. Quando existe um apreciável "bias" de correlação os autores sugerem a utilização das estimativas produzidas apenas por três das células adequadamente es-

colhidas, obtendo para estimativa de n um valor \hat{n} cuja esperança matemática, no exemplo dado, é:

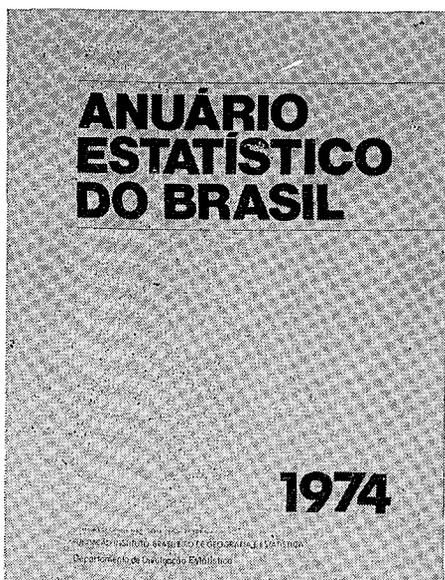
$$0,992N$$

apresentando assim uma redução sensível do "bias". Muitos outros tópicos importantes são desenvolvidos no capítulo 7 ao qual se seguem, como em todos os demais capítulos, importantes notas adicionais (Footnotes). Entre este último capítulo e a extensa bibliografia fornecida pelos autores existe um Glossário de termos e conceitos com 151 verbetes, com amplas explicações.

O livro tem dimensões 26 x 20 cm e 481 páginas de texto inclusive o Glossário (20 p. apr.), as referências bibliográficas (13 p. apr.) e o índice (16 p. apr.) Trata-se como se vê, de um livro volumoso que analisa os problemas das estimativas de tipo dual sob todos os seus aspectos, e apresenta uma vasta experiência internacional sobre a matéria. Cremos poder dizer que é o melhor e mais completo trabalho escrito até o presente sobre o método dual de estimativa de estatísticas vitais, importantes para o bom conhecimento da demografia de qualquer país e, em particular, no Brasil.

Prof. Lyra Madeira

ANUÁRIO ESTATÍSTICO DO BRASIL — 1974



O IBGE vem de entregar ao público consumidor de estatísticas e estudiosos em geral, o ANUÁRIO ESTATÍSTICO DO BRASIL, em seu XXXV volume, correspondente ao ano de 1974.

Referida publicação constitui a mais importante contribuição da entidade à divulgação das estatísticas brasileiras, e sua elaboração, dentro do mais rigoroso critério técnico, integra o plano de oferecer aos usuários de dados estatísticos, com apreciável atualização, os principais resultados dos levantamentos estatísticos realizados no País, para cuja efetivação contou com a colaboração de todos os ór-

gãos que formam o sistema estatístico nacional.

O presente volume, com 960 páginas, rico em gráficos e tabelas representativos dos aspectos que enfoca, se constitui num valioso repositório de dados estatísticos, capaz de interessar a todos os setores econômicos e sociais, industriais e financeiros do País, pois retrata com fidelidade os seus principais campos de atividade: silvicultura, pesca, agropecuária, indústria, comércio interior e exterior, seguros, serviços, transportes, comunicações, mercado monetário e financeiro, propriedade, consumo e contas nacionais.

Cumprе ressaltar a apresentação de informações elaboradas pelos diversos setores do IBGE, tais como índices da produção real das Indústrias de Transformação, Armazenagem e Silagem, Produção Extrativa Vegetal, Agropecuária e Pesquisa Industrial.

O Anuário de 1974 apresenta, igualmente, os dados atualizados sobre Situação Física, Demográfica, Social e Cultural, bem como Administrativa e Política, além dos resultados definitivos dos Censos Demográficos, Industrial e Predial, e os dados preliminares do Censo Agropecuário.

**PUBLICAÇÕES EDITADAS PELOS ÓRGÃOS
DE ESTATÍSTICA DO IBGE NO TRIMESTRE
JANEIRO-MARÇO DE 1975 ***

**CENTRO DE DOCUMENTAÇÃO E INFORMAÇÃO ESTATÍSTICA
— CENDIE**

016:3(81)

BOLETIM BIBLIOGRÁFICO —
ÍNDICE 1970/1971. Rio de Ja-
neiro, 1974. 446 p.

014.3:05

SUMÁRIOS DE PERIÓDICOS
CORRENTES EM ESTATÍSTI-
CA E ECONOMIA. Rio de Ja-
neiro, v. 2, n. 2, dez. 1973.

DEPARTAMENTO DE DIVULGAÇÃO ESTATÍSTICA — DEDIVE

Obras

31

ROCHA, Marcos Vinicius da. Cur-
so de estatística. 3 ed. rev. e

ampl. Rio de Janeiro, 1975. 248
p., il.

Periódicos

31(81) (05)

ANUÁRIO ESTATÍSTICO DO
BRASIL 1974. Rio de Janeiro, v.
35, 1975.

311(81) (05)

REVISTA BRASILEIRA DE ES-
TATÍSTICA. Rio de Janeiro, v.
35, n. 138, abr./jun. 1974 Tri-
mestral.

Séries

911.6(813 22 CRA)

Crateús, Ceará. Texto de Rilza F.
Saldanha. Rio de Janeiro, 1975.
20 p., il. (Coleção de Monogra-
fias, 576)

911 6(816.52 ARR GRA)

Arroio Grande, Rio Grande do Sul.
Texto de Rilza F. Saldanha. Rio
de Janeiro, 1975. 20 p., il. (Cole-
ção de Monografias, 578)

* Bibliografia preparada no Setor de Bibliografia do Centro de Documentação e Informação Estatística do IBGE

911.6(815.12 DIA)
Diamantina, Minas Gerais. Texto de Daisy Costa Lima. Rio de Janeiro, 1975. 32 p., il. (Coleção de Monografias, 580)

911.6(815.32 TRÊS RIOS)
Três Rios, Rio de Janeiro. Texto de Umberto Cognac. Rio de Janeiro, 1975. 28 p., il. (Coleção de Monografias, 582)

911.6(816.12 ITU)
Itu, São Paulo. Texto de Umberto Cognac. Rio de Janeiro, 1975. 36 p., il. (Coleção de Monografias, 586)

911.6(816.52 MON)
Montenegro, Rio Grande do Sul. Texto de Maria de Lourdes F. Cianella. Rio de Janeiro, 1975. 24 p., il. (Coleção de Monografias, 574)

911.6(817.22 RON)
Rondonópolis, Mato Grosso. Texto de Rilza F. Saldanha. Rio de Janeiro, 1975. 25 p., il. (Coleção de Monografias, 575)

911.6(817.32 RIO VER)
Rio Verde, Goiás. Texto de Maria de Lourdes F. Cianella. Rio de Janeiro, 1975. 28 p., il. (Coleção de Monografias, 577)

DEPARTAMENTO DE ESTATÍSTICAS INDUSTRIAIS, COMERCIAIS E DE SERVIÇOS — DEICOM

31:66/69(81)
Indústria de transformação. Pesquisa mensal — janeiro/setembro de 1974. Rio de Janeiro, 1974. 48 p., tab. Mimeografado.

—— — *janeiro/out. de 1974*. Rio de Janeiro, 1975. 48 p., tab. Mimeografado.

31:69(81)
Indústria da construção. Inquérito mensal sobre edificações — julho de 1974. Rio de Janeiro, 1974. 257 p., tab.

—— — *agosto de 1974*. Rio de Janeiro, 1974. 255 p., tab.

—— — *setembro de 1974*. Rio de Janeiro, 1974. 251 p., tab.

338.5:31(81)
——. *Preços de material de construção no comércio ataca-*

disto. Salários na indústria da construção — abril a setembro de 1974. Rio de Janeiro, 1974. 317 p., tab.

——. *julho a dezembro de 1974*. Rio de Janeiro, 1975. 323 p., tab.

338.5:31(81)
Inquérito nacional de preços. Gêneros alimentícios. Comércio varejista das Capitais — 1972 a março de 1974. Rio de Janeiro, 1974. 58 p., tab.

—— — *1972 a abril de 1974*. Rio de Janeiro, 1974. 60 p., tab.

—— — *1972 a maio de 1974*. Rio de Janeiro, 1974. 60 p., tab.

—— — *1972 a junho de 1974*. Rio de Janeiro, 1974. 60 p., tab.

- — 1973 a julho de 1974. 381(814.2)
Rio de Janeiro, 1974. 60 p., tab. *Comércio interestadual. Exportação por vias internas — Bahia*
—— — 1973 a agosto de 1974. 1972. Rio de Janeiro, 1974. 92 p.,
Rio de Janeiro, 1974. 60 p., tab. tab. Mimeografado.

DEPARTAMENTO DE CENSOS — DECEN

- 311.213.1:63(813.5)
Censo Agropecuário — Alagoas.
Rio de Janeiro, 1975. XXXVIII
+ 319 p., tab., questionário em
anexo. (VIII Recenseamento Ge-
ral — 1970. Série Regional v. 3,
t. 11)
- 311.213.1:63(814.1)
—— — *Sergipe.* Rio de Janeiro,
1974. XXXVIII + 321 p., tab.,
questionário em anexo (VIII
Recenseamento Geral — 1970.
Série Regional v. 3, t. 12)
- 311 213.1:63(815.2)
—— — *Espírito Santo.* Rio de
Janeiro, 1974. XXXVIII + 240
p., tab., questionário em anexo
(VIII Recenseamento Geral —
1970. Série Regional v. 3, t. 15)
- 311.213.1:63(815.3)
—— — *Rio de Janeiro.* Rio de
Janeiro, 1974. XXIX + 333 p.,
tab., questionário em anexo
(VIII Recenseamento Geral —
1970. Série Regional v. 3, t. 16)
- 311.213.1:63(816 4)
—— — *Santa Catarina.* Rio de
Janeiro, 1975. XLI + 482 p.,
tab., questionário em anexo
(VIII Recenseamento Geral —
1970. Série Regional v. 3, t. 20)

CENSO AGROPECUÁRIO

Alagoas, Sergipe, Santa Catarina, Rio Grande do Sul e Distrito Federal

Prosseguindo na divulgação dos resultados definitivos dos levantamentos censitários que integraram o Recenseamento Geral — 1970, o IBGE vem de divulgar os Tomos referentes aos Estados de Alagoas, Sergipe, Santa Catarina,

Rio Grande do Sul e ao Distrito Federal, relativos ao Censo Agropecuário nas referidas Unidades, e cuja coleta de informações foi iniciada no segundo semestre de 1971.

Os referidos resultados censitários assinalam, além do confron-

to de dados dos censos de 1920, 1940, 1950, 1960 e 1970, com exceção do Distrito Federal, os seguintes aspectos: Propriedade das terras, segundo a sua condição legal de ocupação e grupos de área total; condição do produtor, forma de administração, utilização, irrigação, atividade econômica, pessoal ocupado, segundo a categoria e sexo; grupos de cabeças de bovinos e de áreas de lavoura; emprego de força nos trabalhos agrários, uso de fertilizantes, máquinas e instrumentos agrícolas; veículos, silos e depósitos; valor dos bens, dos investimentos e dos financiamentos obtidos; despesas;

combustíveis consumidos; produção animal e vegetal; valor da produção agrícola, colheita, efetivo das plantações, segundo a condição do produtor, atividade econômica, grupos de área etc.

A apresentação dos dados é efetuada através de um conjunto de cento e nove tabulações, sendo cinquenta e sete a nível de Unidade da Federação e cinquenta e duas a nível de Microrregiões Homogêneas e Municípios.

É oportuno destacar que, pela primeira vez nos censos agropecuários nacionais, são divulgadas informações sobre o valor da produção.

Todas as publicações do IBGE, inclusive as referentes as notas insertas nesta Revista, poderão ser adquiridas na sede da Instituição, à Avenida Franklin Roosevelt, 146 — loja, Rio de Janeiro — RJ, ou nas respectivas Delegacias de Estatística, nas demais Unidades da Federação.

RBE: ÍNDICE ALFABÉTICO E REMISSIVO

Ano 1974

INTRODUÇÃO

O presente índice, elaborado no Setor de Bibliografia do Centro de Documentação e Informação Estatística do IBGE, abrange a matéria publicada nos números 137 a 140 da Revista Brasileira de Estatística, correspondente a 1974.

Conforme a norma brasileira para a citação de artigos de periódicos (NB-66, in Normalização da Documentação no Brasil, 2. ed., Rio de Janeiro, Associação Brasileira de Normas Técnicas, 1964, p. 93-116), o número do fascículo precede o das páginas inicial e final do artigo, separado das mesmas por dois pontos (:). Ex. 114:149-57 = fascículo 114, p. 149 a 157.

FASCÍCULOS PUBLICADOS E PERÍODOS DE REFERÊNCIA

Ano XXXV — 137 — jan./mar. 1974

Ano XXXV — 138 — abr./jun. 1974

Ano XXXV — 139 — jul./set. 1974

Ano XXXV — 140 — out./dez. 1974

AGROPECUÁRIA

estatística da produção
legislação (137):125

AGRICULTURA

desenvolvimento agrícola
características
América Latina (139):454-63

ALIMENTAÇÃO

consumo alimentar
Estudo Nacional da Despesa Familiar
pesquisa (139):445-53
Reunião CEPAL/FAO
estudo apresentado
América Latina (140):587-606

AMAZÔNIA

bibliografia (140):607-37

AMÉRICA LATINA

agricultura
desenvolvimento (139):454-63

alimentação

Reunião CEPAL/FAO
estudo apresentado (140):587-606

Conferência Latino-americana de Industrialização
notícia (140):584-6

ANO MUNDIAL DE POPULAÇÃO

notícia (139):441-2

BIBLIOGRAFIA (137):83-121; (138):248-60; (139):468-80; (140):607-48

Bremaeker, François Eugene Jean de
(139):271-8

Caetano, Maria Mônica P. Barbosa
Vieira (139):411-28

CARTOGRAFIA

atividades
Plano Geral (138):243

Carvalho, Alceu V. W de (139):339-56

— Cazal, Lindolfo (137):3-12

CENTRO DE INFORMAÇÃO E DISTRIBUIÇÃO EROS
criação (138): 244-5

COLONIZAÇÃO

área da Transamazônica
estudo (137):13-28

CONFERÊNCIA LATINO-AMERICANA DE INDUSTRIALIZAÇÃO
notícia (140):584-6

CONFERÊNCIA MUNDIAL DE POPULAÇÃO

presença do Brasil (137):57-8
reuniões preparatórias (137):57-8

Cova, Hervey Guimarães (138):131-78;
(140):507-28

CUSTO DE VIDA

índices
evolução (137):3-12

DEMOGRAFIA

estudos (137):13-56; (138):179-225;
(139):289-428; (140):529-82
ver também POPULAÇÃO

DEPARTAMENTO DE ESTATÍSTICA DO ESPÍRITO SANTO
reestruturação (138):246-7

DISTRITO FEDERAL

população
estimativa (140):575-82

ENCONTRO BRASILEIRO DE ESTUDOS POPULACIONAIS
notícia (139):429-41

ENSINO

estatísticas educacionais
alguns problemas (138):131-78
superior
ENCE
bacharelados (140):583-4

ESCOLA NACIONAL DE CIÊNCIAS ESTATÍSTICAS — ENCE
bacharelados (140):583-4

ESPÍRITO SANTO

Departamento de Estatística
reestruturação (138):246-7

ESTATÍSTICA

derivadas
Plano Geral (138):240-2
educacional
alguns problemas (138):131-78
órgãos de
publicações editadas (137):112-5;
(138):255-8; (139):477-9; (140):
644-8
primárias
Plano Geral (138):235-40
produção agropecuária
- legislação (137):125
- Reuniões Internacionais
- calendário (137):59-64
Sociedade Brasileira de Estatística
nova diretoria (139):443-4
teórica e aplicada
independência estatística (140):
507-28
vital
coleta (138):179-202

ESTUDO NACIONAL DA DESPESA FAMILIAR — ENDEF (139):445-53

FAMÍLIA

participação na força-de-trabalho
Brasil (139):371-96

Fialho, Eduardo Olímpio Mota (140):
575-82

FORÇA-DE-TRABALHO

membros da família
participação
Brasil (139):371-96

Frias, Luiz Armando de Medeiros
(137):35-56; (140):529-44

FUNDAÇÃO DE ECONOMIA E ESTATÍSTICA DO RIO GRANDE DO SUL
criação e finalidades (138):246

Fundação IBGE, Rio de Janeiro. Centro Brasileiro de Estudos Demográficos
(139):357-70

FUNDAÇÃO IBGE

Centro de Informação e Distribuição EROS
criação (138):244-5
Dia do Ibgeano
comemorações (138):227-34
Plano Geral de Informações Estatísticas e Geográficas (138):235-44

reformulação e modernização (137):
67-77

Geiger, Pedro Pinchas (139):411-28

GEOGRAFIA

estudos e pesquisas

Plano Geral (138):244

GRANDE RIO, região (139):289-338

GUANABARA

mortalidade

estudo (140):545-74

INDEPENDÊNCIA ESTATÍSTICA

estudo (140):507-28

INDICADORES DE SAÚDE

revisão analítica (139):339-56

ÍNDICE

custo de vida

evolução (137):3-12

INDÚSTRIA

desenvolvimento

Conferência Latinoamericana de
Industrialização

notícia (140):584-6

Irwin, Richard (138):179-202; (140):
575-82

Jordan Filho, Leon (139):339-56

LEGISLAÇÃO

Leis e Decretos (137):125

Leite, Valéria da Motta (138):203-26

Madeira, João Lyra (139):289-338

Magalhães, Augusta Baptista de, *trad.*
(137):77-81; (139):454-63; (140):
584-606

MATRIZES DE RELAÇÕES INTERSE- TORIAIS

de Bens e Serviços (140):483-506

MIGRAÇÃO

interna

e urbanização

Brasil (139):411-28

nordestino na Transamazônica

fixação (137):13-28

Monteiro, Mário F. G. (139):339-56;
(140):545-74

MORTALIDADE

distorções

estudo

Brasil (137):35-56

Guanabara (140):545-74

tábuas de

Grande Rio, região (139):289-338

Salvador, BA (139):289-338

São Paulo, SP (140):529-44

NECROLÓGIO

José Ayres de Souza Filho (137):123-
4

NORDESTINO

migração interna

fixação (137):13-28

NOTICIÁRIO (137):57-66; (138):227-
47; (139):429-44; (140):583-6

Oliveira, Luiz Antônio Pinto de (138):
179-202

ORÇAMENTO FAMILIAR

Estudo Nacional da Despesa Fami-
liar (139):445-53

PESQUISA

Superintendência de Pesquisa e De-
senvolvimento

função, estudos (139):463-7

Pimentel, Maria de Lourdes Sá Barre-
to (137):13-28

PLANO DE AÇÃO MUNDIAL

Objetivos (137):58

PLANO GERAL DE INFORMAÇÕES
ESTATÍSTICAS E GEOGRÁFICAS
(138):235-44

POPULAÇÃO

algumas considerações (137):77-81

Ano Mundial de População

notícia (139):441-2

Conferência Mundial de População

reuniões preparatórias (137):57

distribuição

Brasil (139):271-8

economicamente ativa

aspectos demográficos e econômi-
cos

Brasil (139):397-410

Encontro Brasileiro de Estudos Po-
pulacionais

notícia (139):429-41
estimativa
Distrito Federal (140):575-82
não residente
e tábuas de mortalidade
Brasil (137):35-56
Processo de Reprodução da População
Grupo de Trabalho
3ª Reunião (139):442-3
projeção por sexo e idade
Brasil (138):203-26; (139):357-70
urbana
distribuição
Brasil (139):271-8

PROCESSO DE REPRODUÇÃO DA POPULAÇÃO

Grupo de Trabalho
3ª Reunião (139):442-3

RECENSÕES

Análise Conjuntural (139):473
Anuário Estatístico da Aviação Civil (140):641-2
Anuário Estatístico do Café (137):110-1
Anuário Estatístico Embratur (137):112
Anuário Estatístico do Estado da Guanabara (137):112
Anuário Estatístico da Indústria Siderúrgica Brasileira — 1974 (138):253
Anuário Estatístico da RFFSA (140):642-3
Anuário Estatístico de Sergipe (140):643
Atas da Conferência Geral de Liege — 1973 (138):249-50
Boletim Estatístico do DNAEE (139):473-4
Boletim Estatístico da Guanabara (138):253-4
Boletim Estatístico do IBGE (140):643-4
Brasil Exportação — 1973 (139):470-1
Censo Agropecuário — Acre (140):638-9
Censo Agropecuário de Goiás (138):251
Censo Agropecuário — Guanabara (140):639-40
Censo Agropecuário — Rio Grande do Norte e Espírito Santo (139):476

Censo Industrial do Brasil — 1970 (140):638
Censo Industrial do Ceará (138):254
Censo Industrial do Estado de Alagoas (137):110
Censo Industrial do Estado do Amazonas (138):252
Censo Industrial do Estado de Santa Catarina (137):111
Censo Industrial — Guanabara (139):474
Censo Industrial — Maranhão, Rio Grande do Norte, Pernambuco, Bahia, Minas Gerais, Rio Grande do Sul e Mato Grosso (139):471-2
Censo Industrial — Rio de Janeiro (139):475-6
Censo Industrial — Rondônia-Roraima-Amapá, Acre, Pará, Piauí, Paraíba, Sergipe, Espírito Santo, Goiás e Distrito Federal (138):259
Censo Industrial — São Paulo (140):640
Censo Predial — Brasil (139):468-9
Censo Predial — Região Nordeste, Sudeste, Sul e Centro-Oeste (139):472
Censo Predial — Região Norte (138):250
Comércio Interestadual — Exportação por Vias Internas (140):640-1
Conheça o seu Município (138):259; (139) 475
Conjuntura Maranhense (137):116
A Conjuntura do Sal em 1973 (137):115
Estatística do Comércio Exterior (137):109
Indústrias de Transformação — pesquisa mensal (137):115-6
Inquérito Nacional de Preços — Gêneros Alimentícios (140):642
Pesquisa de Rendimentos — PNAD, 2 (139):469
Sinopse Estatística do Brasil (138):248-9

RECURSOS HIDRÁULICOS

Simpósio Internacional sobre as variabilidades nos Sistemas Hidrológicos e Recursos Hidráulicos (137):64-6

- REGISTRO CIVIL**
e estatísticas vitais
Brasil (138):179-202
- RELAÇÕES INTERSETORIAIS**
de Bens e Serviços (140):483-506
- RENDA**
bibliografia (137):83-108
- RESENHA** (137):67-82; (139):445-68;
(140):587-606
- REUNIÃO CEPAL/FAO**
alimentação
estudo apresentado
América Latina (140):587-606
- REUNIÕES INTERNACIONAIS DE ESTADÍSTICA**
calendário (137):59-64
- REVISTA BRASILEIRA DE ESTATÍSTICA**
índice (137):117-21
- RIO GRANDE DO SUL**
Fundação de Economia e Estatística do Rio Grande do Sul
criação e finalidades (138):246
- SALVADOR, BA** (139):289-338
- Sanson, João Rogério (139):397-410
- Sant'Ana, Maristela Afonso de Andrade
(140):483-506
- SÃO PAULO, SP**
tábuas de mortalidade
estudo (140):529-44
- SAÚDE**
indicadores
revisão analítica (139):339-56
- Sauvy, Alfred (137):77-81
- Silva, Diva Cristina da (139):411-28
- Simões, Celso Cardoso da Silva et alii
(139):371-96
- SIMPÓSIO INTERNACIONAL SOBRE AS VARIABILIDADES NOS SISTEMAS HIDROLÓGICOS E RECURSOS HIDRÁULICOS**, Tucson, Arizona, 11/14 dez. 1972
notícia (137):64-6
- SOCIEDADE BRASILEIRA DE ESTATÍSTICA**
nova diretoria (139):443-4
- Souza, Alberto Alexandre de (138):261-6
- SOUZA FILHO, José Ayres de**
necrológico (137):123-4
- Spielman, Evelyn (138):203-26
- SUPERINTENDÊNCIA DE PESQUISA E DESENVOLVIMENTO**
função, estudos (139):463-7
- TÁBUAS DE MORTALIDADE**
Grande Rio, região (139):289-338
Salvador, BA (139):289-338
- TRANSAMAZÔNICA**
área de colonização
migrante nordestino
fixação (137):13-28
- URBANIZAÇÃO**
e migração interna
Brasil (139):411-28
- VIDA MÉDIA**
estimativa
uso das tábuas de vida modelo
(137):29-34
- VULTOS DA ESTATÍSTICA BRASILEIRA**
Paulo de Assis Ribeiro, prof. (138):261-6

Composto e impresso nas
oficinas do Serviço Gráfico do
IBGE - Avenida Brasil, 15 671
- Lucas - Rio de Janeiro - RJ

IBGE

Presidente: ISAAC KERSTENETZKY

Diretor-Geral: EURICO DE ANDRADE NEVES BORBA

Diretor-Técnico: AMARO DA COSTA MONTEIRO

DEPARTAMENTO DE DIVULGAÇÃO ESTATÍSTICA

Chefe-Substituto: Mário Fernandes Paulo

SECRETARIA DA REVISTA BRASILEIRA DE ESTATÍSTICA

Chefe: Fernando Pereira Cardim