Presidente da República
José Sarney

Ministro-Chefe da Secretaria de Planejamento e Coordenação
João Batista de Abreu

Secretário-Geral
Ricardo Luís Santiago

FUNDAÇÃO INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA — IBGE

Presidente
Charles Curt Mueller

Diretor-Geral
David Wu Tai

Diretor de Pesquisas
Lenildo Fernandes Silva

Diretor de Geociências
Mauro Pereira de Mello

Diretor de Informática
José Sant' Anna Bevilaqua

REVISTA BRASILEIRA DE ESTATÍSTICA

Editor Responsável
Djalma Galvão Carneiro Pessoa

Conselho Editorial

Cláudio Considera
Diretoria de Pesquisas

Elisa Caillaux
Diretoria de Pesquisas

Helio Mignon
Escola Nacional de Ciências Estatísticas

Kaizô Beltrão
Escola Nacional de Ciências Estatísticas

Marilourdes Lopes Ferreira
Diretoria de Geociências

Pedro Luis N. Silva
Diretoria de Pesquisas

Victor Hugo Gouvêa
Diretoria de Pesquisas

Valéria da Motta Leite
Diretoria de Pesquisas

SECRETARIA DE PLANEJAMENTO E COORDENAÇÃO DA PRESIDÊNCIA DA REPÚBLICA
FUNDAÇÃO INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA — IBGE

REVISTA BRASILEIRA DE ESTATÍSTICA

ISSN 0034-7175723X

R. bras. Estat. Rio de Janeiro, ano 48, nºs 189/190, p. 1 — 104, jan./dez. 1987

REVISTA BRASILEIRA DE ESTATÍSTICA

Órgão oficial do IBGE e da Sociedade Brasileira de Estatística

Publicação semestral, editada pelo IBGE, que se destina a promover e ampliar o uso de métodos estatísticos (quantitativos) na área das ciências econômicas e sociais, através de divulgação de artigos inéditos.
Temas, abordando aspectos do desenvolvimento metodológico, serão aceitos desde que relevantes para os órgãos produtores de estatísticas.
Os originais para publicação devem ser endereçados para:
Revista Brasileira de Estatística/Escola Nacional de Ciências Estatísticas
Rua André Cavalcanti, 106 — Bairro de Fátima
Rio de Janeiro — RJ — CEP: 20.231
Tel.: (021) 242-0137

Os pedidos de assinatura e número avulso ou atrasado devem ser endereçados para:
Centro de Documentação e Disseminação de Informações
Av. Beira Mar, 436 — 8.º andar — Centro
Rio de Janeiro — RJ — CEP: 20.021
Tel.:(021) 533-3094

A Revista não se responsabiliza pelos conceitos emitidos em matéria assinada.

Criação: Programação Visual e Capa
Pedro Paulo Machado

© IBGE — Copyright

Revista brasileira de estatística / Fundação Instituto Brasileiro de Geografia e Estatística — ano 1, n. 1 (jan./mar., 1940)-
Rio de Janeiro : IBGE, 1940-
Trimestral (1940-1986), semestral (1987-)
Órgão oficial do IBGE e da Sociedade Brasileira de Estatística.
Continuação de : Revista de economia e estatística
Sumários e índices acumulados de autor — assunto: v.1/40 (1940) no v. 43, n. 169 (jan./mar. 1982)
ISSN 0034-7175 = Revista brasileira de estatística
1. Estatística — Periódicos. I. IBGE

IBGE. Gerência de Documentação e Biblioteca
RJ-IBGE/88-05

CDU 31(05)

Impresso no Brasil/Printed in Brazil

SUMÁRIO

EDITORIAL 5

ARTIGOS

PREVISÕES DO IPCA 7
Pedro L. Valls Pereira
Sérgio Portugal

PESQUISA ELEITORAL — UMA ANÁLISE BAYESIANA 25
Isabel G.S. Furtado de Mendonça
Helio S. Mignon

APLICAÇÃO DA AMOSTRAGEM NA COLETA DOS CENSOS DEMOGRÁFICOS NO BRASIL 35
Luiz Nery da Costa

USO DA TÁBUA DE MORTALIDADE EM ESTUDOS DE SEGUIMENTO 65
David Dorigo

MULHER BRASILEIRA: MUITO PRAZER 71
André Cezar Medici

OBSERVAÇÕES SOBRE O CÁLCULO DO ÍNDICE DE GINI (E SUA COMPOSIÇÃO) COM DADOS INDIVIDUAIS DA RENDA 99
José W. Rossi

INSTRUÇÕES BÁSICAS PARA PREPARO DOS ORIGINAIS 103

ISSN 0034-7175

R. bras. Estat. Rio de Janeiro, v.48, nº 189/190, p.1 – 104, jan./dez. 1987

EDITORIAL

Em seus quarenta e oito anos de existência, a REVISTA BRASILEIRA DE ESTATÍSTICA tem sido depositária da história da evolução da ciência estatística no País e um dos principais veículos de divulgação do progresso das técnicas estatísticas.

Sob este último aspecto, a fase inicial da REVISTA foi particularmente fecunda e, num certo sentido, a política editorial que está sendo agora implantada pretende fazer revivê-la, na medida em que buscará recolocar a REVISTA BRASILEIRA DE ESTATÍSTICA na condição de uma importante referência dos temas metodológicos mais valiosos para os órgãos produtores de estatísticas no Brasil. Além dessas matérias contemplando aspectos teóricos da estatística, a REVISTA privilegiará os trabalhos que versem aplicações da análise estatística, notadamente aos fenômenos econômicos e sociais.

Como quer que seja, nosso compromisso maior, na fase que aqui se inaugura, é o de que, pela competência de seus colaboradores e pela atualidade da temática que aborde, a REVISTA BRASILEIRA DE ESTATÍSTICA — agora semestral e com nova feição gráfica — reassuma a posição de um periódico técnico de excelência na estatística e em suas extensas áreas conexas.

Esclarecemos que circunstâncias diversas nos obrigaram a reunir, num só volume, os números 189 e 190 correspondentes a 1987.

Superada esta fase, a nossa REVISTA retomará sua periodicidade semestral.

PREVISÕES DO IPCA*

Pedro L. Valls Pereira**
Sergio Portugal***

1 — INTRODUÇÃO

Alguns exercícios sobre previsões da inflação são pertinentes no atual momento de incerteza em que se encontra a economia brasileira. Neste sentido, este trabalho procura gerar alguns números preliminares para obter intervalos aceitáveis para previsões da inflação.

Embora o indexador da economia seja atualmente o IPC, utilizou-se a série do IPCA. Isto foi feito porque o objetivo é comparar os resultados aqui obtidos com os de outros estudos já feitos no passado. A série do IPCA utilizada não incorpora as correções metodológicas previstas após o Plano Cruzado, que compatibilizam as séries, pré e pós-cruzado, quanto ao período de coleta. Num trabalho posterior pretende-se utilizar a nova série, mas para isto será preciso um estudo mais detalhado, já que se faz necessário identificar um novo modelo para os dados.

Os diversos métodos de previsão da inflação podem ser divididos em dois grupos:

1) "Planilha de Cálculo": são aqueles que utilizam o sistema de ponderação do índice de preços e algumas suposições, muitas vezes *ad hoc,* sobre o comportamento futuro de cada um desses itens e do sistema de ponderação; e

2) *Model-Based:* são aqueles que utilizam algum modelo para o processo de geração dos dados (por exemplo, ajustam a série de preços por algum modelo de série temporal da classe Box e Jenkins).

O primeiro método tem a desvantagem de não levar em consideração a estória existente na série temporal. Porém, sua principal vantagem é a possibilidade de interferência no processo de previsão, dado que o modelador pode incorporar informações relevantes às suas previsões. Por esta característica — incorporação de subjetivida-

* Este trabalho foi escrito utilizando-se as informações disponíveis até 16/02/87. Os autores desejam agradecer os comentários e sugestões de Ricardo Paes de Barros e Hélio Migon, de dois pareceristas desta Revista, a revisão de Nilson Souto Maior e a Marcelo Landau pela ajuda computacional

** Do Instituto de Pesquisas do IPEA e do Departamento de Economia da UFF.
*** Do Instituto de Pesquisas do IPEA.

de —, este método pode tornar-se um pouco mais científico ao se introduzir algum grau de incerteza, tanto na intervenção quanto na utilização das informações. Se isto fosse possível, a tendência seria chamar este método de *Ad Hoc Bayesiano*.

O segundo método possui características diametralmente opostas em relação ao primeiro.

Este estudo apresenta uma abordagem *Model-Based* para o problema de como prever a taxa de inflação quando esta é medida pelo IPCA. Na Seção 2, a seguir, procede-se a uma descrição do método de "Planilha de Cálculo", na Seção 3, no título Previsões do IPCA via Modelos ARIMA, faz-se uma introdução aos modelos de previsão do tipo ARIMA, assim como descrevem-se as previsões obtidas com estes modelos e atualizam-se os modelos utilizados em Brasil et alii (1986); e, finalmente, na Seção 4, apresenta-se uma introdução aos Modelos Estruturais de Séries Temporais, assim como os resultados obtidos com a sua utilização, e faz-se uma comparação das duas abordagens.

2 — PLANILHA DE CÁLCULO

Os objetivos dos métodos *Model-Based* e "Planilha de Cálculo" para calcular a inflação são distintos, pois no primeiro o que se quer estimar são as taxas mensais de inflação, enquanto que no segundo o interesse primordial é mensurar impactos de variações de preços de um conjunto de bens ou serviços na taxa de inflação. Esse interesse, em geral, resulta dos órgãos responsáveis pela política de controle de preços.

Sob o ponto de vista macroeconômico, é útil, muitas vezes, simular o impacto inflacionário de vetores alternativos de variações de preços. Numa economia hipotética, onde o governo controlasse todos os preços, seria possível, com a "Planilha de Cálculo", antecipar perfeitamente a taxa mensal da inflação. No Brasil, em ocasiões de estrito controle de preços, o governo chegou a controlar, através de uma cesta de bens e serviços, 60% da taxa mensal de inflação.

A "Planilha de Cálculo" é um instrumento que visa a mensurar o impacto da variação de preços de um conjunto de bens na taxa mensal de inflação, podendo, a partir de hipóteses para os demais bens, antecipar por um ou no máximo dois períodos a taxa mensal de inflação. Os impactos podem ser separados em:

1) diretos ou de primeira ordem, que resultam diretamente da variação de preços de um determinado bem sobre a taxa mensal de inflação; e

2) indiretos ou de ordens superiores à primeira, que resultam de efeitos secundários na variação de preços de um dado produto sobre os demais e destes na taxa de inflação.

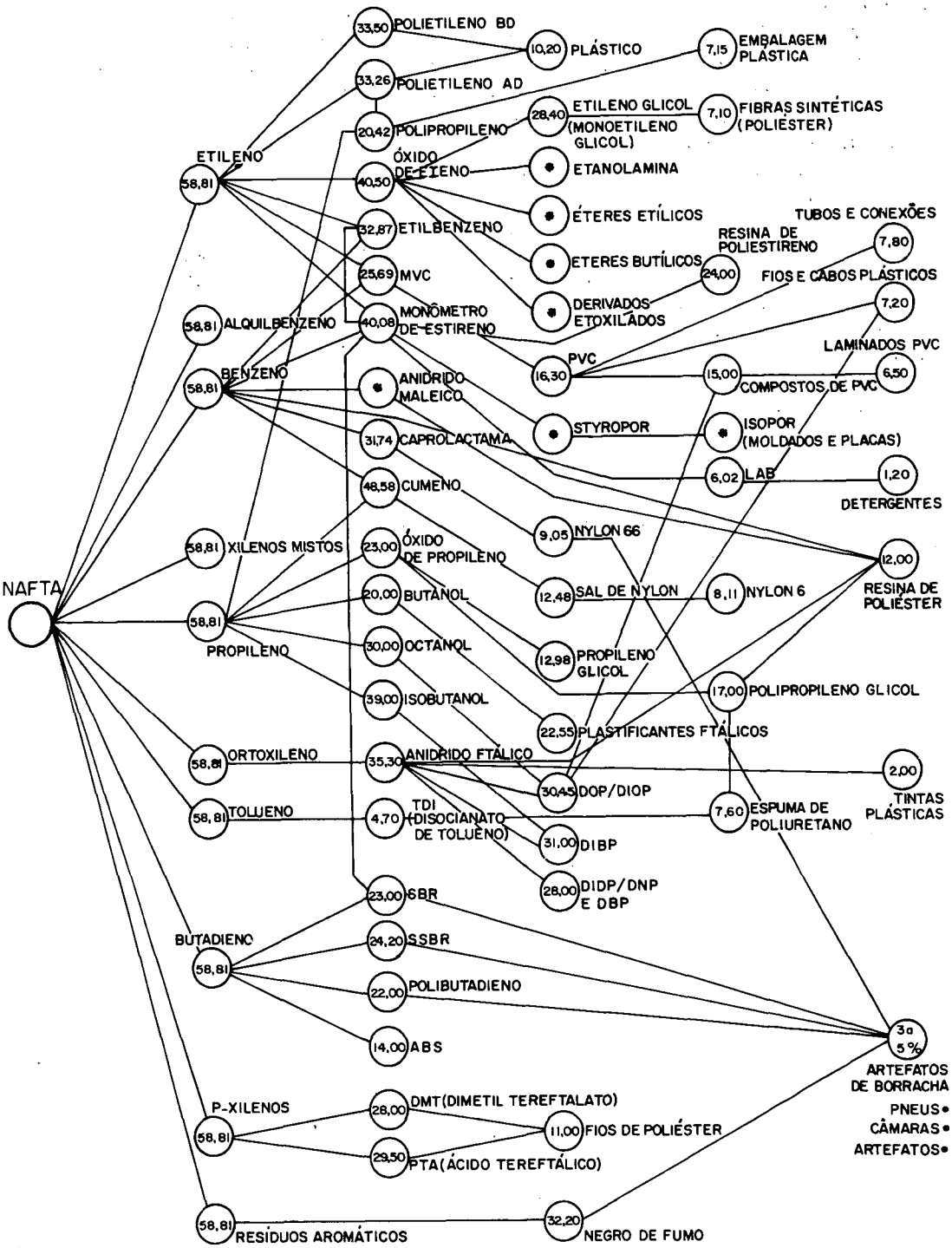
Um exemplo pode esclarecer a distinção: quando o governo autoriza o aumento do preço da gasolina, o efeito direto é medido pela variação de preço vezes o peso relativo da gasolina no índice em questão: as elevações de preços que derivam do aumento da gasolina, tais como tarifas de táxi, transporte escolar, etc., são os efeitos indiretos da elevação do preço da gasolina.

Para calcular o impacto indireto de um determinado produto, é necessário conhecer a estrutura de custos de cada serviço ou produto específico. Um exemplo para os produtos petroquímicos derivados da nafta pode ser observado no Quadro 1: uma elevação de 100% no preço da nafta eleva em 58,81% o preço de etileno, que por sua vez eleva em 40,50% o do óxido de eteno, em 28,40% o do etileno glicol e, finalmente, em 7,10% o das fibras sintéticas de poliéster. Nesse caso, o efeito de primeira ordem será determinado pelo aumento de 100% da nafta no índice de preço e os de ordem superior serão dados pela variação de preços dos demais produtos.

A "Planilha de Cálculo" é construída a partir de um vetor de aumento de preços autorizado pelo governo para o período $(t+1)$, cujos produtos compõem o índice de preço relevante para medir a inflação. Na prática, como se verá, é usual trabalhar com um conjunto restrito e significativo de produtos controlados pelo governo, para os quais se conhece antecipadamente as variações autorizadas, e supor que os demais produtos têm variações próximas da média da taxa de inflação do período anterior.

QUADRO 1

PRODUTOS PETROQUÍMICOS DERIVADOS DA NAFTA E ALGUMAS APLICAÇÕES



•INEXPRESSIVOS NO SETOR

FONTE: SEPLAN/CIP (o quadro foi montado a partir da estrutura de custo dos produtos considerados.

A Tabela 1 apresenta a saída de um programa de "Planilha de Cálculo" para três índices de preços calculados pela FGV—IPA, ICV/RJ e INCC.

Os produtos são divididos em controlados e não-controlados pelo governo. Dos controlados, a título de exemplo, são explicitados alguns na primeira coluna. As colunas subseqüentes apresentam:

1) as ponderações dos produtos nos índices indicados;

2) o vetor de variação de preços autorizados pelo governo;

3) o impacto nos respectivos índices, dado o vetor de preços autorizados;

4) o "fator" representando o relativo do número de dias do mês em que cada aumento vigorou; e

5) o impacto no IGP/DI[1].

As colunas em grifo representam os vetores de entrada de dados da "Planilha de Cálculo".

No exemplo, a taxa de inflação prevista para o mês, medida pelo IGP/DI, é de 15,944%, sendo que 5,1210% representa o impacto direto dos produtos controlados indicados na primeira coluna e 3,9984% de outros produtos controlados pelo governo (cujo aumento médio considerado foi de 15%). No exercício hipotético, 60% (9,1185%) da taxa prevista de inflação resultariam da política de preços praticada pelo governo.

3 — PREVISÕES DO IPCA VIA MODELOS ARIMA

3.1 — Introdução[2]

Antes de considerar-se como uma série temporal pode ser prevista. É necessário introduzir alguns modelos, isto é, métodos que determinam o processo gerador dos dados. O mais simples que se pode conceber é o de uma série puramente aleatória, usualmente chamada de "ruído branco"[3]. Uma

TABELA 1

MODELO DE DECOMPOSIÇÃO DA TAXA DE INFLAÇÃO

	PONDERAÇÕES			VARIAÇÃO	IMPACTOS			IMPACTO NO IGPDI		
	IPA	ICV	INCC	DE PREÇO	IPA	ICV	INCC	TOTAL	FATOR	NO MÊS
MARGARINA	0.1864	0.4856	0.0000	11.78	0.0220	0.0572	0.0000	0.0303	1.00	0.0303
SABÃO	0.2847	0.7396	0.0000	22.66	0.0645	0.1676	0.0000	0.0890	1.00	0.0890
DETERGENTE	0.1034	0.0884	0.0000	21.24	0.0220	0.0188	0.0000	0.0188	1.00	0.0188
SABONETE	0.0923	0.5249	0.0000	12.43	0.0115	0.0652	0.0000	0.0265	1.00	0.0265
BEBIDAS E REFRIG.	1.2759	0.6047	0.0000	20.00	0.2552	0.1209	0.0000	0.1894	1.00	0.1894
LEITE	0.0000	5.1672	0.0000	47.00	0.0000	2.4286	0.0000	0.7286	1.00	0.7286
PÃO FRANCÊS	0.0000	3.3281	0.0000	14.30	0.0000	0.4759	0.0000	0.1428	1.00	0.1428
CARNE	3.4285	8.9651	0.0000	20.00	0.6857	1.7930	0.0000	0.9493	1.00	0.9493
COMBUSTÍVEIS E DER.	6.0225	2.8221	0.0000	20.00	1.2045	0.5644	0.0000	0.8920	0.10	0.0892
AÇO	3.9779	0.0000	0.0000	15.00	0.5967	0.0000	0.0000	0.3580	1.00	0.3580
ENERGIA	0.0000	3.0741	0.0000	15.00	0.0000	0.4611	0.0000	0.1383	0.33	0.0457
TRANSP. URB.	0.0000	6.5834	0.0000	15.00	0.0000	0.9875	0.0000	0.2963	0.33	0.0978
CORRESPONDÊNCIAS	0.0000	0.0017	0.0000	20.00	0.0000	0.0003	0.0000	0.0001	1.00	0.0001
TELEFONE	0.0000	0.0632	0.0000	39.00	0.0000	0.0246	0.0000	0.0074	1.00	0.0074
IMPOSTOS	0.0000	0.6711	0.0000	34.30	0.0000	0.2302	0.0000	0.0691	1.00	0.0691
ALUGUEL	0.0000	5.3876	0.0000	68.80	0.0000	3.7067	0.0000	1.1120	1.00	1.1120
CASA PRÓPRIA	0.0000	0.2179	0.0000	112.00	0.0000	0.2440	0.0000	0.0732	1.00	0.0732
TOTAL	15.3716	38.7247	0.0000	24.57	2.8620	11.3462	0.0000	5.1210	0.79	4.0271
OUTROS CONTROLADOS	44.2284	0.3753	0.0000	15.00	6.6343	0.0563	0.0000	3.9974	1.00	3.9974
TOTAL CONTROLADOS	59.6000	39.1000	0.0000	19.20	9.4962	11.4025	0.0000	9.1185	0.88	8.0245
NÃO-CONTROLADOS	40.4000	60.9000	100	13.00	5.2520	7.9170	13.0000	6.8263	1.00	6.8263
TOTAL	100	100	100	16.10	14.7482	19.3195	13.0000	15.9448	0.93	14.8508

Vetor entrada =

[1] Considera-se o IGP = 0,6 IPA + 0,3 ICV + 0,1 INCC.

[2] Muitas das idéias contidas nesta seção são devidas a Granger (1980).

[3] Este nome vem da literatura de engenharia. Difícil tentar explicá-lo sem algumas técnicas de análise espectral.

série é "ruído branco" se a média e a variância da série são independentes do tempo e não existe correlação temporal.

Suponha-se que o mercado de fumo de cachimbo no Brasil seja pequeno, que este produto seja negociado numa bolsa de mercadorias na Bahia, que não tenha atraído a atenção de especuladores e que este mercado esteja em equilíbrio, isto é, todos os que nele estejam envolvidos concordem sobre o preço correto a ser praticado num instante do tempo. Suponha-se, ainda, que em algum instante do tempo uma informação inesperada chegue ao mercado, como por exemplo, a respeito de uma tempestade tropical que tenha destruído a plantação de fumo em Cuba. Esta nova informação provavelmente acarretará um aumento do preço do tabaco na Bahia. Suponha-se que o preço seja denotado por P_t e que o novo preço seja determinado por:

$$P_{t+1} = P_t + \epsilon_{t+1} \qquad (3.1.1)$$

onde ϵ_{t+1} é o efeito desta informação inesperada nos preços. Denotando-se a variação dos preços por Y_t, pode-se escrever (3.1.1) da seguinte forma:

$$Y_{t+1} = \epsilon_{t+1} \qquad (3.1.2)$$

Neste caso, o modelo que representa a variação do preço do tabaco é "ruído branco".

Como estamos interessados em projeções, pode-se mostrar que o melhor[4] previsor dos valores futuros de uma série do tipo "ruído branco" é a média da série. Neste caso, a função de previsão deste modelo é a média da série.

Suponha-se que as informações inesperadas continuem chegando à bolsa de mercadorias. Por exemplo, no instante $t = 2$ chega a informação de que a chuva em Cuba continua, mas que a do dia anterior não foi tão devastadora quanto levava a crer a primeira informação. Neste caso, a variação dos preços no instante $t + 2$ é dada por:

$$Y_{t+2} = \epsilon_{t+2} + \theta\,\epsilon_{t+1} \qquad (3.1.3)$$

onde ϵ_{t+2} é o efeito da nova informação e $\theta\,\epsilon_{t+1}$ reflete a perda de importância da informação anterior. Por exemplo, se $\theta = 0{,}5$, somente 50% da informação inesperada no instante $t + 1$ é repassada para a mudança dos preços no instante $t + 2$.

O modelo (3.1.3), para fins da previsão, é mais informativo do que um "ruído branco"[5]. Como já foi visto antes, a função de previsão para um "ruído branco" é a média da série. Quando esta for zero, tem-se que a função de previsão seja sempre zero, qualquer que seja o horizonte de previsão. No caso do modelo (3.1.3), pode-se mostrar que a função de previsão um passo à frente é dada por:

$$f_{n,1} = v(y_{n} - f_{n-1,1}) \qquad (3.1.4)$$

onde $f_{n,h}$ representa a previsão, $\underline{h}$ passos à frente, feita no instante $\underline{n}$. Pode-se mostrar que:

$$f_{n,h} = 0 \qquad h > 1 \qquad (3.1.5)$$

A propriedade (3.1.5) implica que a função de previsão "perde memória" para um horizonte acima da ordem do modelo. No exemplo acima, como a ordem do modelo é 1, se houver interesse em obter previsões para a variação dos preços mais de um período à frente, chega-se à conclusão de que esta variação é nula.

Os modelos do tipo média-móvel satisfazem a propriedade (3.1.5) e, em conseqüência, são de uso limitado em termos de previsão. Deseja-se obter modelos cuja estória contida na série seja informativa.

Considere-se o seguinte modelo: o número de pessoas desempregadas num mês[6] é composto de uma proporção fixa ϕ dos de-

[4] O critério de otimalidade usado é o de mínimos quadrados.

[5] Em geral, se a variabilidade do modelo (3.1.2) fosse a mesma de (3.1.3), seria escolhido pelo princípio da parcimônia, o modelo mais simples, isto é, (3.1.2). Em geral, modelos com mais parâmetros podem ter uma variabilidade maior se a variância dos erros for a mesma nos modelos. Mas se há uma redução na variância dos erros, quando se utiliza um modelo com mais parâmetros — como, por exemplo, se a variância dos erros de (3.1.3) é menor do que a de (3.1.2) —, as previsões com o modelo mais geral serão melhores.

[6] Este modelo não é necessariamente bom ou realista para tentar explicar o número de desempregados.

sempregados no mês anterior (as outras pessoas obtiveram emprego naquele mês) e um novo grupo de trabalhadores que está procurando emprego. Supondo-se que as novas entradas na força de trabalho, que ainda não conseguiram emprego, podem ser representadas por um "ruído branco" com média $\mu(1-\phi)$, tem-se que o número de desempregados, denotados por x_t, pode ser modelado por:

$$x_t = \phi x_{t-1} + \mu(1-\phi) + \epsilon_t \qquad (3.1.6)$$

onde ϵ_t é um "ruído branco" com média zero. Pode-se escrever (3.1.6) da seguinte forma:

$$x_t - \mu = \phi(x_{t-1} - \mu) + \epsilon_t \qquad (3.1.7)$$

que é um modelo chamado de auto-regressivo de primeira ordem. Pode-se mostrar que a função de previsão para (3.1.7) é dada por:

$$f_{n,1} = \phi x_n \qquad (3.1.8)$$

$$f_{n,h} = \phi^h x_n \qquad h > 1 \qquad (3.1.9)$$

Suponha-se agora que o objetivo seja modelar o número total de desempregados de duas cidades e que o modelo gerador seja o mesmo para cada uma delas, isto é, se y_t representa o número de desempregados da cidade do Rio de Janeiro e x_t o da cidade de Manaus e ambas as séries, x_t e y_t, podem ser modeladas por um modelo do tipo (3.1.7), deseja-se modelar:

$$z_t = x_t + y_t \qquad (3.1.9)$$

Por hipóteses admita-se que x_t seja independente de y_t[7]. Neste caso, o modelo gerador de z_t não é mais auto-regressivo de primeira ordem, que é o que gera cada um dos componentes de (3.1.9). O modelo gerador da série z_t é chamado de auto-regressivo-média-móvel[8], que, neste caso, tem ordem 2 para a parte auto-regressiva e ordem 1 para a parte média-móvel, sendo denotado por ARMA (2,1). Sua função de previsão é uma composição das funções de previsão para as partes auto-regressiva e média-móvel.

Os modelos considerados até aqui são chamados de estacionários e têm uma característica em comum: as médias, variâncias e correlações temporais são independentes do tempo.

Em algumas séries de índices de preço ocorre uma tendência crescente ao longo do tempo e, neste caso, sua média não é independente do tempo. Em outras a dispersão da série depende do nível da mesma, isto é, a variância não é independente do tempo, já que está relacionada ao nível da série, que supostamente não é constante.

Estes exemplos mostram que séries econômicas, em geral, são não-estacionárias. Em alguns casos, uma simples transformação da série original induz estacionaridade. No exemplo do preço do tabaco na bolsa de mercadorias da Bahia, pode-se mostrar que a série p_t – do nível dos preços – tem variância $t\sigma_\epsilon^2$, onde σ_ϵ^2, é a variância de ϵ_t. Neste caso, a série P_t é não-estacionária, e este modelo é chamado de passeio aleatório *(random walk)* [9]. Mas a série y_t – variação dos preços – era um "ruído branco" (logo, estacionária). Observe que a relação entre as duas séries nos dá a transformação necessária para induzir estacionaridade. Neste caso, esta transformação é a primeira diferença da série. Formalmente tem-se que:

$$y_t = P_t - P_{t-1} = (1 - L)P_t \qquad (3.1.10)$$

onde L e o operador defasagem, definido por:

$$L^j P_t = P_{t-j} \qquad (3.1.11)$$

Considera-se agora o exemplo do preço mensal do trigo na Argentina, que provavelmente será menor logo após a colheita, da-

[7] Como os custos de transporte de uma cidade para outra são extremamente elevados, fazendo com que a migração das pessoas ocupadas seja insignificante, a hipótese de independência das duas séries é razoável.

[8] Existe um teorema que diz se $x_t \sim$ ARMA (p_1,q_1) e $y_t \sim$ ARMA (p_2,q_2) com x_t e y_t independentes, então $z_t = x_t + y_t \sim$ ARMA (P,Q), onde $P = p_1 + p_2$ e Q = máximo $[p_1 + q_2, p_2 + q_1]$.

[9] Este modelo é denotado por ARIMA (p,d,q), onde ARIMA significa Auto-regressivo-integrado-média-móvel e, neste caso, $p = q = 0$ e $d = 1$.

do que existe uma grande quantidade ofertada deste produto, mas tenderá a subir à medida que a oferta disponível vá diminuindo até uma nova colheita. Esta série de preço pode ser descrita por uma tendência linear determinística, que representa os aumentos após a colheita mais o preço do trigo no mês da colheita anterior, isto é, se a série mensal do preço do trigo é denotada por PT_t, tem-se o seguinte modelo:

$$PT_t = PT_{t-12} + \alpha + \beta t \qquad (3.1.12)$$

que pode ser reescrito, usando-se (3.1.11), da seguinte forma:

$$PT_t^* = (1 - L^{12})\, PT_t = \alpha + \beta t \qquad (3.1.13)$$

o que implica que a variação dos preços em relação ao mesmo mês do ano anterior é dada por uma tendência determinística. Observa-se que a variação da série PT_t^* em relação ao mês anterior, é constante e igual a β. Como é provável que esta variação esteja sujeita a choques aleatórios, tem-se que o modelo para o preço do trigo é dado por:

$$PT_t^* = PT_{t-1}^* + \beta + \epsilon_t \qquad (3.1.14)$$

onde ϵ_t é um "ruído branco" que reflete os choques aleatórios nos preços. Observa-se que (3.1.14) equivale a:

$$PT_t^+ = (1 - L)(1 - L^{12})\, PT_t = \beta + \epsilon_t t \qquad (3.1.15)$$

e que a série PT_t^+ é estacionária e apresenta uma variação sazonal regular, dada pelo componente $(1 - L^{12})$, e uma variação não-sazonal (1-L), também regular. Este modelo é denotado por ARIMA $(0,1,0) \times (0,1,0)_{12}$.

Embora a série PT_t seja não-estacionária, utilizando-se duas transformações — uma para retirar a variação sazonal regular e a outra para retirar a variação não-sazonal — foi possível obter uma série estacionária, PT_t^+.

A função de previsão para o modelo (3.1.5) é dada por:

$$f_{n,h} = f_{n,h-1} + f_{n,h-2} - f_{n,h-13} + \beta \qquad h > 3 \qquad (3.1.16)$$

com as seguintes condições iniciais:

$$f_{n,1} = PT_n + PT_{n-11} - PT_{n-12} + \beta \qquad (3.1.17\text{-a})$$

$$f_{n,j} = f_{n,j-1} + PT_{n-j+12} - PT_{n-j+13} + \beta \qquad j = 2,\ldots,12 \qquad (3.1.17\text{-b})$$

$$f_{n,13} = f_{n,12} + f_{n,1} - PT_n + \beta \qquad (3.1.17\text{-c})$$

Para o caso geral, os modelos que melhor aproximam dados mensais, segundo Box e Jenkins (1976), são do tipo:

$$\phi(L)\Phi(L^s)\, \nabla^d\, \nabla_s^D y_t = \Theta(L)\Theta(L^s)\epsilon_t \qquad (3.1.18)$$

onde $\Phi(L)$ é o polinômio em L, o operador defasagem, de ordem p, que representa a parte auto-regressiva não-sazonal, $\Phi(L^s)$ é o polinômio em L^s de ordem P, que representa a parte auto-regressiva sazonal, $\Theta(L)$ é o polinômio em L de ordem q, que representa a parte média-móvel não-sazonal, $\Theta(L^s)$ é o polinômio em L^s de ordem Q, que representa a parte média-móvel, s é o período de sazonalidade, ∇^d é o operador diferença, definido por $\nabla^d = (1 - L)^d$, que permite tornar a parte não-sazonal estacionária, e ∇_s^D é o operador diferença sazonal, definido por $\nabla_s^D = (1 - L^s)^D$, que permite tornar a parte sazonal estacionária. Supõe-se que $\epsilon_t \sim NID(0, \sigma^2)$, que pode ou não ser adquado. O modelo (3.1.19) é usualmente chamado de ARIMA $(p,d,q) \times (P,Q,D)_s$.

3.2 — Modelos ARIMA para o IPCA

Num estudo anterior [Brasil et alii (1986)], o modelo identificado para o IPCA[10] foi $(0,1,1) \times (0,1,1)_{12}$[11], onde o período amostral era de 1980.I e 1986.I[12]. Estes resultados são apre-

[10] Utiliza-se a taxa de variação desta variável para se estabilizar a variância desta série.

[11] Este modelo é conhecido, na literatura de séries temporais, por *airline passenger* [ver Box e Jenkins (1976, pp. 305-6)].

[12] Utiliza-se a notação M.N., onde M é o ano e N = I,..., XII são os meses do ano M.

sentados abaixo para uma comparação com aqueles que figuram a seguir:

$$\nabla \nabla_{12} y_t = (1 - 0{,}82L)(1 - 0{,}77L^{12})\epsilon_t \qquad (3.2.1)$$

$$\sigma_w^2 = 7{,}44;\ \sigma_\epsilon^2 = 3{,}09;\ Q(23) = 26{,}3$$

onde $w = \nabla \nabla_{12} y$ e Q é a estatística de Box e Pierce para se testar a independência dos resíduos. Reestimando este modelo para o período 1980. I a 1986. XII, obteve-se:

$$\nabla \nabla_{12} y_t = (1 - \underset{(3,50)}{0{,}42}L)(1 - \underset{(4,73)}{0{,}74}L^{12})\epsilon_t \qquad (3.2.2)$$

$$\sigma_w^2 = 8{,}18;\ Q(22) = 18{,}85;\ AIC = 352{,}69;\ SBC = 357{,}22$$

onde AIC e SBC são dois critérios de seleção de modelos e os números entre parênteses são as estatísticas t de Student para os parâmetros.

Comparando-se os dois modelos, observa-se que os parâmetros sazonais da parte MA são bastante estáveis, enquanto que os da parte MA não-sazonal são estatisticamente diferentes nos dois modelos. Como o SAS, pacote utilizado para estimar o modelo (3.2.2), não apresenta σ_ϵ^2, não é possível comparar a aderência do modelo dentro da amostra. Uma vez que o interesse é fazer previsões, comparam-se as previsões dos dois modelos um passo à frente.

No caso do modelo (3.2.1), utilizando-se Brasil et alii (1986), a previsão para 1986. I foi de 14,29[13], enquanto que o observado foi de 16,20. Tem-se, portanto, que o erro de previsão foi de 1,91 e, em termos percentuais, de 11,79%.

No caso do modelo (3.2.2), a Tabela 2 abaixo, apresenta as previsões para 1987. I a 1987. XII.

A previsão para 1987. I foi de 10,8973 com desvio padrão de 2,8030, enquanto que o valor observado foi de 13,21%. Tem-se, portanto, que o erro de previsão é de 2,31% e o erro percentual de 21,22%. Adotando-se este número como "observado" para 1987. I, tem-se:

1) o intervalo de previsão para 1987.I[14] (5,2912, 16,5034), que, portanto, recobre o valor "observado", indicando uma melhoria em termos de poder preditivo; e

TABELA 2

PREVISÃO PARA O ANO DE 1987
Usando o Modelo (3.2.2)
Método Box e Jenkins

ANO	PREVISÃO	DESVIO PADRÃO	INTERVALO INFERIOR	INTERVALO SUPERIOR
87.I	10,8973	2,8030	5,2912	16,5034
87.II	8,0452	3,2386	1,5681	14,5224
87.III	5,5107	3,6221	−1,7336	12,7549
87.IV	4,6095	3,9688	−3,3280	12,5470
87.V	4,6897	4,2875	−3,8852	13,2647
87.VI	4,5366	4,5841	−4,6315	13,7048
87.VII	6,1580	4,8626	−3,5673	15,8832
87.VIII	5,3638	5,1260	−4,8883	15,6159
87.IX	6,1992	5,3766	−4,5540	16,9524
87.X	5,9780	5,6160	−5,2540	17,2099
87.XI	6,1112	5,8456	−5,5798	17,8023
87.XII	8,7639	6,0664	−3,3690	20,8967
ACUMULADA PARA 1987	110,28		−29,53	470,36

[13]Não é apresentado o erro-padrão desta previsão, prejudicando uma comparação que tenha um teor estatístico mais rigoroso.

[14]Definido por $\hat{y}$ + 2E.P.P, onde $\hat{y}$ é a projeção da taxa do IPCA para 1987.I e E.P.P é o erro-padrão da projeção.

2) por outro lado, em termos do erro de previsão, há uma piora para o modelo (2.2.2).

Por (1) e (2) acima tem-se que o modelo foi incapaz de se adaptar ao novo patamar de inflação que se apresenta após dezembro de 1986. Isto é decorrência da própria "filosofia" dos modelos de Box e Jenkins, que não respondem a mudanças repentinas do nível da série.

A Tabela 3 apresenta as estimativas dentro da amostra.

Pelas estimativas dentro da amostra, tem-se que o valor observado está fora do Intervalo de Confiança para os meses de 1986. III e 1986. XII, nos quais há mudança de nível na série. Deste modo, já era de se esperar que as previsões para 1987 com este modelo não fossem boas, dado que ele foi incapaz de se adaptar à mudança de nível ocorrida em 1986. XII. Foi feito, então, um exercício com a hipótese de que os dados de 1987. I e 1987. II fossem "observados", adotando-se para estes meses, respectivamente, os valores de 15,00 e 13,00[15].

Estimou-se, então, o modelo (0,1,1) × × (0,1,1) para o período 1980.I e 1987.II e previu-se 12 períodos à frente. Os resultados são dados abaixo.

$$\nabla \nabla_{12} y_t = (1 - \underset{(2,72)}{0,32} L)(1 - \underset{(4,56)}{0,72} L^{12})\epsilon_t \qquad (3.2.3)$$

$\sigma_w^2 = 8,20$; Q(22) = 16,73; AIC = 362,71; SBC = 367,29

Pelas estimativas acima tem-se que há um aumento, embora marginal, na variância do IPCA estimado. A Tabela 4 abaixo apresenta as previsões dentro da amostra.

Através da Tabela 4, observa-se que os valores previstos dentro da amostra apresentam os mesmos problemas com relação aos anos de mudança de nível, embora em menor escala. A Tabela 5, abaixo, apresenta as previsões para o ano de 1987 com 1987.I e 1987.II "observado".

TABELA 3

ESTIMATIVA DENTRO DA AMOSTRA
Usando o Modelo (3.2.2)
Método Box e Jenkins

ANO	TAXA DO IPCA	PREVISÃO	INTERVALO INFERIOR	INTERVALO SUPERIOR
86.I	16,20	15,01	9,40	20,61
86.II	14,36	12,49	6,88	18,10
86.III	−0,11	15,24	9,63	20,85
86.IV	0,78	4,83	−0,73	10,43
86.V	1,40	2,37	−3,23	7,98
86.VI	1,27	1,65	−3,96	7,26
86.VII	1,19	3,65	−1,96	9,26
86.VIII	1,68	0,98	−4,63	6,59
86.IX	1,72	2,50	−3,11	8,11
86.X	1,90	1,69	−3,92	7,29
86.XI	3,29	1,50	−4,11	7,11
86.XII	11,65	3,18	−2,43	8,79
87.I	★	10,90	5,29	16,50
87.II	★	8,05	1,57	14,52

[15]Uma vez que não existe nenhuma justificativa para estes números, talvez fosse necessário um *experimento* adotando diversas estimativas para estes dois meses. Uma possível explicação para a utilização destes números é o fato de pertencerem à parte superior do Intervalo de Confiança para as previsões de 1987.I e 1987.II. Como é esperada uma aceleração inflacionária para estes dois meses, a adoção destes números não é de todo arbitrária embora este exercício já esteja desatualizado, resolveu-se mantê-lo porque o objetivo deste documento é fornecer uma abordagem alternativa aos modelos ARIMA. Na Seção 5 são apresentadas as previsões para 1987, adotando-se 1987.I como observado.

TABELA 4
ESTIMATIVA DENTRO DA AMOSTRA
Usando o Modelo (3.2.3)
Método Box e Jenkins

ANO	TAXA DO IPCA	PREVISÃO	INTERVALO INFERIOR	INTERVALO SUPERIOR
86.I	16,20	15,39	9,78	21,00
86.II	14,36	12,69	7,08	18,30
86.III	−0,11	15,51	9,90	21,23
86.IV	0,78	3,39	−2,23	9,00
86.V	1,40	1,44	−4,17	7,05
86.VI	1,27	1,29	−4,32	6,90
86.VII	1,19	3,54	−2,07	9,15
86.VIII	1,68	0,74	−4,87	6,35
86.IX	1,72	2,49	−3,12	8,10
86.X	1,90	1,54	−4,08	7,15
86.XI	3,29	1,50	−4,11	7,11
86.XII	11,65	3,40	−2,21	9,01
87.I	15,00	11,86	6,25	17,47
87.II	13,00	11,12	5,51	16,73

TABELA 5
PREVISÃO PARA O ANO DE 1987 COM 87.I E 87.II
"Observados" usando o Modelo (3.2.3)
Método Box e Jenkins

ANO	PREVISÃO	DESVIO PADRÃO	INTERVALO INFERIOR	INTERVALO SUPERIOR
87.III	9,6075	2,8058	3,9960	15,2190
87.IV	8,7238	3,3869	1,9499	15,4976
87.V	8,7645	3,8820	1,0003	16,5286
87.VI	8,6345	4,3234	−0,0021	17,2812
87.VII	10,2520	4,7189	0,8140	19,6899
87.VIII	9,5135	5,0861	−0,6586	19,6856
87.IX	10,3292	5,4284	−0,5275	21,1859
87.X	10,0658	5,7503	−1,4348	21,5664
87.XI	10,2455	6,0552	−1,8648	22,3559
87.XII	13,0604	6,3454	0,3696	25,7513
87.I	16,0710	6,6230	2,8251	29,3169
87.II	13,4495	6,8894	−0,3291	27,2282
ACUMULADA PARA 1987	234,40		34,58	677,43

Segundo a Tabela 5, que contém as previsões para 1987. III a 1988. II, pode-se obter a projeção da inflação para 1987, condicional aos valores de 1987.I e 1987.II. Obtém-se então, uma inflação de 234,40%, na média, com um Intervalo de Confiança que vai de 34,58 a 677,43%[16].

Dada a tendência atual dos reajustes de preços nos primeiros dois a três meses do ano de 1987, o disparo do gatilho em janei-

[16] Adotou-se para os valores *observados* de 1987.I e 1987.II um erro-padrão nulo, isto é, os limites inferiores e superiores para os intervalos de confiança foram calculados supondo estes valores como dados.

ro e, possivelmente, em março[17], o intervalo mais razoável seria a parte superior do mesmo, isto é, a inflação deveria estar entre 234,40 e 677,43%.

A seguir, apresentam-se os resultados deste exercício utilizando-se uma abordagem alternativa: a dos Modelos Estruturais de Séries Temporais.

4 — PREVISÕES DO IPCA VIA MODELOS ESTRUTURAIS

A principal desvantagem dos modelos ARIMA é a sua característica de tratar os diversos componentes da série como fixos ao longo do tempo. Por exemplo, se o componente de tendência nos modelos de Box e Jenkins é suposto como linear, este crescerá sempre a uma taxa fixa. Isto parece pouco provável quando se utiliza a série do IPCA, pois o Plano Cruzado fez com que a tendência desta série, pelo menos nos primeiros meses, fosse reduzida consideravelmente.

Dever-se-ia, portanto, ter modelos que pudessem se adaptar a estas mudanças do nível da tendência, assim como mudanças do padrão sazonal. A classe de modelos que tem esta característica é chamada de Modelos Estruturais de Séries Temporais[18].

A idéia desta modelagem é tratar uma série temporal com soma (ou produto) de cada um dos componentes da mesma, a saber, de tendência, de ciclo, de sazonalidade e irregular. Desta forma, a série temporal está na sua forma estrutural, no sentido econométrico, e pode-se mostrar que a forma reduzida é um modelo ARIMA. Portanto, tem-se uma racionalidade para a utilização dos modelos ARIMA de Box e Jenkins. Ao se escrever uma série temporal em forma estrutural, informações *a priori* são usadas para se especificar cada um dos componentes, implicando um modelo com maior aderência aos dados e principalmente menor variabilidade. Isto se deve à utilização de informação *a priori* para se especificar os componentes do modelo, diminuindo-se portanto, a variância das estimativas.

O componente de tendência em vez de ser global, é modelado de forma local. Por exemplo, em vez de adotar como hipótese uma tendência linear global, isto é, uma reta para todo o período amostral, supõe-se que o nível e a taxa de crescimento desta reta variam suavemente ao longo do tempo, ou seja, tem-se uma tendência linear local[19]. O componente sazonal no modelo *airline* de Box e Jenkins é bastante semelhante àquele suposto no modelo estrutural[20]. O componente de ciclo, no modelo estrutural, é especificado para captar as flutuações de longo prazo da série. A especificação deste componente é tal que a sua forma reduzida corresponde a um modelo ARMA(2,1), onde as raízes do polinômio auto-regressivo são complexas com módulo igual ao coeficiente de amortecimento do ciclo. A racionalidade deste modelo é dada pelo componente auto-regressivo, já que um AR(2) com raízes complexas pode ser escrito com soma de senóides e co-senóides que geram movimentos cíclicos. A introdução do componente MA(1) permite que este ciclo varie ao longo do período amostral, dado que este componente funciona como um realimentador dos "choques" estocásticos.

Partiu-se do modelo estrutural geral, isto é:

$$Y_t = \mu_t + \psi_t + \gamma_t + \epsilon_t \qquad (4.1)$$

onde μ_t é o componente de tendência, ψ_t o de ciclo, γ_t o de sazonalidade e ϵ_t o irregular, que é tal que $\epsilon_t \sim NID(0,\sigma_\epsilon^2)$.

[17]Uma hipótese "otimista" com relação ao gatilho é que o seu disparo implica um aumento de 11% nos preços, dado que a mão-de-obra na indústria representa 55% dos custos variáveis. Mas supondo que este aumento da folha salarial é totalmente repassado aos preços, é razoável assumir que o novo nível da taxa vá para 20%.

[18]Para uma descrição destes modelos, tanto com a abordagem Bayesiana quanto clássica, ver Migon e Pereira (1986).

[19]Um modelo possível seria permitir que o nível, α, e a taxa de crescimento, β, da tendência variassem segundo um *random walk*. Este tipo de modelo é inadequado, já que gera descontinuidade na função. Pode-se mostrar, ver Harvey e Pereira (1987), que o modelo que será adotado aqui induz continuidade na função de previsão.

[20]Para uma descrição de como definir os componentes de tendência e de sazonalidade em modelos estruturais, ver Harvey e Pereira (1987).

É possível, através deste tipo de modelagem, como subproduto, estimar cada um dos componentes da série[21]. Estimou-se o modelo (4.1) para o período 1980.I a 1986.XII sem o componente de ciclo[22]. Os resultados foram[23]:

$\sigma_\epsilon^2 = 2.5313\ (0{,}4898)$

$\sigma_\gamma^2 = 0{,}1906 * 10^{-2}(0{,}6877) * 10^{-2})$

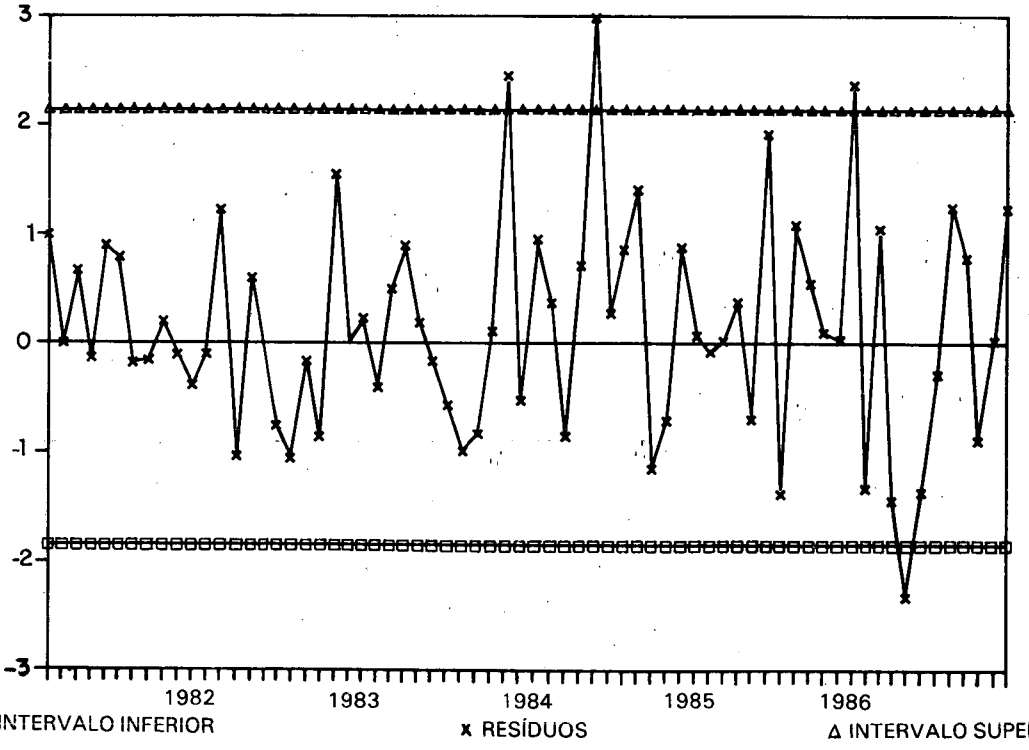
$\sigma_\beta^2 = 0{,}6218 * 10^{-5}(0{,}8794 * 10^{-05})$

$\sigma_\mu^2 = 0{,}5687 * 10^{-1}(0{,}6749 * 10^{-1})$

Observa-se que os componentes de nível, de taxa de crescimento da tendência (isto é, σ_μ^2 e σ_β^2) e de sazonalidade (isto é, σ_γ^2) não apresentam variâncias estatisticamente diferentes de zero,mas optou-se por mantê-los, dado que, ao se retirar o componente de sazonalidade, a aderência do modelo piora[24].

Os resíduos foram testados para normalidade e não se rejeitou esta hipótese[25]. Pelo gráfico abaixo, dos resíduos padronizados,

GRÁFICO 1
RESÍDUOS PADRONIZADOS



[21] Esta é outra das vantagens da modelagem estrutural, já que através do algoritmo de suavizamento [para uma descrição teórica deste algoritmo ver Migon e Pereira (1986) e para uma aplicação ver Pereira (1986)], pode-se obter a "melhor" estimativa dos componentes (não observados) da série. É possível portanto, recompor os componentes de tendência, sazonalidade e ciclo de série.

[22] Tentou-se, inicialmente, o modelo geral (4.1), mas as estimativas do componente de amortecimento do ciclo e da variância do componente de sazonalidade implicavam um ciclo não-estacionário e uma sazonalidade determinística. Optou-se, então, por excluir o componente de ciclo, o que é razoável pelo fato de se ter somente cinco anos de dados mensais.

[23] Para este tipo de modelo são apresentados como resultados as variâncias dos erros de cada um dos componentes. Para uma descrição do método de estimação, ver Migon e Pereira (1986).

[24] Pode-se definir, para modelos estruturais, uma estatística semelhante ao coeficiente de determinação de uma regressão, isto é, R^2, que neste caso é definido utilizando-se resíduos recursivos, que são calculados através da previsão um passo à frente. E é comparando estes valores que se julga a aderência dos modelos.

[25] Os testes de normalidade são: de *skewness*, cuja estatística teve o valor de 1,3807: de *kurtosis*, com valor de 0,4243; e normalidade propriamente dita, que testa simultaneamente *skewness* e *kurtosis* e cujo valor foi de 1,8050. Estas estatísticas têm distribuições Chi-quadrado com graus de liberdade 1, 1 e 2, respectivamente.

observa-se que os meses de 1984.I, 1984.VII, 1986.I e 1986.III são os que apresentam valores extremos, isto é, estão fora de dois desvios padrão; para o mês de 1984.I, o valor extremo é decorrência dos Decretos-Leis n.os 2.024, 2.025 e 2.065 para a política de reajustes salariais; para o mês de 1984. VII, uma possível justificativa para este valor seria o expurgo feito no índice; e, para os meses de 1986.I e 1986.III, as mudanças nos patamares da inflação (no primeiro, para cima, e, no segundo, para baixo) justificam os valores extremos observados.

Testou-se também a independência dos resíduos através da estatística de Q de Box e Pierce e não se rejeitou esta hipótese. O gráfico dos resíduos padronizados apresenta um aumento na variância dos mesmos, principalmente a partir de 1984, o que é confirmado através do teste de heterocedasticidade[26], já que a hipótese de variância constante é rejeitada[27].

Embora se tenha consciência de que o modelo não passa por todos os testes estatísticos a que foi submetido, utilizou-se o mesmo para prever a inflação para 1987. As previsões, com os respectivos Intervalos de Confiança, encontram-se na Tabela 6, abaixo.

A inflação acumulada foi de 275,12%, na média, com um Intervalo de Confiança de 148,32 a 460,41%. Comparando-se estes resultados com os do modelo ARIMA correspondente, observa-se que:

1) pela Tabela 2, a previsão da inflação acumulada para 1987, utilizando-se o modelo ARIMA (3.2.2), era de 110,28%, na média, com um Intervalo de Confiança de −29,53%, 470,36%, valores que correspondem a uma inflação mensal de 6,42%, na média, com um Intervalo de Confiança de −2,87%, 15,61%;

2) para o modelo estrutural (3.1), os valores previstos para 1987 correspondem a uma inflação mensal, na média, de 11,66%, com um Intervalo de Confiança de 7,87%, 15,45%;

3) mesmo considerando-se somente a parte superior do Intervalo de Confiança para o modelo ARIMA, chega-se a 110,28%, 470,36%, que nos parece subestimado; e

TABELA 6
PREVISÃO PARA O ANO DE 1987
Usando o Modelo Estrutural

ANO	PREVISÃO	DESVIO PADRÃO	INTERVALO INFERIOR	INTERVALO SUPERIOR
87.I	11,591	1,7689	8,0532	15,1288
87.II	11,493	1,8022	7,8886	15,0974
87.III	11,339	1,8325	7,6739	15,0041
87.IV	11,191	1,8574	7,4763	14,9057
87.V	11,112	1,8762	7,3597	14,8643
87.VI	11,146	1,8903	7,3654	14,9266
87.VII	11,307	1,9022	7,5026	15,1114
87.VIII	11,574	1,9146	7,7448	15,4032
87.IX	11,898	1,9300	8,0379	15,7581
87.X	12,215	1,9508	8,3135	16,1165
87.XI	12,464	1,9781	8,5079	16,4202
87.XII	12,599	2,0117	8,5756	16,6224
ACUMULADA PARA 1987	275,12		148,32	460,41

[26]Este teste verifica se a variância dos resíduos é constante ao longo do período amostral. A estatística do teste tem uma distribuição F que, neste caso, tem valor crítico de 2,07, dado que é uma F (22,22).

[27]Uma possível correção seria supor que a variância dos erros varia com o nível da série do instante anterior. No estágio atual do programa utilizado para estimar modelos estruturais, tal correção não é possível. Uma alternativa seria a utilização de um modelo Bayesiano de previsão que permite modelar a variância em função do nível da série, ver Brasil e Migon (1986).

4) considerando-se somente a parte superior do Intervalo de Confiança, para o modelo estrutural, chega-se ao intervalo 275,62%, 460,41%, que nos parece mais realista, dado que a taxa mensal, pelo Item 2 fica entre 11,66 e 15,45%.

Pelos itens acima, observa-se que a discrepância entre valores mínimos e máximos para os intervalos de confiança, utilizando-se o modelo estrutural, são bem menores do que para o modelo ARIMA. Além disso, considerando-se o Item 4 e dada a aceleração inflacionária dos últimos meses, é bastante razoável a previsão de uma taxa mensal média no intervalo citado.

A seguir reaplica-se o experimento feito para o modelo ARIMA supondo que 1987.I e 1987.II são "observados". O modelo estimado também foi do tipo tendência mais sazonalidade, tendo sido obtidos os seguintes resultados:

$\sigma_\epsilon^2 = 2{,}6016\ (0{,}4860)$

$\sigma_\mu^2 = 0{,}000$

$\sigma_\beta^2 = 0{,}2818 * 10^{-3} (0{,}2998 * 10^{-3})$

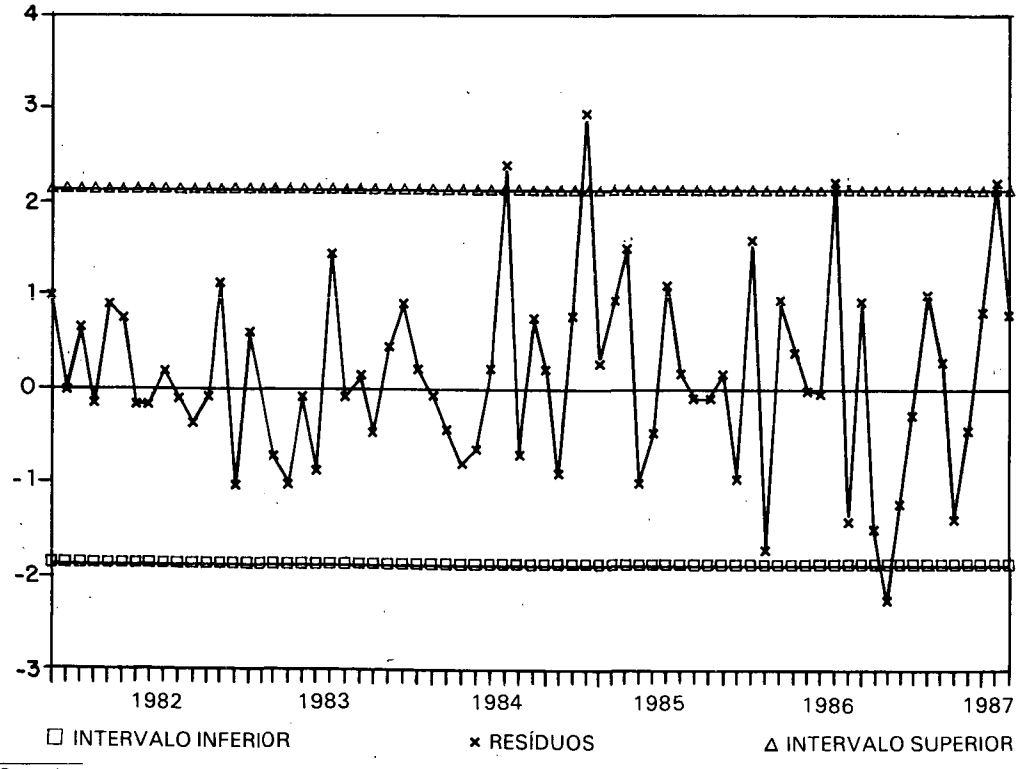
$\sigma_\gamma^2 = 0{,}2820 * 10^{-1} (0{,}2478) * 10^{-1})$

Observa-se que a variância da taxa de crescimento da tendência não é estatisticamente diferente de zero, mas, como o nível deste componente tem variância não-nula, segue-se que o componente da tendência é estocástica.

Testou-se independência, normalidade e heterocedasticidade dos resíduos[28], não se rejeitando o fato de que os erros são independentes, normais, mas não têm variância constante. Alguma correção seria necessária, mas, conforme já mencionado acima, no estágio atual isto não é possível.

Pelo Gráfico 2 abaixo, dos resíduos padronizados, nota-se que os valores extremos observados no caso anterior permanecem, mas neste caso 1987.I também é um valor extremo, o que é decorrência da mudança de patamar do nível da inflação.

GRÁFICO 2
RESÍDUOS PADRONIZADOS



[28]Os valores para as estatísticas dos testes de *skewness, kurtosis,* normalidade e heterocedasticidade foram, 1,2854, 0,46123, 1,7466 e 3,292, respectivamente, cujas distribuições são Chi-quadrada para as três primeiras (com 1,1 e 2 graus de liberdade, respectivamente) e F (23,23) para a última.

As projeções para 1987.III e 1988.II são apresentadas na Tabela 7.

Tem-se uma inflação acumulada para 1987, supondo 1987.I e 1987.II como "observados", de 326,78%, com um Intervalo de Confiança de 196,71%, 505,83%, que corresponde a uma taxa mensal de 12,85%, na média, com um intervalo de 9,05%, 16,20%. Comparando-se com os resultados do modelo ARIMA correspondente, observa-se que:

1) o Intervalo de Confiança para o modelo ARIMA tem uma dispersão muito grande, mesmo quando só se considera a parte superior do intervalo; e

2) as taxas mensais de inflação previstas pelo modelo ARIMA, considerando-se somente a parte superior do Intervalo de Confiança, estão entre 10,58 e 18,64%, apresentando também uma dispersão maior do que as obtidas pelo modelo estrutural, entre 12,85 e 16,20%.

5 — CONCLUSÕES

Apresentaram-se neste trabalho alguns exercícios de previsões da taxa de inflação para 1987, que foram feitas utilizando-se modelos ARIMA e Modelos Estruturais de Séries Temporais. Para os primeiros, o modelo base foi o *airline* de Box e Jenkins, enquanto que, para os modelos estruturais, partiu-se do modelo geral, verificando-se que, após alguns testes, o melhor modelo para os dados era do tipo tendência + sazonalidade.

As previsões para o modelos apresentavam os seguintes padrões:

1) os modelos ARIMA tendem a subpredizer a taxa de inflação se a amostra terminar em 1986.XII, enquanto que, para os modelos estruturais o Intervalo de Confiança para as previsões está mais próximo da realidade;

2) considerando-se somente a parte superior dos intervalos de previsão, observa-se que, para os modelos ARIMA, é 110,28%, 470,36%, enquanto que para os modelos estruturais é de 275,62%, 460,51%, o que nos parece razoável;

3) adotou-se, então, os meses de 1987.II como "observados"; a previsão da taxa de inflação acumulada para 1987, condicionada a estes dois meses, foi de 234,40%, com um Intervalo de Confiança de 34,58%, 677,43%, para o modelo ARIMA, e de 326,78%, com um Intervalo de Confiança de 196,71%, 505,83%, para o modelo estrutural;

TABELA 7
PREVISÃO PARA O ANO DE 1987
COM 87.I E 87.II "OBSERVADOS"
Usando o Modelo Estrutural

ANO	PREVISÃO	DESVIO PADRÃO	INTERVALO INFERIOR	INTERVALO SUPERIOR
87.III	12,731	1,8643	9,0024	16,4596
87.IV	12,103	1,9288	8,2455	15,9605
87.V	11,555	1,9822	7,5906	15,5194
87.VI	11,271	2,0168	7,2373	15,3046
87.VII	11,364	3,0343	7,2954	15,4326
87.VIII	11,848	2,0416	7,7648	15,9312
87.IX	12,629	2,0452	8,5387	16,7193
87.X	13,537	2,0492	9,4386	17,5859
87.XI	14,366	2,0588	10,2483	18,4837
87.XII	14,931	2,0833	10,7644	19,0976
88.I	15,119	2,1330	10,8531	19,3849
88.II	14,916	2,2116	10,4926	19,3393
ACUMULADA PARA 1987	326,78		196,71	505,83

4) mesmo considerando-se somente as partes superiores dos intervalos de confiança, os resultados para o modelo estrutural são mais razoáveis; e

5) considerando-se somente as partes superiores dos Intervalos de Confiança, as taxas mensais de inflação para 1987 ficam entre 10,58 e 18,64% para o modelo ARIMA e entre 12,85 e 16,20% para o modelo estrutural.

6 — POST SCRIPTUM

Com a divulgação dos índices de janeiro, fevereiro, março, abril e maio de 1987, reestimou-se o modelo estrutural. A previsão para a taxa de inflação acumulada para 1987 foi de 785,14% com um Intervalo de Confiança de 408,66%, 1375,79%. A taxa mensal média de 19,93, com um intervalo de 14,52%, 25,15%. Os resultados para as previsões da taxa de inflação para os meses de 1987 estão apresentados na Tabela 8 e no Gráfico 3.

Cabe notar que, devido à aceleração inflacionária dos últimos meses, a incerteza traduz-se no modelo pelo aumento das variâncias de predição acarretando intervalos de confiança com uma variabilidade muito grande. Considerando-se a previsão para 1987.VI, é possível que esta avaliação seja otimista, dado que o próprio IBGE divulgou uma previsão acima do intervalo superior, que é de 26,65%.

Caso o Plano Bresser não tivesse sido declarado, estas previsões deveriam servir de alerta aos tomadores de decisão, uma vez que, se a economia continuasse descontrolada, seria bem possível que uma taxa acumulada de 1.375,79%, que é o intervalo superior das previsões obtidas, fossem otimistas.

Com a declaração do Plano Bresser foi possível reduzir a atual tendência da taxa de inflação. Esperava-se que a taxa de inflação se estabilizasse num novo patamar de 3%. Fez-se uma intervenção nas previsões, permitindo que o nível e a taxa de crescimento da tendência fossem reduzidas[29]. Como após noventa dias de congelamento voltará

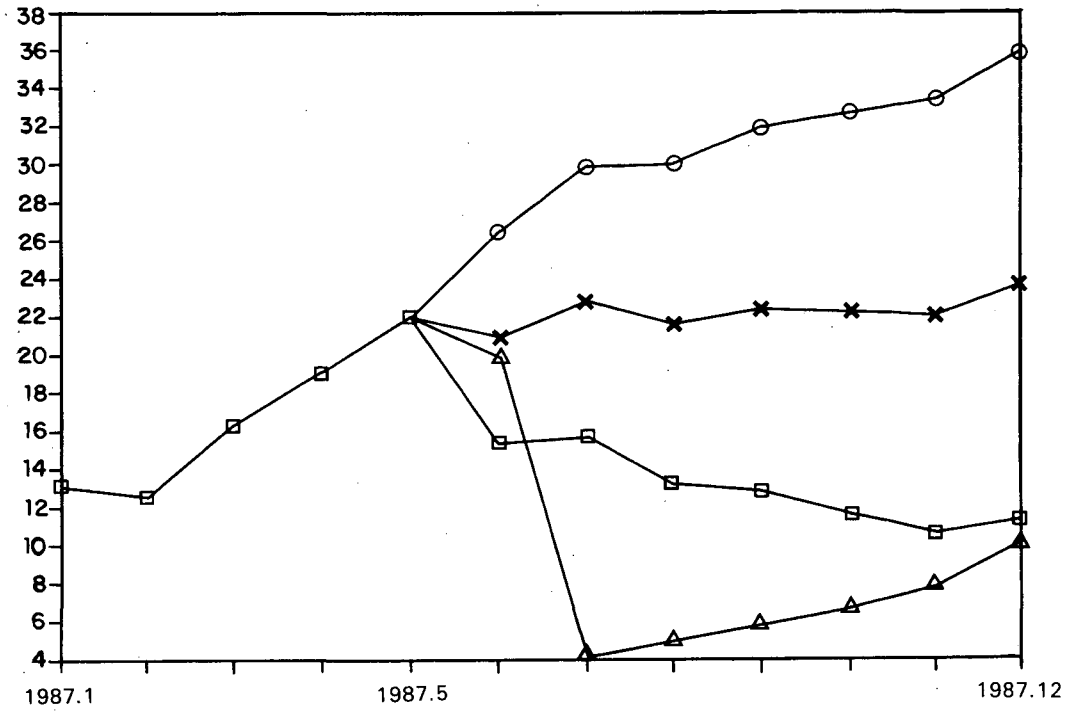
TABELA 8
PREVISÃO PARA O ANO DE 1987
COM 87.I, 87.II, 87.III, 87.IV E 87.V OBSERVADOS

Usando o Modelo Estrutural

ANO	PREVISÃO	DESVIO PADRÃO	INTERVALO INFERIOR	INTERVALO SUPERIOR
87.I	13,2150	–	–	–
87.II	12,6400	–	–	–
87.III	16,3700	–	–	–
87.IV	19,1000	–	–	–
87.V	22,09	–	–	–
87.VI	21,070	2,7886	15,4928	26,6472
87.VII	22,937	3,5748	15,7874	30,0866
87.VIII	21,772	4,2215	13,329	30,215
87.IX	22,494	4,7795	12,935	32,053
87.X	22,278	5,2788	11,7204	32,8356
87.XI	22,150	5,7349	10,6802	33,6198
87.XII	23,696	6,1573	11,3814	36,0106
ACUMULADA PARA 1987	785,14		408,66	13,7579

[29]Este exercício foi feito logo após a divulgação do índice de junho. O Gráfico 3 apresenta os resultados denotados por p.c.c. (previsões com choque).

GRÁFICO 3
PREVISÃO DO IPCA
Com 1987.1 até 1987.5 observados



□ INTERVALO INFERIOR SEM CHOQUE
ΔPREVISÕES COM CHOQUE
INTERVALO SUPERIOR SEM CHOQUE
x PREVISÕES SEM CHOQUE

a flexibilização dos preços, assumiu-se uma nova intervenção em outubro, que permitiu uma recuperação no nível e na taxa de crescimento da tendência. A previsão da taxa de inflação acumulada para 1987 foi, neste caso, de 282,79%, o que dá uma taxa de 4,27; 5,17; 5,98; 6,87; 7,96; 10,17 para os meses de julho, agosto, setembro, outubro, novembro, dezembro, respectivamente.

BIBLIOGRAFIA

BOX, G. E. P. ; JENKINS, G. M. *Time series analysis:* forecasts and control. San Francisco, Holden-Day, 1976.

BRASIL, G. H. ; MIGON, H. S. *Modelagem estrutural e previsão bayesiana:* uma aplicação e comparação com modelos ARIMA. Anais do VIII Encontro Brasileiro de Econometria, p. 413-40.

_____;_____; SOUZA, R. C.; PORTUGAL, S. *Modelos de previsão para séries de produção e preços:* metodologia bayesiana e Box e Jenkins para séries temporais. Rio de Janeiro, IPEA/INPES, 1986 (Texto para Discussão Interna, 92).

GRANGER, C. W. J. *Forecasting in business and economics.* New York, Academic Press, 1980.

HARVEY, A. C. ; PEREIRA, P. L. V. *Trend, seasonality and seasonal adjustment.* Artigo apresentado no Seminar in Applied Statistics: "Statistical Methods for Cyclical and Seasonal Analysis", Mar del Plata, Argentina, 1987.

MIGON, H. S. ; PEREIRA, P. L. V. *Modelagem estrutural:* abordagem bayesiana e clássica. Texto do curso do I Encontro Regional de Econometria da Região Sul. Santa Maria, RS, 1986.

PEREIRA, P. L. V. Estimação do hiato do produto via componentes não observados. *Revista de Econometria,* v. 6, n° 2, p. 47-68, nov. 1986.

RESUMO

Os métodos de previsão de inflação podem ser divididos em dois grupos: Planilhas de Cálculo e *Model-Based*. Este artigo apresenta uma descrição destes dois grupos de métodos. Na classe dos *Model-Based* duas metodologias são descritas: modelos globais do tipo Box e Jenkins e modelos locais do tipo Estrutural. Argumenta-se que os modelos estruturais têm uma performance superior aos modelos do tipo Box e Jenkins, devido a sua característica local que permite uma adaptação mais rápida a mudanças estruturais como as que ocorreram durante o período do Plano Cruzado.

ABSTRACT

Methods for predicting inflation could be divided into two groups: Worksheet and Model-Based. In the Model-Based class, two methodologies are presented: global models (Box and Jenkins) and local models (Structural Models). The Structural Models, to predict inflation, perform better than the Box and Jenkins's models because they adapt more quickly to structural changes as has been the case after the Cruzado Plan.

PESQUISA ELEITORAL — UMA ANÁLISE BAYESIANA

Isabel G.S. Furtado de Mendonça *
Helio S. Migon **

1 – INTRODUÇÃO

O objetivo deste trabalho foi antecipar, com boa estimativa, a proporção de votos a ser obtida por determinado candidato, nas eleições para Prefeito da Cidade do Rio de Janeiro, em 1985.

A parte teórica está diretamente ligada à análise bayesiana de um modelo hierárquico apropriado, inspirado no trabalho feito por Bernardo (1984), onde a amostra, composta por eleitores de determinados locais de votação, foi selecionada através de uma medida de informação logarítmica (Kullback, 1968) tal que possibilitasse identificar os locais de votação representativos do comportamento político da população. Outro ponto a salientar é que nessa proposta metodológica não se consideram as "heranças mercadológicas " para o dimensionamento da amostra, isto é, não se supõe, no dimensionamento da amostra, que o voto esteja fortemente relacionado com distribuições etárias, de renda e, nem mesmo, de sexo. Por outro lado, as facetas multidimensionais do problema são devidamente consideradas. Uma vez identificados os locais de votação, a amostra aleatória de eleitores é selecionada em cada um desses pontos.

Na Seção 2, discute-se o desenho de amostra, que consiste na escolha das Zonas Eleitorais mais representativas e dos locais de votação onde seriam entrevistados os eleitores. O modelo bayesiano para análise do resultado está descrito na Seção 3 sob o título Modelo bayesiano para Análise dos Resultados. As estimativas dos percentuais de voto para os cinco primeiros candidatos, com os respectivos intervalos de credibilidade, estão descritos na Seção 4 — Resultados Obtidos. Finalmente, na Seção 5 — nos Comentários e Extensões, fazem-se várias observações e propõem-se extensões ao método apresentado.

2 – DESENHO DE AMOSTRA

Todo o trabalho de identificação da amostra foi feito a partir dos dados disponíveis sobre as eleições de 1982, para o Governa-

* Aluna de mestrado do IMUFRJ
** Do SERPRO/UFRJ, atualmente professor da Escola Nacional de Ciências Estatísticas/IBGE.

dor do Estado. Usou-se a divisão natural do município em Zonas Eleitorais (25 em 1982) e assumiu-se que há certa estabilidade temporal no comportamento eleitoral.

Para identificar as Zonas Eleitorais mais representativas da população votante, isto é, em que a proporção mais se assemelhou ao resultado final da eleição anterior, foi usada a medida l_i correspondente à distância entre as distribuições de probabilidade de cada área e o resultado real obtido.

2.1 – Escolha das Zonas Eleitorais mais representativas

Seja n_{ij} – número de votos para o partido j, na Zona Eleitoral i, referente às eleições de 1982;

m – número de partidos que disputaram a eleição anterior;

$N_i = \sum_{i=1}^{m} n_{ij}$ = número de votos da zona eleitoral i, e

γ_j = proporção de votos do candidato j na última eleição.

Então, o estimador de Bayes da probabilidade de um eleitor da i-ésima Zona Eleitoral votar no j-ésimo candidato, como será visto posteriormente, é dado por:

$$\hat{Y}_{ij} = (n_{ij} + 1/2)/(N_i + (m+1)/2); \quad j = 0,1,\ldots,m$$

e a medida de distância desejada é:

$$\ell_i = \sum_{j=0}^{m} \gamma_j \log (\gamma_j/\hat{\gamma}_{ij})$$ (Kullback-1968, Bernardo-1979).

De acordo com a definição anterior, quanto menor for a medida ℓ_i, mais representativa do total é a i-ésima Zona Eleitoral.

Neste trabalho, ordenaram-se as vinte e cinco Zonas Eleitorais existentes em 1982 e, então, escolheram-se as cinco mais representativas da cidade, ou seja, aquelas com os menores valores de ℓ_i.

Essas Zonas Eleitorais acham-se descritas no apêndice e correspondem aos bairros que se distribuem ao longo das linhas férreas da Central do Brasil e da Leopoldina.

2.2 – Obtenção da amostra final

Por coerência, deseja-se utilizar o mesmo critério para escolher quatro Seções Eleitorais, em cada uma das Zonas Eleitorais previamente selecionadas.

Por questões operacionais, como desagregação geográfica da amostra e pequeno número de entrevistadores disponíveis, decidiu-se trabalhar com "locais de votação", isto é, prédios públicos, em geral, onde uma ou mais Seções são localizadas.

Isso gerou uma dificuldade a mais: como selecionar os locais de votação mais representativos de cada uma das Zonas Eleitorais já escolhidas?

A solução encontrada foi ordenar todas as Seções Eleitorais dentro de cada uma das respectivas Zonas, de acordo com a medida de distância já apresentada. Feito isso, agregaram-se as Seções Eleitorais em "locais de votação", a partir de dados fornecidos pelo TRE/SERPRO, e calculou-se a média $\bar{\ell}$ para cada local de votação na respectiva Zona Eleitoral.

Dessa maneira, em cada Zona Eleitoral selecionada escolheram-se quatro locais de votação representativos, ou seja, com menor medida ℓ. Finalmente, em cada local de votação, foram entrevistados aleatoriamente cinqüenta eleitores, totalizando 1.000 entrevistas de "boca de urna".

Nesse ponto, é oportuno lembrar que o problema da composição da amostra, isto é, número de locais "similares" versus número de eleitores a entrevistar em cada local, merece ser estudado de forma adequada. Esse é um problema de decisão da maior relevância, pois depende de um compromisso entre custo e precisão.

Vale recordar que as Zonas Eleitorais foram escolhidas de forma a garantir homogeneidade por serem as mais "parecidas", em termos eleitorais, com a realidade municipal no pleito de 1982. Considerando, ademais, que em cada "local de votação" os eleitores devem apresentar comportamento heterogêneo, para representar o universo, concluiu-se pela necessidade de maior fração amostral nesse estágio (eleitores), comparativamente à fração amostral do primeiro estágio (locais de votação).

Partindo-se, então, da disponibilidade de 20 entrevistadores, apenas, julgou-se mais adequado distribuí-los entre "locais de votação" distintos, cada um com 50 questionários, número compatível com o tempo destinado, pela legislação, ao exercício do voto.

Experiência piloto foi realizada semanas antes das eleições para se avaliar tanto o critério de seleção das Zonas Eleitorais como o questionário a ser usado. Os resultados obtidos, embora satisfatórios, suscitaram o aprimoramento do questionário.

3 – MODELO BAYESIANO PARA ANÁLISE DOS RESULTADOS

3.1 – Definições necessárias

Sejam: θ_{ij} – probabilidade de um eleitor do local de votação i votar no candidato j;

N_i – tamanho da amostra no local de votação i;

n_{ij} – número de votos para o candidato j no local de votação i;

m – número de candidatos.

Note-se que $N_i = n_{io} + n_{il} + \ldots n_{im}$, onde n_{io} é o número de respostas não especificadas.

O vetor aleatório $(n_{io}, n_{il} \ldots, n_{im})'$ tem distribuição multinomial com parâmetros $\theta_{io}, \theta_{il}, \ldots, \theta_{im}$, onde $\theta_{ij} \geq o$ e $\sum_{j=1} \theta_{ij} = 1$.

3.2 – Estimador Bayesiano usando perda quadrática

a) Para se obter $\hat{\theta}_{ij}$, o estimador de Bayes para θ_{ij}, deve-se começar pela distribuição *a priori* para os θ's.

Como não se deseja usar outra informação a não ser aquela fornecida pelos dados amostrais e, também, por questões de simplificação do trabalho operacional, optou-se por uma distribuição *a priori* não informativa de Jeffreys (Box-Tiao, 1973), apropriada para os parâmetros de uma distribuição multinomial como a descrita no apêndice. Dessa maneira encontra-se:

$\pi(\theta) \quad \alpha \prod_{i,j} \theta_{ij}^{-1/2}$,i – local de votação e

j – número de candidato.

b) Determinada a distribuição *a priori* e a função de verossimilhança, encontra-se a distribuição *a posteriori* para θ, como sendo proporcional ao produto da distribuição *a priori* pela função de verossimilhança.

Combinando-se a distribuição *a priori* não informativa de Jeffreys e a distribuição multinomial, chega-se à distribuição *a posteriori*, que pertence à família de distribuições de Dirichlet (ver apêndice) com parâmetros $n_{ij} + 1/2$, j = 0, 1, 2, ... m.

c) O estimador de Bayes, para a proporção de eleitores do candidato j no local de votação i, é dado pela média de distribuição *a posteriori*:

$$\hat{\theta}_{ij} = E[\theta_{ij}/\text{dados}] = (n_{ij} + 1/2)/\left(N_i + \frac{(m+1)}{2}\right)$$

$i = 1,2,\ldots n$

$j = 0,1,\ldots m$

Note-se que, mesmo no caso em que m = 2 (partidos), não se chega ao estimador usual n_{ij}/N_i.

3.3 – Descrição do comportamento eleitoral do Município do Rio de Janeiro

Fazendo-se uma transformação de escala, para se trabalhar com toda a reta real, definiu-se $\hat{\theta}_{ij}$, na escala da razão de chances, correspondente ao vetor $(n_{io}, n_{il}, \ldots, n_{im})$ como sendo:

$$x_{ij} = \log [\hat{\theta}_{ij}/(1-\hat{\theta}_{ij})] = \log [(n_{ij} + \frac{1}{2})/(N_i - n_{ij} + (m - \frac{m}{2})/2)]$$

Depois de determinado o plano amostral, tem-se que o tamanho da amostra N, em cada local de votação, é fixo; além disso, cada local de votação foi escolhido de modo que a distribuição de votos fosse parecida com aquela de sua Zona Eleitoral (representativo da respectiva Zona Eleitoral) e essa, por sua vez, foi selecionada de tal maneira que tivesse a sua distribuição de votos similar à distribuição de votos da Cidade do Rio de Janeiro, ou seja, de modo que representasse o comportamento político de toda a cidade.

Logo, pode-se supor que a distribuição de votos de cada local de votação é, também,

representativa da Cidade do Rio de Janeiro e que a distribuição de votos em cada um desses locais é similar. Com esses conhecimentos, e usando o teorema central do limite, duas suposições são viáveis:

– atribuir-se ao vetor $x_i = (x_{i1}, x_{i2}, \ldots x_{im})'$ que, mesmo estando em outra escala, representa a força política de cada partido no local de votação i, uma distribuição normal multivariada com vetor de médias μ_i e matriz de precisão constante H_o.

$$p(x_i/\mu_i, H_o) \sim N_m(\mu_i, H_o)$$

– o vetor da média $\mu_i = (\mu_{i1}, \ldots, \mu_{im})$ é aproximadamente normal em cada local de votação.

$$p(\mu_i/\delta, H_1) \sim N_m(\delta, H_1),$$

onde a matriz de precisão H_1 também é constante, e a média δ, além de constante, por se tratar de locais politicamente similares entre si, descreve o comportamento político da Cidade do Rio de Janeiro na escala das razões de chance.

Com essas duas probabilidades combinadas, encontra-se um modelo hierárquico que resulta na distribuição conjunta dos vetores x_i e μ_i de cada local de votação.

$$p(x_i, \mu_i/\delta, H_o, H_1) \sim p(x_i/\mu_i, H_o)p(\mu_i/\delta, H_1)$$

Considerando agora que t é o número total de locais de votação, tem-se:

$$\prod_{i=1}^{t} p(x_i, \mu_i/\delta, H_o, H_1) = p(x_1 \ldots x_t, \mu_1 \ldots \mu_t/\delta, H_o, H_1)$$

Logo, para se encontrar a distribuição de $x_1, \ldots, x_t$ integra-se a probabilidade encontrada acima em relação a cada μ_i.

Usando as distribuições apropriadas (Lindley & Smith, 1972 e Smith, 1973) chega-se a:

$$p(x_1 \ldots x_t/\delta, H_o, H_1) = \int \cdots \int \prod_{i=1}^{t} N_m(x_i/\mu_i, H_o)$$

$$N_m(\mu_i/\delta, H_1) d\mu_1 \ldots d\mu_t$$

$$\prod_{i=1}^{t} \int N_m(x_i/\mu_i, H_o)\, N_m(\mu_i/\delta, H_1)\, d\mu_i$$

$$\prod_{i=1}^{t} N_m(x_i/\delta, H_o[H_o + H_1]^{-1}H_1)$$

Para se encontrar a média e a matriz de precisão da distribuição de cada x_i, lembre-se que:

$$E[x_i] = E[E(x_i/\mu_i)] = \delta$$

$$V(x_i) = E[V(x_i/\mu_i)] + V[E(x_i/\mu_i)] =$$

$$= E[H_o^{-1}] + V(\mu_i) = H_o^{-1} + H_1^{-1}$$

Então a matriz de precisão H, fica sendo:

$$H = [H_o^{-1} + H_1^{-1}] = H_o[H_o + H_1]^{-1}H_1$$

Esse resultado mostra que cada vetor x_i é amostra aleatória de uma distribuição normal multivariada, com vetor de média δ e a matriz de precisão constante H.

Análise bayesiana

O objetivo é encontrar uma distribuição que descreva o comportamento político do Rio de Janeiro, ou seja, uma distribuição para o vetor δ e, através dessa distribuição, seja determinado o estimador de Bayes para δ. A distribuição *a posteriori* de δ que descreve o comportamento do parâmetro, conhecidos os dados, é determinada, como já foi visto, através de função de verossimilhança dos dados amostrais e de uma distribuição *a priori* atribuída a δ.

Como já foi determinado, a função de verossimilhança é uma normal multivariada com média δ e a matriz de precisão H. Resta encontrar uma distribuição *a priori* adequada para δ.

Pelos motivos citados anteriormente, decidiu-se por uma distribuição *a priori* não informativa para δ e H.

$$\pi(\delta, H) \alpha\ |H|^{-(m+1)/2}$$

Com isso determinado, encontra-se a distribuição *a posteriori* de δ como sendo uma distribuição t multivariada com t-m graus de liberdade:

$$P(\delta/\text{dados}) \alpha [S + (\bar{x} - \delta)(\bar{x} - \delta)']^{-t/2}$$

com média $\bar{x} = (\bar{x}_i \ldots \bar{x}_m)'$, onde $\bar{x}_j =$

$$= \sum_{i=1}^{t} x_{ij}/t;\ j = 1, 2 \ldots m$$

e matriz de precisão $\Sigma = \dfrac{S}{t-m}$, onde $S =$

$$= \sum_{i=1} \{(x_i - \bar{x})(x_i - \bar{x})'/t\}$$

Como δ está na escala da razão de chances, para obter-se o estimador desejado (média da distribuição *a posteriori* usa-se a transformação inversa, encontrando:

$$\psi = e\ \bar{x}_j/(1 + e^{\bar{x}_j}); j = 1, 2, ...m$$

onde ψ_j representa a proporção de votos a ser obtida pelo candidato j na Cidade do Rio de Janeiro.

O intervalo de credibilidade de γ% para δ é obtido diretamente da distribuição t-Student com t-m graus de liberdade.

4 – RESULTADOS OBTIDOS

O Município do Rio de Janeiro possuía, em novembro de 1985, 3.222.494 eleitores, distribuídos em 26 Zonas Eleitorais, subdivididas em 7.634 Seções (urnas).

Cada local de votação comportava, em média, cinco Seções, existindo ao todo 1.500 locais, aproximadamente.

No dia da eleição, realizaram-se 1.000 entrevistas de "boca de urna" distribuídas pelos 20 locais de votação selecionados (aproximadamente 1% dos locais), pelo critério descrito.

A principal pergunta do questionário (ver apêndice) referia-se à preferência do eleitor em relação ao candidato para a Prefeitura da Capital do Estado. Informações de natureza sócio-econômica e política também foram obtidas para possibilitar a caracterização do perfil do eleitorado carioca e resolver o problema do voto indeciso através de classificações probabilísticas (Sabbag – 1987).

A tabela 1 apresenta os percentuais referentes aos resultados encontrados pelos principais candidatos.

Como mostra a tabela 1, a única classificação fora de ordem foi a 5ª, onde deveria ter aparecido o candidato Alvaro Valle (6,75% dos votos válidos). Observou-se, no entanto, que o desempenho dos dois candidatos foi estatisticamente equivalente. Além disso, a estimativa para o primeiro colocado foi cerca de dois pontos percentuais acima do real, enquanto que, para o segundo colocado, ocorreu o oposto. Especula-se que isto possa ter ocorrido pela forma como se conduziu a seleção casual dos eleitores, bem como pelo possível efeito dos não respondentes. Estes foram simplesmente desprezados e substituídos por outros. A conjectura é que os possíveis "perdedores" estariam menos propensos a revelar suas preferências. Os pesquisadores foram orientados para entrevistar os eleitores na saída do local de votação e que, tão logo terminada uma entrevista, se iniciasse a seguinte. Apesar desses comentários, percebe-se a excelente qualidade dos resultados: todos os intervalos contêm o verdadeiro valor do parâmetro.

Vários são os institutos de pesquisa que, em convênio com Jornais e Revistas, ou a serviço de partidos políticos, fazem pesquisa de opinião eleitoral ao longo da campanha. Destes, destacam-se o GALLUP e o IBOPE. O primeiro utiliza esquema de amostragem probabilística impregnada das ditas "heranças mercadológicas" já menciona-

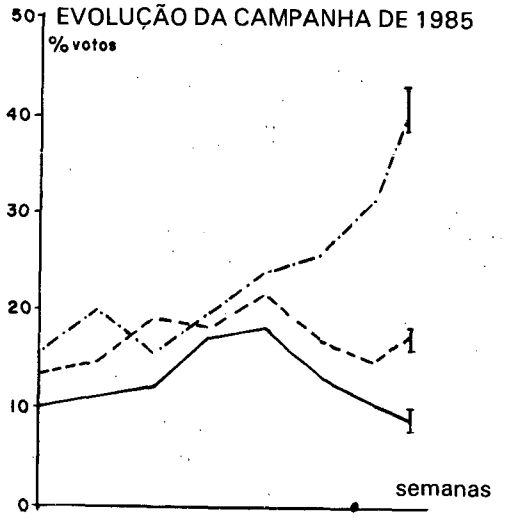
TABELA 1

RESULTADOS DA PESQUISA DE "BOCA DE URNA" NA ELEIÇÃO PARA PREFEITO DA CIDADE DO RIO DE JANEIRO – 1985
INTERVALO DE CREDIBILIDADE DE 95% E MÉDIA DA DISTRIBUIÇÃO *A POSTERIORI*

CANDIDATO	Limite inferior	Predição	Limite superior	Real
Saturnino	38,2	41,6	47,4	39,3
Medina	12,8	15,1	17,7	17,1
Jorge Leite	6,8	8,6	11,4	8,6
Cerqueira	5,0	7,1	10,0	7,0
Aarão S.	5,0	6,4	8,3	6,0

das, enquanto que o segundo emprega procedimento de amostragem por cotas. Considerando-se que, por restrições de orçamento, não se fizeram pesquisas de acompanhamento da campanha eleitoral, serão aproveitados os dados publicados pela "Revista VEJA", nos números de novembro de 85, para ilustrar graficamente a evolução da campanha.

FIGURA 1:
EVOLUÇÃO DA CAMPANHA DE 1985



-.- Saturnino --- Medina__ J. Leite
I Intervalo de Credibilidade de 95%

5 – COMENTÁRIOS E EXTENSÕES

Primeiramente, vale ressaltar o excelente desempenho do método. Apesar de baseadas em amostra de apenas 1.000 eleitores, as previsões são extremamente precisas. Talvez se obtivessem resultados com intervalos de credibilidade de comprimento ainda menores, caso houvesse sido explorado melhor o compromisso entre o número de locais e o de entrevistas a realizar.

Algumas dificuldades foram encontradas no processo de entrevistas como, por exemplo, as não-respostas. Alguns eleitores negaram-se a colaborar com o entrevistador, ora alegando desinteresse, ora evitando declarar o seu voto. É óbvio que, no futuro, precisa-se pensar em forma mais adequada de realizar as entrevistas.

As grandes vantagens do procedimento são: a clareza conceitual com que se constroem os intervalos de credibilidade e o baixo custo de obtenção das informações.

Dentre as possíveis extensões que se pretende desenvolver, destaca-se a aplicação de métodos de classificação probabilística ao problema do voto indeciso. Ao longo de uma campanha eleitoral, quando muitos eleitores ainda estão indecisos, os resultados das pesquisas pouco ajudam a determinar entre os reais candidatos. Com base num vetor de características sócio-econômica e política e de posse de uma função de classificação, a cada eleitor indeciso será imputado o seu voto. Para calibrar-se a função de classificação, precisa-se de uma amostra de treinamento. Os primeiros experimentos realizados com os dados dessa pesquisa são bastante promissores.

Ainda com relação às pesquisas de acompanhamento da campanha, imagina-se que seria interessante realizar previsões, de forma a antecipar o resultado final. Sem dúvida o instrumental adequado para tratar este problema é o "método bayesiano de Previsão". Os modelos são de natureza local, adaptam-se rapidamente a mudanças estruturais e, sobretudo, permitem, que se incorporem informações subjetivas.

6 – APÊNDICE

Neste apêndice serão apresentados uma cópia do questionário, as Zonas Eleitorais selecionadas pelo critério de divergência descrito e um sumário da análise bayesiana em modelos multinomiais.

6.1 — A — Zonas eleitorais selecionadas

A seguir, descrevem-se as zonas eleitorais selecionadas e os bairros onde se localizam os locais de votação escolhidos para as entrevistas:

1ª Zona Eleitoral
Centro
Galeão
Jardim Guanabara
Bancários

8ª Zona Eleitoral
Engenho Novo (2 locais)
Sampaio (2 locais)

9ª Zona Eleitoral
São Cristóvão
Caju
Benfica (2 locais)

11ª Zona Eleitoral
Ramos
Bonsucesso (3 locais)

22ª Zona Eleitoral
Penha (2 locais)
Jardim América
Vicente de Carvalho

6.2 — B — Cópia do questionário utilizado na pesquisa

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
CENTRO DE CIÊNCIAS MATEMÁTICAS DA NATUREZA
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE MÉTODOS ESTATÍSTICOS
LABORATÓRIO DE ESTATÍSTICA

QUESTIONÁRIO Nº _____

1. EM QUE CANDIDATO A PREFEITO VOCÊ VOTARÁ?
Jorge Leite ☐ Saturnino ☐ Medina ☐ Outro
Branco ☐ Nulo ☐ Não informado ☐ Qual?_____

2. IDADE: Até 21 ☐ De 22 a 25 ☐ De 26 a 30 ☐ De 31 a 40 ☐ De 41 a 60 ☐ Mais de 61 ☐

3. SEXO: Masc. ☐ Fem. ☐

3. ESCOLARIDADE

	Completo	Incompleto
Primário		
Secundário		
Universitário		

4.a) RENDA / SALÁRIO FAMILIAR: (Em milhões de Cr$) Até 1.0 ☐ De 1.0 a 3.5 ☐ De 3.5 a 7.0 ☐ Mais de 7.0 ☐

5.b) Nº PESSOAS NA FAMÍLIA: ____

6. PROFISSÃO:_____

7. EM QUEM VOTOU NA ELEIÇÃO PASSADA (Governador — 1982)?
Brizola ☐ Sandra ☐ Moreira ☐ Miro ☐ Lisâneas ☐ Outro ☐
Franco Teixeira

8. QUAL O PARTIDO DE SUA PREFERÊNCIA?_____

9. QUAL O PRINCIPAL PROBLEMA A SER ATACADO PELO FUTURO PREFEITO?_____

10. DOS POLÍTICOS ABAIXO, QUAL É O DE SUA PREFERÊNCIA?
Lula ☐ Brizola ☐ Ulisses ☐ Aureliano ☐ Maluf ☐ Outro
Qual?_____

TELEGRAMA: IMUFER-RJ-CAIXA POSTAL 685 DO CEP 21944-TEL:230-0017-280-9322 R.342 -- RIO DE JANEIRO — RJ

6.3 — C — Análise bayesiana do modelo multinomial

As principais alternativas para modelar-se a distribuição *a priori* dos parâmetros de um modelo multinomial são discutidas a seguir:

6.3.1 — 1) Família conjugada de distribuições *a priori*

A família de distribuições de Dirichlet tem a propriedade de reproduzir-se quando combinada a verossimilhança do modelo multinomial.

Seja a função de verossimilhança de uma amostra aleatória de tamanho n de uma distribuição multinomial:

$$\ell\,(p/x_i \ldots x_k) \alpha \prod_{i=1}^{k} p_i^{x_i}$$

onde:

$$\sum_{i=1}^{k} x_i = n \text{ e } \sum_{i=1}^{k} p_i = 1,$$

e considere a família de distribuições de Dirichlet com parâmetro $\alpha = (\alpha_1 \ldots \alpha_k)$:

$$\pi(p) \alpha \prod_{i=1}^{k} p_i^{\alpha_i - 1}$$

É fácil verificar que a distribuição *a posteriori* será

$$\pi(p)/\underline{x}) \alpha \prod_{i=1}^{k} p_i^{\theta_i - 1},$$

onde: $\theta_i = x_i + \alpha_i$

As principais características da distribuição (k-1) dimensional de Dirichlet (Bergeo, 1985) são:

$$\text{Média: } E[p_i/\underline{x}] = \gamma_i / \sum_{j=1}^{k} \gamma_i = 1, \ldots, k$$

$$\text{Variância: } VAR[p_i/\underline{x}] = \gamma_i \{\sum_{j=1}^{k} \gamma_j - \gamma_i\} /$$

$$/ \{(\sum_{j=1}^{k} \gamma_j^2)\ (\sum_{j=1}^{k} \gamma_j + 1)\}$$

Covariância: $COV[p_i, p_j/\underline{x}] =$

$$= -\gamma_i\gamma_j/\{(\sum_{\ell=1}^{k}\gamma_\ell^2)\ (\sum_{j=1}^{k}\gamma_\ell + 1)\}$$

6.3.2 — 2) Distribuição *a priori* não informativa

Em várias situações deseja-se utilizar a maquinaria bayesiana, isto é, realizar análises condicionais, mesmo sem se dispor de informações *a priori* sobre os parâmetros. As distribuições *a priori* vagas são importantes, pois não contêm informação a cerca do parâmetro.

Utilizando-se as regras de Jeffreys, obtém-se a distribuição não informativa para os parâmetros da multinomial:

$$\pi(p/x) \alpha \prod_{i=1}^{k} p_i^{\theta_i - 1/2},$$

Ainda, neste caso, tem-se uma distribuição *a posteriori* para p na família Dirichlet:

$$\pi(p) \alpha \prod^{k} p_i^{-1/2}$$

isto é: $\theta_i = x_i + 1/2$

6-4 — D — Resultados oficiais da eleição municipal de 15 de novembro de 1985

RESULTADO
DA
1ª a 25ª ELEIÇÃO MUNICIPAL DE 15 DE NOVEMBRO DE 1985

e 117ª Zonas Estado RIO DE JANEIRO Município RIO DE JANEIRO

Seções		Eleitorado	
Organizadas 7.634	Que funcionaram 7.634	Apto a votar 3.222.494	
Que não funcionaram —	Anuladas 1	Votantes 2.699.474	

I — Para PREFEITO

a) Votação obtida pelos Partidos ou Coligações — Consolidação da votação:

Partido ou Coligação	Votação
PDT — PFL — PS —	1.059.614 462.486
PMDB — PSB — PCB — PC do B	231.208 186.088
PL — PASART —	182.219 162.351
PTB — PTN —	47.063 31.552
PT — PDS —	26.432 21.428
PH — PND —	17.567 12.311
PMN — PN —	9.790 6.997
PPB — PNR —	4.785 4.512
PJ — PRT —	3.879 3.677
PRP —	2.364
TOTAL	2.478.323

Consolidação	
Total de votos dos Partidos ou Coligações	2.478.323
Votos em branco	59.711
Subtotal (votos válidos)	2.538.034
Votos nulos	161.440
TOTAL (comparecimento)	2.699.474

b) Votação obtida pelos candidatos:

Nomes	Votação	Partido ou Coligação
ROBERTO SATURNINO BRAGA	1.059.614	PDT —
Jó Antonio de Rezende	"	"
RUBEM MEDINA	462.486	PFL — PS —
Sebastião Augusto de Souza Nery	"	"
JORGE CORDEIRO LEITE	231.208	PMDB —
Aloysio Maria Teixeira Filho	"	"
MARCELLO AUGUSTO DINIZ CERQUEIRA	188.088	PSB — PCB — PC do B
João Alves Saldanha	"	" "
ALVARO BASTOS VALLE	182.219	PL —
Herculano Leal Carneiro	"	"
AARÃO STEINBRUCH	162.351	PASART —
Armando Cabral de Castro	"	"
FERNANDO SOUZA RIBEIRO DE CARVALHO	47.063	PTB —
Francisco Luiz Cavalcanti da Cunha Horta	"	"
CARLOS EDUARDO CORTE IMPERIAL	31.552	PTN —
Sidney Pereira Domingues	"	"
WILSON RODRIGUES DE FARIAS	26.432	PT —
Mirian Limoeiro Cardoso	"	"
HEITOR BAPTISTA FURTADO	21.428	PDS —
Fleing Baptista Furtado	"	"

TOTAL de todos os candidatos (deve ser igual ao "Total de votos dos Partidos ou Coligações")

BIBLIOGRAFIA

BERGER, James O. *Statistical decision theory; foundations, concepts and methods.* New York, Springer-Verlag. 1980, 425 p. (Springer series in statistics).

BERNARDO, J. M. Monitoring the 1982 spanish socialist victory; a bayesian analysis. *Journal of the American Statistical Association.* Washington , *79* (387): 510-15, sept. 1984.

BOX, G.E.P.; TIAO G.C. *Bayesian inference in statistical analysis.* Reading, Mass., Addison-Wesley. 1973, 588 p. (Addison-Wesley series in behavioral science; quantitative methods).

DEGROOT, M. H. *Optimal statistical decision.* New York. McGraw-Hill. 1970, 489 p.

KULLBACK, S. *Information theory and statistics.* New York, J. Wiley. London, Chapman & Hall. 1959, 395 p. (Wiley publication in mathematical statistical).

LINDLEY, D. V.; SMITH, A. F. M. Bayes estimates for the linear model. *Journal of the Royal Statistical Society; series B: methodological.* London, *34* (1): 1-42, 1972.

PEREIRA, C. A. B.; VIANA, M. A. G. *Elementos de inferência bayesiana.* São Paulo, USP. 1982, 98 p. (Documento apresentado no V Simpósio Nacional de Probabilidade e Estatística).

SABBAG, A. *Classificação probabilística.* Dissertação-mestrado, UFRJ, 1987.

SMITH, A. F. M. A general bayesian linear model. *Journal of the Royal Statistical Society; series B: methodological.* London, *35* (1): 67-75, 1973.

RESUMO

Com intuito de prever o resultado das eleições de 1985, para a Prefeitura da Cidade do Rio de Janeiro, entrevistaram-se, ao todo, 1.000 eleitores em 20 diferentes *locais de votação*, criteriosamente escolhidos. O desenho da amostra, intimamente associado ao modelo estatístico para análise dos dados, é bastante diferente dos comumente utilizados. A abordagem de natureza bayesiana, que permite descrever probabilisticamente os parâmetros, gerou resultados muito satisfatórios.

ABSTRACT

A sample of 1000 electors from different pooling stations conveniently selected, was interviewed in order to predict the final result of the 1985 election for the Mayor of RJ. The sample desingning is rather different from the commonly used and strongly related to the proposed statistical model.

The Bayesian approach besides enabling one to describe probabilistically parameters of the model, did bring up very satisfactory results.

Os autores agradecem a todos os alunos e Professores do Departamento de Estatística da Universidade Federal do Rio de Janeiro pela colaboração empregada a este trabalho, bem como ao Laboratório de Estatística pelo suporte financeiro que proporcionou.

Menções especiais devem ser conferidas a Sonia Helt, pela informatização do quadro amostral, e a Mauricio Koki Matsutani, pela coordenação da pesquisa de opinião.

APLICAÇÃO DA AMOSTRAGEM NA COLETA DOS CENSOS DEMOGRÁFICOS NO BRASIL

Luiz Nery da Costa *

1 — INTRODUÇÃO

Nos últimos 50 anos, a população humana apresentou substancial aumento populacional. Durante o período de 1940 a 1980, o contingente demográfico mundial cresceu de 2,3 bilhões de habitantes; no Brasil, o crescimento foi de 77,8 milhões de pessoas.

A realização de Censos Demográficos (População e Habitação) de forma exaustiva ou completa tornou-se uma operação de custo muito elevado, devido aos grandes contingentes populacionais de muitos países.

Durante este período, a Teoria de Amostragem foi aperfeiçoada e o seu conhecimento atingiu maior número de estudiosos e países. Neste período foram editados os principais livros exclusivamente de amostragem, e pontificaram os estudos científicos da eficácia dessa técnica.

Devido às crescentes exigências de informações estatísticas para o planejamento, à necessidade de redução ou limitação no custo da realização de um censo e ainda à premência na obtenção de resultados com mais rapidez, a adoção da amostragem nas tarefas pertinentes a um censo vem sendo ampliada, apesar de já largamente adotada por diversos países.

Em 1940 os Estados Unidos utilizaram, pela primeira vez, a técnica de amostragem na coleta do Censo de População e Habitação; desde então, essa técnica vem sendo aperfeiçoada e utilizada cada vez mais nas diversas tarefas censitárias.

No Brasil empregou-se, pela primeira vez, a amostragem na coleta do Censo Demográfico de 1960, e posteriormente o seu emprego vem sendo aperfeiçoado e ampliado nas diversas fases do Censo Demográfico.

Atualmente, diversos países utilizam a amostragem na coleta do Censo Demográfico, como por exemplo: Canadá, Inglaterra, Argentina, Peru, além dos Estados Unidos e do Brasil.

Com o extraordinário desenvolvimento tecnológico na área da computação eletrônica e o grande incremento da capacidade de armazenagem de dados e de cálculo, os processos de estimação vêm se tornando cada vez mais sofisticados.

A literatura existente sobre a utilização da técnica de amostragem nos Censos é relati-

* Analista especializado da Fundação Instituto Brasileiro de Geografia e Estatística — IBGE.

vamente escassa e específica. Assim sendo, julguei ser interessante reunir algumas notas técnicas sobre o assunto para que os estatísticos e técnicos que estão iniciando os seus trabalhos nessa atividade possam fazê-los com mais facilidade, corrigindo os aspectos ora apresentados, com os quais não concordem, e sobretudo, aperfeiçoando-os.

Embora a utilização de técnicas de amostragem seja possível em quase todas as fases dos trabalhos censitários, neste estudo serão focalizadas apenas as aplicações na coleta e obtenção dos resultados definitivos do Censo.

2 — RAZÕES DO USO DA AMOSTRAGEM

No Brasil, até o Censo Demográfico de 1950, todas as informações investigadas foram coletadas de forma exaustiva ou completa, através do preenchimento de um único modelo de questionário para cada domicílio. No Censo Demográfico de 1960 utilizou-se, pela primeira vez, a técnica de amostragem na coleta das informações. Desde então, a amostragem vem sendo empregada nos trabalhos censitários cada vez em maior escala. Sem dúvida, o apoio técnico propiciado pela USAID ao IBGE contribuiu de modo expressivo para a aplicação da amostragem neste Censo.

A redução nos custos da coleta e de outras fases dos trabalhos censitários, a possibilidade de obtenção dos resultados com maior rapidez e precisão, pois a quantidade de dados a ser trabalhada é menor, constituem as principais razões do uso da amostragem na coleta do Censo (17).

É oportuno salientar que em outros países, no passado, alguns censos foram realizados por amostragem devido à impossibilidade de uma enumeração completa decorrente da falta de recursos, carência de pessoal, condições materiais deficientes, etc. (14). A situação contemporânea é diferente, pois esses fatores impeditivos não são predominantes, apenas procura-se otimizar o aproveitamento dos recursos disponíveis, dentro de uma concepção tecnológica mais avançada.

3 — PLANEJAMENTO DE AMOSTRAGEM

A adoção do uso da amostragem na coleta do Censo requer que sejam apreciados diversos pontos que influem na escolha da metodologia e nos resultados finais, dentre eles destacando-se:

- Definição dos objetivos do Censo;
- Plano de divulgação dos resultados;
- Uso de dois ou mais questionários;
- Critérios para inclusão dos quesitos nos questionários;
- Precisão desejada do Censo;
- Unidade básica de amostragem;
- Tamanho da amostra;
- Processo de seleção da amostra;
- Processo de estimação;
- Diretrizes para a construção do processo de estimação; e
- Cálculo e apresentação dos erros.

A seguir, são apresentadas considerações a respeito de cada um dos pontos assinalados anteriormente, onde a ordenação dessas considerações não necessita ser seguida tal como se apresenta.

3.1 — Definição dos Itens da Investigação

A formulação dos itens a serem investigados nos Censos visa a atender às necessidades nacionais e às recomendações e princípios apontados pelas Nações Unidas, a fim de possibilitar estudos comparativos entre os diversos censos realizados pelos países. A investigação censitária procura a manutenção de séries históricas, o estabelecimento de tendências e mudanças na composição de características, principalmente da população.

Atualmente, os objetivos do Censo visam a conhecer três grandes conjuntos de características: de pessoas, de famílias e de domicílios ou habitações. Inúmeros tópicos são investigados. Dentre eles destacam-se as características sobre: condições de presença, migração, alfabetização e escolaridade, fecundidade, mortalidade, ocupação das pessoas, emprego e rendimentos, estruturas etárias, composição das famílias, espécie de família, condições dos domicílios, etc.

A necessidade cada vez maior de informações demográficas e sócio-econômicas, tem acarretado uma sobrecarga na investigação censitária. No Brasil, em 1960, foram indagados 24 quesitos de pessoas, em 1970 a investigação cresceu para 32 perguntas e, finalmente, atingiu 57 itens em 1980 (15).

Não é aconselhável aumentar, na proporção que vem ocorrendo, o número de itens investigados, pois os erros cometidos na investigação censitária crescem substancialmente com a ampliação do número de quesitos formulados.

3.2 – Divulgação dos Resultados

O plano de divulgação dos resultados, entendido como sendo os resultados das tabulações finais contendo as estimativas e os dados da apuração do universo, é influenciado pelo nível de detalhamento geográfico segundo o qual os resultados são apresentados e pelo volume de tabelas cruzadas que contém. Implicitamente ligadas aos objetivos do Censo, a maioria das tabelas é concebida para apresentar resultados referentes às Unidades da Federação, ao Município e ao Distrito. Por outro lado, as tabelas elaboradas muitas vezes apresentam cruzamentos de diversos itens pesquisados, algumas delas com grandes detalhamentos.

A dificuldade de investigação de certo item, o nível de detalhamento geográfico e ainda cruzamentos previstos no plano tabular devem ser adequadamente ponderados na fase do planejamento da operação censitária, em função da qualidade e precisão desejadas dos resultados, principalmente quando obtidos a partir da amostra.

Algumas tabelas são elaboradas cruzando informações para subsidiarem estudos enfocados em âmbito nacional ou quando muito regional. Não há razão de concebê-las a nível de detalhamento menor, pois não possibilitarão decisões político-administrativas em áreas menores do que as Unidades da Federação.

Entretanto, muitas outras tabulações devem ser concebidas por município para auxiliarem no planejamento regional e muito poucas por distrito. Em tais casos, os resultados funcionam como indicadores gerais, possibilitando comparações, sem tabulações cruzadas detalhadas.

Quando for julgado necessário um detalhamento profundo dos dados, há que se pensar na sua conveniência, pois isto é difícil de ser conseguido com o uso da amostragem.

3.3 – Uso de Dois ou Mais Questionários

A aplicação da amostragem na coleta do Censo quase sempre é realizada utilizando-se dois modelos de questionário. Um questionário contendo poucos itens da investigação censitária é aplicado às unidades que não foram selecionadas para a amostra; outro contendo esses mesmos itens e todos os demais é aplicado apenas às unidades selecionadas para a amostra.

Dessa forma, para uns poucos quesitos a investigação é realizada de forma exaustiva ou completa para todas as unidades recenseadas, enquanto que para os demais quesitos a investigação é realizada por amostragem, geralmente através de uma única amostra, ou então de duas amostras. Neste caso, com o emprego de duas amostras na coleta, seriam utilizados três questionários. O primeiro modelo de questionário contendo poucos itens seria aplicado às unidades não selecionadas para as amostras; esses poucos itens também figurariam nos segundo e terceiro modelos de questionários, os quais conteriam partes inteiras ou parcialmente distintas dos demais quesitos da investigação censitária. O segundo e o terceiro questionários seriam aplicados a amostras distintas.

Quando o elenco de quesitos da investigação censitária for grande ou envolver muita complexidade, sendo indispensável tal investigação, o uso de dois questionários aplicados a duas amostras distintas contribuiria para a redução do custo da investigação e para diminuir os erros alheios à amostragem, os quais são mais freqüentes e importantes quando o número de quesitos do questionário é grande.

3.4 – Critérios para Inclusão dos Quesitos nos Questionários

A realização de um Censo Demográfico constitui uma operação de custo elevado, complexa e, no Brasil, a coleta demorou cerca de 4 (quatro) meses em 1980. Além do mais, envolve grande contingente de pes-

soas na sua realização. Em decorrência dessa complexidade, vários fatores afetam o resultado final do Censo; alguns deles podem ser atenuados ou controlados tomando-se algumas decisões no planejamento dos questionários. Assim, após a definição dos temas ou itens a serem pesquisados no Censo, o estabelecimento do nível de detalhamento geográfico dos resultados e do grau de desagregação das características investigadas, necessários a atender aos objetivos do Censo, devem, então, ser definidos os critérios de inclusão dos quesitos nos diferentes questionários a serem utilizados na coleta do Censo.

Por ocasião do planejamento dos Censos Demográficos no Brasil de 1960 a 1980, o assunto foi devidamente apreciado e, levando-se em consideração a experiência do "Bureau of the Census" (18 e 19), pode-se delinear regras gerais para a escolha dos quesitos da enumeração completa e para a investigação por amostragem.

As características que devem figurar obrigatoriamente em todos os questionários do Censo, isto é, na enumeração completa, constituem, portanto, os itens básicos e devem atender algumas condições, principalmente:

a) variáveis fundamentais — idade, sexo, condição de presença, situação geográfica do domicílio, caracterização da família;

b) dados destinados a atender dispositivos legais — em alguns países a declaração de idade e sexo serve a este propósito;

c) características que, em função do objetivo definido, se destinam a apresentar resultados para subdivisões geográficas administrativas de pequena população, distritos, subdistritos, etc.;

d) variáveis indispensáveis à montagem do processo de expansão da amostra e para controlar possíveis tendenciosidades passíveis de ocorrerem na coleta. No Censo de 1980, além das variáveis fundamentais, as variáveis condição de presença no domicílio e tamanho do domicílio (número de pessoas recenseadas) foram utilizadas para esse fim; e

e) variáveis consideradas importantes e que representam proporções rarefeitas da população.

Estas condições, de um modo geral, são válidas nos países que utilizam a amostragem na coleta do Censo. Varia, entretanto, a importância atribuída às variáveis. Por exemplo, no Censo de População de 1976 do Canadá, as variáveis estado civil, língua falada pela mãe, etc., foram consideradas para a investigação completa. Já no Censo de População de 1960 dos Estados Unidos, as variáveis cor das pessoas e condição de domicílio, próprio ou alugado, etc., foram consideradas básicas.

Em decorrência dos objetivos do Censo, das dificuldades da coleta, da necessidade de redução dos custos, dos planos tabulares, a seleção dos quesitos a serem investigados por amostragem deve considerar:

a) variáveis difíceis de serem apuradas, especialmente quanto à codificação e crítica.

No Censo Brasileiro de 1980, entre as características sobre migração, o quesito "... nome do Município e a sigla da Unidade da Federação ou nome do país em que morava antes" e o item "Município em que trabalha ou estuda" exigem a aplicação de um código de 7 dígitos;

b) características sujeitas a grande variabilidades de respostas. A investigação das características sobre ocupação, rendimento e nível de instrução, apresentou no Censo Brasileiro de 1970 os maiores índices de inconsistência (11). Estas variabilidades, alheias ao processo de amostragem, apresentam erros bastante superiores àqueles decorrentes da amostragem; e

c) características de difícil investigação. Os resultados do Censo de 1980 sobre a mortalidade das pessoas moradoras no domicílio que faleceram nos últimos 12 (doze) meses foram bastante imprecisos. Embora aparentemente o falecimento de pessoas do domicílio seja um fato marcante na vida das pessoas, os resultados para esse quesito foram insatisfatórios, revelando, possivelmente, dificuldades na sua investigação.

A escolha dos temas a serem selecionados para a investigação por amostragem deve considerar: a dificuldade da codificação; a complexidade da crítica; a variabilidade das respostas, principalmente em função dos erros alheios à amostragem; e a extensão da seqüência dos quesitos pertinentes à caracterização do assunto estudado, como é o caso da Migração, Fecundidade,

Ocupação, Rendimento, etc. Os quesitos portadores de uma dessas restrições devem ser investigados por amostragem, exceto quando, por algum motivo, dados da investigação completa sejam indispensáveis ao processo de expansão, como é o caso da investigação da idade.

3.5 — Precisão do Censo

Os levantamentos por amostragem convivem com os erros alheios à amostragem e os decorrentes do uso dessa técnica.

Os erros alheios à amostragem, como o próprio nome indica, independem das técnicas de amostragem utilizadas, mas são oriundos de falhas de planejamento, coleta defeituosa, treinamento insatisfatório, codificação falha, preparação e processamento de dados deficientes,etc. Já fazem parte do levantamento censitário, seja ele realizado com a amostragem na coleta ou sem o uso dela. Entretanto, é possível atenuar alguns desses erros mediante processos de investigação mais detalhada, o uso de crítica mais elaborada, indagações mais cuidadosas o que é possível com a utilização da amostragem na coleta. As Pesquisas de Avaliação dos Censos Brasileiros de 1970 e 1980 propiciaram estimativas preliminares das taxas de omissão da população da ordem de 7% e 4%, respectivamente.

Enquanto os erros alheios à amostragem são quase sempre inponderáveis, os erros de amostragem são previamente fixados, previstos ou esperados, em função das condições técnicas do trabalho, dos recursos financeiros e dos procedimentos de amostragem adotados.

As especificações técnicas implicam em definir o grau de precisão e o nível de confiança dos resultados em termos probabilísticos, estabelecidos no planejamento, propiciando o conhecimento da dimensão dos erros antes do início da operação. Os procedimentos de amostragem implicam na definição do modelo da amostra adotado e nos processos de estimação, fatores que determinam os erros de amostragem. Portanto, não é demais esclarecer que os erros de amostragem podem ser reduzidos, rigorosamente previstos e controlados, o que não ocorre em relação às tendenciosidades, que podem ser deliberadas e imprevisíveis com erros maiores. Geralmente não podem ser medidas objetivamente e algumas vezes são desconhecidas as suas causas e ocorrências, aqui no Brasil, não obstante a existência de alguns estudos já realizados (15 e 13).

3.5.1 — Ordem de grandeza do erro de amostragem

A estimação de qualquer parâmetro Θ (total, média, proporção, etc.) é realizada por um estimador $\hat{\Theta}$, com uma margem de flutuação, dentro de intervalo de confiança, obtido a partir do erro de amostragem. Essas medidas de erro podem ser absolutas ou relativas, e são definidas por:

$$E = Z \sqrt{V^2\{\hat{\Theta}\}} \quad \text{ou} \quad E_r = Z \sqrt{V_r^2\{\hat{\Theta}\}} \quad [1]$$

onde:

E – Máximo erro absoluto previsto;

E_r – Máximo erro relativo previsto;

Z – Múltiplo do desvio padrão, inerente ao coeficiente de confiança;

$V^2\{\hat{\Theta}\}$ – Variância do estimador $\hat{\Theta}$ do parâmetro Θ; e

$V_r^2\{\hat{\Theta}\}$ – Variância relativa do estimador $\hat{\Theta}$ do parâmetro Θ.

Os valores mais usuais de Z, provenientes da distribuição Normal (0;1), são:

TABELA 1
ABSCISSAS SELECIONADAS DA DISTRIBUIÇÃO NORMAL (0;1) E RESPECTIVAS ÁREAS

VALORES DE Z	ÁREAS DA NORMAL (%)	COEFICIENTE DE CONFIANÇA
1,00	68,26	0,68
1,65	90,00	0,90
1,96	95,00	0,95

No caso particular de $Z = 1{,}96$, valor mais usual, o correspondente coeficiente de confiança é $(1-\alpha) = 0{,}95$. Isto significa que, se amostras de mesmo tamanho e desenho são selecionadas cem vezes, resultam 100 intervalos de confiança baseados nos estimados $\{\hat{\Theta}\}$, dos quais 95% devem conter o valor do parâmetro Θ.

Substituindo-se em [1] o valor de $V^2\{\hat{\Theta}\}$ pela fórmula apropriada para o desenho de

amostra correspondente, chega-se à expressão que possibilita o cálculo do tamanho n da amostra. O tamanho da amostra depende, principalmente, da variância do estimador do parâmetro Θ, do tamanho da população (até certo ponto), do coeficiente de confiança pré-fixado para atender o erro desejado e do custo. Por outro lado, a expressão de $V^2\{\hat{\Theta}\}$ varia conforme o modelo de amostragem seguido e o processo de estimação.

No Censo são pesquisadas inúmeras variáveis e são diversos os fatores que condicionam o tamanho da amostra para diferentes universos. Assim sendo, para dar uma idéia imediata da ordem de grandeza do erro de amostragem esperado, utilizou-se o esquema de amostragem aleatória simples, pois as aplicações da amostragem na coleta do Censo são aproximadamente equivalentes a esse modelo em termos de precisão.

A seguir, são apresentadas as ordens de grandeza dos erros absolutos e relativos de amostragem esperados, concernentes a diversos tamanhos de universo e de amostra [1]; no presente caso utilizamos a fração de amostragem $f = \frac{n}{N}$ como alternativas de n.

A apreciação dos valores inseridos nas Tabelas 2 e 3 possibilita as seguintes observações, entre outras:

— fixados o tamanho da estimativa e o tamanho da amostra, os erros absolutos e relativos aumentam à medida que o tamanho do universo cresce, até um certo valor, depois o erro permanece inalterado, embora o universo aumente;

— fixado o tamanho da estimativa, os erros absolutos e relativos decrescem com intensidade menor, à medida que o tamanho da amostra aumenta;

— as estimativas de totais que representam proporções pequenas do universo contêm erros elevados; e

— os erros referentes às frações 1/4 e 1/5 são muito parecidos.

As diversas alternativas de tamanho de amostras, com as diferentes frações de amostragem, foram apreciadas por ocasião do planejamento do Censo Brasileiro 1980. Maiores detalhes sobre o assunto podem ser obtidos em (17 e 19).

3.6 — Unidade Básica de Amostragem

Os Censos Demográficos já realizados por alguns países mostram que têm sido utilizados modelos de amostragem sistemática equivalentes à amostragem simples e amostragem de áreas. A seguir, são relatadas algumas dessas aplicações.

Os Estados Unidos da América empregaram, pela primeira vez, a amostragem na coleta de dados do Censo de População de 1940; desde então os Censos Americanos de População e Habitação passaram a utilizar cada vez mais a amostragem nos traba-

TABELA 2

ERROS ABSOLUTOS DE AMOSTRAGEM NA ESTIMAÇÃO DE ALGUNS TOTAIS POR DIFERENTES TAMANHOS DE AMOSTRA, SEGUNDO ALGUNS TAMANHOS DE UNIVERSOS

TAMANHO DO UNIVERSO	TAMANHOS DAS ESTIMATIVAS PARA DIVERSAS FRAÇÕES DE AMOSTRAGEM															
	50				100				250				500			
	1/10	1/8	1/5	1/4	1/10	1/8	1/5	1/4	1/10	1/8	1/5	1/4	1/10	1/8	1/5	1/4
500	20	18	13	12	27	24	18	16	34	30	22	19	–	–	–	–
1000	21	18	14	12	28	25	19	17	40	38	27	23	47	42	32	27
5000	21	19	14	12	30	26	20	17	46	41	31	27	64	56	42	37
10000	21	19	14	12	30	26	20	17	47	42	32	27	66	58	44	38
50000	21	19	14	12	30	26	20	17	47	42	32	27	67	59	45	39
100000	21	19	14	12	30	26	20	17	47	42	32	27	67	59	45	39

Nota - Cálculos obtidos segundo o esquema de amostragem simples com coeficiente de confiança de 68%.

TABELA 3

ERROS RELATIVOS DE AMOSTRAGEM NA ESTIMAÇÃO DE ALGUNS TOTAIS, POR DIFERENTES TAMANHOS DE AMOSTRA, SEGUNDO ALGUNS TAMANHOS DE UNIVERSOS

TAMANHO DO UNIVERSO	TAMANHOS DAS ESTIMATIVAS PARA DIVERSAS FRAÇÕES DE AMOSTRAGEM															
	50				100				250				500			
	1/10	1/8	1/5	1/4	1/10	1/8	1/5	1/4	1/10	1/8	1/5	1/4	1/10	1/8	1/5	1/4
500	40	36	26	24	27	24	18	16	14	12	9	8	–	–	–	–
1000	42	36	28	24	28	25	19	17	16	15	11	9	9	8	6	5
5000	42	38	28	24	30	26	20	17	18	16	12	11	13	11	8	7
10000	42	38	28	24	30	26	20	17	19	17	13	11	13	12	9	8
50000	42	38	28	24	30	26	20	17	19	17	13	11	13	12	9	8
100000	42	38	28	24	30	26	20	17	19	17	13	11	13	12	9	8

Nota - Cálculos obtidos supondo o esquema de amostragem simples com o coeficiente de confiança de 68%.

lhos censitários. No Censo de 1940, a pessoa constituiu a unidade de amostragem: uma em cada cinco pessoas do domicílio foi incluída na amostra se ele possuísse menos de 5 pessoas, duas pessoas foram selecionadas se o domicílio possuísse de 6 a 10 pessoas.

Novamente, no Censo Demográfico de 1950, os Estados Unidos utilizaram a pessoa como unidade de amostragem (14). Em 1960, o domicílio foi empregado como unidade de amostragem, ao invés da pessoa. Todas as pessoas do domicílio selecionado para a amostra também foram incluídas nela. A razão básica dessa mudança de unidade de amostragem foi devido ao fato de que o domicílio possibilitou melhor obtenção de estatísticas sobre as características do domicílio e da família; além disso, facilitou a cooperação dos informantes através da auto-enumeração das pessoas do domicílio. A partir deste Censo, o domicílio passou a ser a unidade de amostragem (18).

O Canadá, desde o Censo de População de 1971 até o último, em 1981, tem aplicado a amostragem como principal método de coleta censitário. A unidade básica de amostragem adotada foi o domicílio (9 e 2).

A Argentina aplicou, pela primeira vez, a amostragem na coleta do Censo Demográfico de 1980, através de uma amostra de áreas. A unidade básica de amostragem adotada foi o segmento censitário (conglomerado de 40 domicílios contíguos, em média), dentro de uma mesma área.

Em todos os domicílios existentes nos segmentos censitários, selecionados aleatoriamente para a amostra de áreas, foi aplicado o questionário completo e mais detalhado, enquanto nos domicílios localizados nos segmentos censitários não selecionados para a amostra de áreas foi aplicado o questionário simplificado.

A garantia de uma seleção aleatória, o fato de cada recenseador trabalhar com um único modelo de questionário, com possível melhoria na qualidade dos dados, e a simplificação do programa de treinamento foram os principais motivos de adoção do segmento censitário como unidade básica de amostragem (10).

O Brasil, nos seus três Censos Demográficos realizados com aplicação de amostragem na coleta (1960, 1970 e 1980), utilizou o domicílio como unidade de amostragem (em domicílios particulares).

A escolha da unidade de amostragem merece cuidadosa apreciação sob diversos aspectos, como as condições de comunicações, o nível de educação, cultura e conscientização das pessoas, os recursos materiais e condições técnicas, e os objetivos do Censo de cada país.

Quando a investigação censitária objetiva conhecer as características das pessoas, dos domicílios e das famílias, tudo se torna

mais simples de se obter adotando-se o domicílio como unidade de amostragem, e incluindo na amostra todas as pessoas nele recenseadas. Aliás, esta foi a principal razão pela qual os Estados Unidos passaram a utilizar o domicílio como unidade de amostragem, a partir do Censo de 1960. Destaque-se, ainda, que a correlação intraclasse da proporção do sexo das pessoas no domicílio é negativa, o que torna mais desfavorável o uso da pessoa como unidade de amostragem.

Nos países em que a população possui um bom nível de educação, cultura, conscientização e tradição para a enumeração através dos Correios, sem dúvida o domicílio é adotado como unidade básica de amostragem. Para o sucesso dessa operação censitária, há que se dispor de cadastro e de listagens bem organizados e controlados sobre as residências existentes, para a maior parte do País.

Com o auxílio dos Correios, é possível fazer chegar previamente aos domicílios, com maior rapidez, os questionários a ele destinados, de modo a serem preenchidos no dia do Censo ou em data mais próxima, com devolução imediata. Exemplo: 90% da coleta do Censo de População de 1981, do Canadá, foi realizada em um mês (16). Além disso, cada responsável pelo domicílio preenche o questionário fazendo a auto-enumeração das pessoas e do domicílio, reduzindo possíveis erros introduzidos pelo recenseador, conseqüentemente obtendo melhor qualidade dos dados.

Este moderno processo de auto-enumeração, com as suas vantagens e desvantagens, só pode ser adotado em países que disponham de condições para tal, como é o caso dos Estados Unidos, Canadá, etc. Nos países menos desenvolvidos, com menores recursos técnicos e materiais, é uma temeridade a auto-enumeração, com distribuição dos questionários, previamente, pelos Correios. É a situação da maioria dos países.

A utilização do domicílio ou do conglomerado de domicílios (setor censitário ou fração deste), como unidade de amostragem, além dos aspectos técnicos envolvidos, depende das condições e recursos de cada país.

Por ocasião do planejamento do Censo Demográfico do Brasil, referente ao ano de 1980, foi examinada a possibilidade da adoção do setor censitário como unidade básica de amostragem ao invés do domicílio. Os motivos favoráveis e desfavoráveis estão registrados em (15). Apesar disso, não é demais acrescentar alguns aspectos adicionais. A adoção de uma unidade conglomerada visa, principalmente, a:

a) maior facilidade operacional, com aplicação pelo recenseador de um único modelo de questionário em cada unidade;

b) maior rapidez na coleta, talvez em um ou poucos dias; e

c) melhor qualidade das informações coletadas, se os recenseadores das unidades conglomeradas selecionadas na amostra forem mais capacitados

Além disso, a precisão das informações coletadas no Censo depende do informante, do recenseador e das condições operacionais. A aplicação de um único modelo de questionário pelo recenseador realmente é uma facilidade operacional, mas a precisão das respostas não depende apenas disto. Os resultados da Pesquisa de Avaliação do Censo Demográfico de 1970 (12) mostraram que a qualidade das informações praticamente foi a mesma para os dois modelos de questionário, mas a precisão variava mais fortemente quando os informantes eram diferentes nos dois levantamentos ou, então, quando o recenseador ou o pesquisador utilizou processos diferentes, por exemplo, na forma de registro da idade.

A coleta é feita com maior rapidez empregando-se um contingente maior de recenseadores. Naturalmente, subdividindo-se os atuais setores censitários em áreas menores e efetuando-se a coleta, simultaneamente em todos eles, na mesma época, a coleta seria mais rápida, independentemente do tipo de unidade de amostragem adotado, mas exigiria maior número de recenseadores, praticamente uns 500 mil no Brasil.

Se, por um lado, a coleta pode ser agilizada e as informações coletadas guardarem mais precisão em relação aos aspectos do dia do Censo, com o aumento do número de recenseadores, a dificuldade de treinamento de todo o pessoal envolvido na coleta au-

TABELA 4
TAMANHO DA AMOSTRA SIMPLES NECESSÁRIO À ESTIMAÇÃO DA PROPORÇÃO P
(para $\delta = 1,96 = 2$)

TAMANHO DA POPULAÇÃO (N)	TAMANHO DA AMOSTRA (n)											
	P = 0,001				P = 0,01				P = 0,05			
	E_r = 0,03	E_r = 0,05	E_r = 0,10	E_r = 0,20	E_r = 0,01	E_r = 0,05	E_r = 0,10	E_r = 0,20	E_r = 0,01	E_r = 0,05	E_r = 0,10	E_r = 0,20
500	500	500	499	498	500	499	494	416	500	492	469	396
1 000	1 000	999	998	990	1 000	994	975	908	999	968	884	655
2 000	1 999	1 998	1 990	1 960	1 999	1 975	1 904	1 664	1 995	1 877	1 583	974
2 500	2 499	2 496	2 484	2 439	2 498	2 462	2 352	1 996	2 492	2 311	1 881	1 080
3 000	2 998	2 994	2 978	2 913	2 998	2 945	2 789	2 302	2 988	2 730	2 151	1 163
4 000	3 996	3 990	3 960	3 846	3 996	3 902	3 633	2 849	3 979	3 535	2 621	1 288
5 000	4 994	4 984	4 938	4 762	4 993	4 841	4 439	3 322	4 973	4 294	3 016	1 376
7 500	7 487	7 465	7 362	6 976	7 486	7 160	6 305	4 267	7 422	6 012	3 775	1 516
10 000	9 978	9 938	9 756	9 090	9 975	9 406	7 984	4 975	9 870	7 525	4 318	1 597
15 000	14 950	14 860	14 458	13 042	24 943	13 687	10 879	5 964	14 710	10 044	5 044	1 686
20 000	19 910	19 752	19 047	16 664	19 899	17 758	13 289	6 622	19 487	12 063	5 507	1 735
25 000	24 860	24 615	23 528	19 996	24 843	21 592	15 325	7 092	24 204	13 718	5 828	1 766
50 000	49 443	48 483	44 440	33 330	49 377	38 004	22 098	8 264	46 914	18 905	6 597	1 830
75 000	73 754	71 639	63 147	42 839	73 606	50 856	25 916	8 746	68 345	21 640	6 901	1 853
100 000	97 797	94 112	79 984	49 975	97 537	61 300	28 367	9 008	88 372	23 313	7 063	1 865
250 000	236 673	216 187	153 787	71 378	235 154	96 964	34 185	9 523	188 119	27 104	7 376	1 886
500 000	449 392	380 861	222 099	83 264	443 946	120 292	36 694	9 708	301 587	28 658	7 486	1 893
1 000 000	816 176	615 147	285 510	90 826	798 387	136 740	38 092	9 803	431 818	29 503	7 542	1 896
∞	4 440 000	1 598 400	399 600	99 900	3 960 000	158 400	39 600	9 900	760 000	30 400	7 600	1 900

TAMANHO DA POPULAÇÃO (N)	TAMANHO DA AMOSTRA (n)											
	P = 0,10				P = 0,20				P = 0,30			
	E_r = 0,01	E_r = 0,05	E_r = 0,10	E_r = 0,20	E_r = 0,01	E_r = 0,05	E_r = 0,10	E_r = 0,20	E_r = 0,01	E_r = 0,05	E_r = 0,10	E_r = 0,20
500	499	484	439	321	498	464	381	222	497	441	326	159
1 000	997	935	783	474	994	865	615	286	989	789	483	189
2 000	1 990	1 756	1 285	621	1 975	1 524	889	333	1 958	1 302	636	209
2 500	2 484	2 130	1 475	662	2 462	1 798	976	345	2 436	1 496	680	213
3 000	2 975	2 483	1 636	692	2 945	2 043	1 043	353	2 907	1 663	712	216
4 000	3 956	3 130	1 895	735	3 902	2 462	1 143	363	3 836	1 931	758	220
5 000	4 932	3 711	2 094	763	4 848	2 807	1 212	370	4 746	2 137	787	223
7 500	7 347	4 932	2 432	804	7 164	3 453	1 319	380	6 942	2 493	830	226
10 000	9 730	5 901	2 647	826	9 412	3 902	1 379	385	9 032	2 718	854	228
15 000	14 400	7 347	2 903	849	13 714	4 466	1 446	390	12 923	2 989	879	230
20 000	18 947	8 372	3 051	861	17 778	4 848	1 481	392	16 470	3 146	892	231
25 000	23 377	9 137	3 147	869	21 622	5 096	1 504	394	19 718	3 248	900	231
50 000	43 902	11 180	3 358	884	38 095	5 674	1 550	397	32 558	3 474	916	232
75 000	62 069	12 081	3 435	889	51 064	5 897	1 567	398	41 584	3 556	922	233
100 000	78 260	12 587	3 474	892	61 538	6 015	1 575	398	48 276	3 599	925	233
250 000	147 541	13 616	3 549	897	97 561	6 240	1 590	399	67 961	3 678	930	233
500 000	209 302	13 997	3 574	898	121 212	6 319	1 595	400	78 652	3 706	932	233
1 000 000	264 706	14 195	3 587	899	137 931	6 359	1 597	400	85 366	3 719	932	233
∞	360 000	14 400	3 600	900	160 000	6 400	1 600	400	93 333	3 733	933	233

Obs.: Tabela elaborada pelo Setor Especial de Amostragem em maio de 1970.

NOTA – *Utilizou-se a seguinte fórmula:*

$$N = \frac{\delta^2 N(1-P)}{\delta^2(1-P) + E_r^2 P(N-1)}$$

onde:

- N = tamanho da população
- n = tamanho da amostra
- δ = múltiplo do desvio padrão, inerentes ao coeficiente de confiança $(1 - \alpha)$
- E_r = erro relativo de amostragem
- P = proporção de atributo x

menta substancialmente. Esse ponto merece uma apreciação cuidadosa para o próximo Censo.

A precisão do Censo, não só quanto à qualidade das respostas obtidas, mas também quanto à cobertura da coleta, pode e deve ser aumentada, principalmente, com a melhoria do recrutamento dos recenseadores, livre de influências políticas; treinamento mais adequado, controle e acompanhamento da coleta com maior rigidez,etc.

No Censo Experimental Demográfico Brasileiro de 1978, foi empregada uma equipe de recenseadores com bom nível de escolaridade (segundo grau), com bom nível de treinamento e acompanhamento constante. Apesar destes cuidados, o nível de precisão das respostas obtidas e de cobertura do Censo não foi substancialmente melhor do que os resultados do Censo de 1970, para a zona urbana do Estado de São Paulo (7).

Por fim, não é demais registrar que, usando-se o conglomerado de domicílios como unidade de amostragem, a amostra deve ser aumentada para alcançar a mesma eficiência de uma amostra de domicílios, pois o grau de homogeneidade dos setores censitários é muito forte, especialmente em relação a certas características. Os resultados da Pesquisa da Avaliação de 1980 (12) indicam valores superiores a 0,6 para o consumo de energia elétrica, lenha, gás liquefeito, etc. Através da Pesquisa de Avaliação de 1970, foi estimada a correlação intraclasse em 0,3 a 0,4 para a renda.

3.7 – Tamanho da Amostra

Teoricamente, o tamanho da amostra é obtido a partir da expressão [1], indicada em 3.5.1, explicitando-se o valor de n, contido em $V^2\{\hat{\Theta}\}$.

Como as características levantadas em um censo são numerosas e para muitos universos, com distribuições distintas, procura-se avaliar a dimensão do erro de amostragem cometido na estimação de proporções p, ou de totais reduzidos. As tabelas inseridas em 3.5.1 possibilitam uma idéia razoável sobre os erros de amostragem e a tabela no anexo 1 propicia a ordem de grandeza do tamanho da amostra n, para estimar diversos níveis de proporções p com diversos tamanhos de universos, para certos níveis de precisão fixados.

A partir de considerações dessa natureza, são avaliadas as respectivas frações de amostragem $f = \frac{n}{N}$ e as conseqüências para as diversas alternativas da fração de amostragem.

Os Estados Unidos utilizaram a fração de amostragem de 1/20 no Censo de População de 1940, posteriormente as frações de amostragens básicas foram 1/5, 1/4, 1/5, respectivamente nos Censos de 1950, 1960 e 1970. Em 1980, nas áreas referentes à menor divisão civil com menos de 2.500 pessoas foi adotada a fração de amostragem 1/2 e 1/6 no restante do País.

O Canadá adotou a fração de amostragem de 1/3 nos Censos de População referentes aos anos de 1971 e 1976. A fração de amostragem foi reduzida para 1/5 no Censo de 1981, principalmente para diminuir o custo, e serviu de argumento também a menor carga de investigação junto ao público (9).

A Argentina, no Censo de População de 1980, dividiu o País em duas áreas distintas de levantamentos. Na área compreendida pela Capital Federal, pelas Províncias de Buenos Aires, Córdoba, Entre Rios e Santa Fé, e mais nove cidades com mais de 100 mil habitantes, foram aplicadas duas frações de amostragem, 1/10 na Região Metropolitana de Buenos Aires e 1/5 no restante dessa área. No restante do País, não se aplicou a amostragem na coleta (10).

O Brasil empregou a fração de amostragem de 1/4 nos Censos de 1960, 1970 e 1980. No planejamento do Censo Demográfico de 1980, foram consideradas várias alternativas para aplicação de frações de amostragem menores e esquemas de levantamento dos dados diferenciados, com o intuito de reduzir os custos de operação censitária. Maiores detalhes estão em (17 e 15).

É oportuno registrar a aplicação de amostragem na realização do Censo Escolar de 1977, pelo Governo do Estado do Espírito Santo. Este levantamento constituiu, na realidade, um Censo Demográfico, onde foram investigadas as características de escolaridade, econômicas, sociais, demográficas, etc. Neste Censo foram utilizadas duas

frações de amostragem, 1/10 aplicada nos municípios com mais de 5.000 pessoas e 1/5 nos demais municípios. Nos domicílios selecionados para a amostra, foram preenchidos o questionário completo e nos demais o boletim simplificado (8).

3.7.1 — Reflexões para os futuros Censos

Viu-se, anteriormente, que as frações de amostragem já aplicadas nos diversos Censos variam de 1/20 a 1/3, com tendências mais atuais de redução a fim de diminuir os custos da operação censitária.

Por outro lado, no Brasil tem havido uma tendência de aumentar o número de quesitos na investigação por amostragem, sobrecarregando o tamanho do questionário, com graves conseqüências na precisão das respostas coletadas e acarretando o surgimento de diversos erros alheios à amostragem. Além disso, a população brasileira continuará a crescer, devendo atingir em 1990, 150 milhões de habitantes com a existência de, aproximadamente, 30 a 35 milhões de domicílios. Neste ano, provavelmente, o número de municípios com menos de 5.000 habitantes deverá ficar em torno de 600 a 700, contribuindo com cerca de 2% da população total. Por outro lado, os municípios com mais de 100 mil habitantes deverão atingir 200, abrigando cerca de 45% da população brasileira, enquanto 1/3 da população estará concentrada nas regiões metropolitanas.

Os resultados das distribuições de diversas variáveis provenientes da amostra nos municípios pequenos (com menos de 5.000 habitantes) são estimados com erros de amostragem mais acentuados, em relação a cada classe ou categoria, conforme é esperado e se depreende das tabelas indicadas em 3.5.1.

Considerando que nos municípios pequenos há carência de recursos, conseqüentemente carência de informações mais detalhadas para o planejamento local, e que os erros relativos de amostragem das estimativas para esses municípios são geralmente maiores do que 10% e, ainda, que o custo da coleta nesses municípios representa uma parcela reduzida do custo total, então uma idéia razoável é a de não se aplicar a amostragem na coleta das informações desses municípios.

Atualmente, e nos próximos anos, haverá carência de recursos financeiros. A redução nos gastos, onde se puder economizar, constitui um dever social, pois o Brasil precisa de grande soma de dinheiro para implementar as reformas sociais e investir nos seus programas de desenvolvimento.

Considerando essa necessidade, dever-se-ia procurar a redução nos custos de todas as operações censitárias. A diminuição no tamanho da amostra do Censo Demográfico contribui para esse objetivo. Considerando, ainda, o grande elenco de itens da investigação censitária, as três últimas modalidades de esquemas de levantamento ou variações destes indicadas em (17 e 15), acrescido do enfoque dado aos pequenos municípios, são plenamente válidas e atuais.

Embora o assunto tenha sido apreciado com certa rapidez no planejamento do Censo Brasileiro de 1980, é necessário avaliar se os argumentos utilizados naquela ocasião, persistem hoje, bem como se os programas de trabalho que seriam desenvolvidos em relação às áreas metropolitanas e microáreas foram efetivados. No Censo Brasileiro de 1980,ocorreram várias tendenciosidades na seleção da amostra e na cobertura, não obstante o processo já viesse sendo repetido por três vezes seguidas.

3.8 — Processo de Seleção da Amostra

A seleção das unidades para serem investigadas por amostragem constitui uma importante fase do Censo, pois durante a mesma podem ocorrer sérios inconvenientes nas informações coletadas, conduzindo a diversos tipos de tendenciosidades.

O princípio fundamental da amostragem é o de ter as suas unidades selecionadas aleatoriamente. Entretanto, numa operação complexa como a coleta do Censo, nem sempre é possível garantir a aleatoriedade através de regras de seleção simples e compreensíveis para serem aplicadas pelos recenseadores.

Quando a unidade de amostragem é a pessoa, as regras de seleção tornam-se mais complicadas de aplicação pelos recenseadores. Porém, quando a unidade de amostragem corresponde à unidade de área (setor censitário, segmento, etc.), a seleção

aleatória é facilmente conseguida previamente, no próprio escritório, empregando-se pessoal mais qualificado.

Utilizando-se o domicílio como unidade de amostragem, e sendo realizada a seleção pelos recenseadores, durante a coleta, as regras de seleção devem ser simples, de fácil aplicação. Por esse motivo, essas regras são processos sistemáticos de seleção, estabelecidas de forma a quebrar ou interromper possíveis seqüências cíclicas do universo, de modo a obter bastante variação nas séries sistemáticas de seleção e a evitar o surgimento de tendenciosidades cometidas durante toda a coleta.

No Censo de População e Habitação dos Estados Unidos de 1960, foi adotado um processo de seleção sistemática, em que os domicílios foram selecionados através de linhas de amostra. O recenseador, ao percorrer sua área de trabalho, assinalava cada linha das listagens de domicílios estabelecida na primeira série de visitas com as letras A, B, C ou D, seqüencialmente. A cada recenseador foi atribuída, aleatoriamente, uma das quatro letras. Os domicílios assinalados com letra atribuída a ele foram selecionados para a amostra (18).

No Censo de 1970, os Estados Unidos empregaram o método sistemático de seleção nas duas áreas em que o País foi dividido. Na área de coleta através dos Correios, a seleção foi feita em computador antes do dia do Censo. Na área restante, a seleção foi feita através de linhas de amostra, sendo indicado para cada recenseador um início aleatório, como ponto de partida da seleção sistemática, de 1 em cada 5 domicílios (19).

O Canadá também utilizou um processo sistemático na seleção dos domicílios para a amostra nos Censos de População e Habitação de 1971 e 1976 (2).

A Argentina, no Censo de População de 1980, utilizou o método de seleção aleatória das áreas de amostra, com a seleção efetivada antes da realização do Censo de modo a garantir a aleatoridade da amostra. Todos os domicílios pertencentes às áreas de amostragem selecionadas foram incluídos na amostra (10).

O Brasil também utilizou um processo sistemático de seleção dos domicílios e grupos conviventes em domicílios coletivos, nos Censos Demográficos de 1960, 1970 e 1980. O processo adotado mantém algumas semelhanças com o método usado em 1960 pelos Estados Unidos.

De fato, as linhas de amostra realçadas, com intervalos regulares de quatro linhas, nos dois modelos de folhas de coleta, se assemelham ao processo citado anteriormente; as unidades registradas nessas linhas eram selecionadas para a amostra.

Apesar dos cuidados tomados no estabelecimento do sistema de seleção dos domicílios para a amostra dos Censos Demográficos brasileiros, infelizmente ocorreu em 1980, com maior intensidade, uma grave tendenciosidade na coleta, principalmente nas áreas menos urbanizadas. Essa tendenciosidade se deu pela inclusão de domicílios com menor número de pessoas nas linhas de amostra da folha de coleta.

É oportuno registrar que a manipulação por parte do recenseador, incluindo ou excluindo da linha de amostra domicílios com menor número de pessoas, domicílios vagos e domicílios fechados, verificada nos Censos Brasileiros, também ocorreu, em 1970, no Censo de População dos Estados Unidos.

Embora algumas tendenciosidades cometidas durante a coleta possam ser atenuadas, através de processo de estimação, esforços e controles devem ser previstos e seguidos rigorosamente durante a coleta, a fim de que essas tendenciosidades não prejudiquem significativamente as demais fases da apuração do Censo.

Por exemplo, os resultados das Tabulações Avançadas no Censo Demográfico de 1980 foram prejudicados pela aludida tendenciosidade. O processo de estimação adotado não corrigiu os efeitos causados por essa falha de coleta. Em conseqüência, principalmente as estimativas das características de famílias e domicílios foram superestimadas.

É oportuno assinalar que o autor deste trabalho, por ocasião do Censo Escolar do Espírito Santo de 1977,encontrou o mesmo tipo de tendenciosidade ocorrida no Censo Demográfico de 1980, porém com menor intensidade. Naquela época, quando foi percebida a tendenciosidade do Censo Escolar, as tabulações com os resultados das características sócio-econômicas estavam sendo

emitidas. Conseqüentemente foram feitas modificações no processo de estimação, a fim de minimizar os efeitos daquela tendenciosidade, e novas tabulações foram emitidas. Infelizmente, por falta de tempo hábil e de recursos financeiros, não foi possível eliminá-la totalmente, da maneira que seria desejável.

Por esse motivo e prevenindo-se contra a possível ocorrência dessa tendenciosidade no Censo Demográfico de 1980, antes de ser realizada a sua coleta, foi concebido o controle dessa tendenciosidade no processo de expansão da amostra, através da variável tamanho do domicílio utilizada como variável de pós-estratificação.

Nesses dois Censos, foram aplicados processos automáticos de seleção sistemática com diferenciações. No Censo Escolar, os domicílios foram selecionados através do número de ordem natural de enumeração que os domicílios receberam no arrolamento das unidades do setor, que em função do algarismo final do seu número, 1 em 5 ou 1 em 10 domicílios, conforme o município, foi selecionado para a amostra. Além disso, a coleta não foi remunerada. No Censo Demográfico, os domicílios foram selecionados através de dispositivos automáticos de seleção, com auxílio de duas folhas de coleta e regras de enumeração, a fim de quebrar possíveis existências de séries cíclicas do universo e de modo a evitar que o recenseador não tivesse outras alternativas de registro nas folhas de coleta; outrossim, a coleta foi totalmente remunerada.

A ocorrência da mesma espécie de tendenciosidade nesses dois Censos é atribuída à possível rapidez com que o recenseador tentou concluir a enumeração de sua área de trabalho, pois a sua tarefa foi concluída em menor tempo, com a inclusão deliberada de domicílios com menor número de pessoas, na seleção da amostra.

Portanto, independentemente do estabelecimento de regras mais elaboradas, com o intuito de evitar as ocorrências de tendenciosidades na coleta, motivadas por falhas do recenseador, inclusive as intencionais, é indispensável o controle rigoroso do trabalho executado pelo recenseador, principalmente no que diz respeito aos critérios e regras de seleção da amostra.

Além disso, o treinamento de todo o pessoal envolvido na coleta deve ser claro e objetivo, mostrando os principais erros alheios à amostragem cometidos por ocasião da coleta e suas conseqüências, a fim de evitar ou diminuir a sua incidência. Pressupõe-se que a rede de coleta, consciente desses problemas, desempenharia sua função com maiores cuidados.

3.9 — Processos de Estimação

Conforme foi visto na exposição anteriormente apresentada, o Censo Demográfico, com aplicação de amostragem na coleta, é concebido de modo que algumas características, consideradas básicas, são levantadas para todas as unidades da população, enquanto as demais características da investigação censitária são coletadas apenas nas unidades selecionadas para a amostra. Os resultados finais das apurações do Censo são obtidos através das tabulações referentes às características investigadas para o universo e pelas estimativas das características dos itens levantados por amostragem.

O processo de estimação objetiva a obtenção de estimativas de todos os itens da investigação censitária, para a população pesquisada, principalmente daqueles investigados por amostragem, separadamente, ou de forma cruzada.

O conhecimento de algumas características básicas de todas as unidades da população possibilita o uso do processo de estimativa de razão com pós-estratificação, aumentando a eficiência da estimação, tornando mais consistentes as estimativas das características investigadas por amostra com os respectivos valores da apuração do universo, trazendo maior credibilidade aos resultados (6).

Após a coleta do Censo geralmente se observa, em cada área de coleta, que a fração de amostragem resultante difere da que foi estabelecida no planejamento, devido a diversas causas e tendenciosidades introduzidas na coleta. No Censo Demográfico brasileiro de 1980, as frações de amostragem dos domicílios da situação urbana e rural foram, respectivamente, 25,8% e 26,9%, e a da população atingiu 24,7%.

Em face dos recursos de informações disponíveis e das particularidades da amostra efetivamente selecionada, o processo de

estimação deve ser concebido de modo a neutralizar a variação ocorrida na fração da amostragem, bem como controlar e minimizar os efeitos de algumas tendenciosidades passíveis de ocorrerem na coleta, pois nem sempre é possível eliminá-las. Portanto, é necessário que o processo de estimação produza estimativas que atendam ao seguinte: controlar ou neutralizar as tendenciosidades ocorridas na seleção da amostra; reduzir a variabilidade das estimativas; aumentar a consistência entre os resultados obtidos na apuração do universo e as estimativas provenientes da amostra; permitir o cálculo de estimativas razoavelmente precisas do erro de amostragem; e minimizar o custo de processamento. Esses pontos, considerados pelos Estados Unidos na escolha do processo de estimação do seu Censo de 1970, são ainda hoje plenamente válidos.

Inúmeros processos de estimação podem ser concebidos. Entretanto, convém que o benefício-custo dos resultados, traduzido pelo aumento da precisão das estimativas em função do custo de processá-las, seja o mais adequado.

A experiência internacional revelada pelas diversas aplicações já adotadas nos Censos dos países anteriormente citados é fundamental para a melhor compreensão da matéria. Assim sendo, relatam-se, a seguir, os processos de estimação já utilizados pelos Estados Unidos, Canadá, Argentina e Brasil.

Os Estados Unidos, no Censo de 1950, utilizaram o método de estimativa simples. No Censo de 1960, com o auxílio de computadores eletrônicos de alta velocidade, foi possível adotar o processo de estimativa de razão por grupos. Nos Censos de 1970 e 1980, foi utilizado o processo iterativo de estimativa de razão por grupos. O Canadá, também, empregou um processo iterativo de estimativa de razão nos Censos de População referentes aos anos de 1971, 1976 e 1981, denominado *Raking Ratio Estimators Procedure* — RREP.

A Argentina adotou no Censo de 1980 o processo de estimativa de razão por grupos. O Brasil, no Censo de 1960, usou um processo de estimativa de razão por grupos, semelhantes ao dos Estados Unidos, para aquele ano, nas tabulações inicialmente estabelecidas e efetivadas para 11 unidades da federação. Posteriormente, para concluir a apuração do Censo, foram utilizadas apenas as informações do Boletim de Amostra; por esse motivo, o processo de estimativa de razão adotado baseou-se na população urbana e rural constante das Sinopses Preliminares.

Novamente, em 1970, o Brasil seguiu um processo de estimação semelhante ao inicialmente previsto em 1960. No Censo de 1980, após vários estudos (19), foi adotado um processo de estimativa de razão, semelhante ao seguido pelo Canadá, que no Brasil recebeu a denominação de Processo Iterativo de Estimação por Totais Marginais — PIETOM.

Conforme se observa, excetuando-se o Censo dos Estados Unidos de 1950, todos os processos de estimação utilizados seguiram o método de estimativa de razão por grupos com pós-estratificação e com diversas variações.

Antes de apresentar a descrição de cada um desses métodos e proceder à análise comparativa deles, é oportuno salientar que todos eles visam estabelecer o peso ou inflator atribuído a cada unidade da amostra, obtido através do método de estimativa de razão.

3.9.1 — Estimativa de Razão

De modo geral, a estimativa do total X da característica x obtida com o uso do estimador de razão é expressa por:

$$\hat{X} = \sum_i \frac{\hat{X}_i}{\hat{Y}_i} \cdot Y_i \qquad [2]$$

onde:

$\hat{X}_r$ — estimativa de razão do total X;

$\hat{X}_i$ — estimativa simples da característica x, no i-ésimo grupo, baseado na amostra;

Y_i — total da característica y, calculado com informações de todas as unidades do universo, no i-ésimo grupo.

Como em [2] as estimativas simples $\hat{X}_i$ e $\hat{Y}_i$ resultam dos totais amostrais de x e y multiplicados pelo inverso da fração de amostragem, em cada grupo i, então o estimador $\hat{X}_r$ pode ser assim representado:

$$\hat{X}_r = \sum \frac{Y_i}{y_i} \cdot x_i = \sum W_i x_i \qquad [3]$$

onde:
x_i — total de característica x, na amostra, para o i-ésimo grupo;
y_i — total da característica y na amostra, para o i-ésimo grupo;
Y_i — total da característica y, no universo, para o i-ésimo grupo.

O valor $W_i = \frac{Y_i}{y_i} = \frac{N_i}{n_i}$ constitui o peso de x_i no i-ésimo grupo, ou seja, é a razão entre o número de unidades N_i, no universo e o respectivo número n_i, na amostra em cada grupo i.

Observa-se em [3] que o total estimado $\hat{Y}_h$ corresponde à soma dos pesos atribuídos a cada unidade de amostragem que possua a característica h, quando Y_h representa uma categoria ou atributo. Quando Y_h é uma variável (número de filhos, renda, etc.), $\hat{Y}_h$ corresponde à soma dos produtos do peso de cada unidade da amostra que possua a característica pelo seu respectivo valor.

Entre os processos de estimação citados anteriormente, destacam-se os seguintes:

— *Ratio Estimation Groups* — conhecido no Brasil como Estimativa de Razão por Grupos de Controle;

— *Iterative Ratio Estimation Groups;*e

— *Raking Ratio Estimators Procedure* (RREP), denominado no Brasil, em 1980, como Processo Iterativo de Estimação por Totais Marginais — PIETOM.

3.9.2 — Ratio Estimation Groups

O processo Estimativa de Razão por Grupos de Controle objetiva aumentar a eficiência da amostragem através do recurso da pós-estratificação e do uso da estimativa de razão em cada pós-estrato. A pós-estratificação possibilita introduzir os efeitos da estratificação, enquanto a estimativa de razão procura assegurar o controle da ponderação da amostra, de modo tal que, combinados, aumenta-se a consistência das estimativas.

As variáveis consideradas como mais importantes, ou através das quais se procura explorar os ganhos da estratificação, por um lado, e aquelas que possibilitem controlar ou eliminar as tendenciosidades cometidas na seleção da amostra constituem as variáveis de estratificação ou pós-estratificação. Através do cruzamento dessas variáveis, é definida a matriz de ponderação cujas células constituem os pós-estratos ou grupos de controle; no Censo Demográfico de 1970 essa matriz possuía 46 células (15).

Este processo de estimação consiste em calcular o peso ou fator de expansão a ser atribuído a cada unidade da amostragem, através da razão entre o número de unidade no universo e o respectivo número na amostra em relação a cada grupo ou pós-estrato.

Como, em cada grupo, o número de unidades no universo N pode ser pequeno, ou então, a razão calculada (w) pode ultrapassar certo valor, são estabelecidas duas restrições que o processo deve atender. Quando uma das restrições não é satisfeita procede-se ao agrupamento dos pós-estratos, de acordo com critérios preestabelecidos, até que as condições sejam satisfeitas ou se esgotem as possibilidades de agrupamento, quando, então, o cálculo dos pesos é efetuado de qualquer modo.

Nos Censos Demográficos brasileiros de 1960 e de 1970, essas duas restrições foram $N < 100$ e $w > 16$. É interessante registrar que no Censo de População dos Estados Unidos, referente ao ano de 1960, em que foi utilizado um processo semelhante ao seguido pelo Brasil, essas restrições foram $N < 150$ e $w > 16$.

A fim de evitar a produção de resultados inconsistentes, a partir da amostra, são utilizados no cálculo das estimativas pesos inteiros imediatamente próximos à razão fracionária encontrada em cada grupo.

Para a realização dos cálculos dos pesos das estimativas da população são utilizadas duas matrizes, uma contendo a contagem da população no universo e a outra com a contagem na amostra. Os passos seguidos no cálculo do peso w_{ij} até a atribuição dos pesos de cada unidade na amostra, em cada célula (i,j), são os seguintes:

I — Cálculo inicial de $w_{ij} = \frac{N_{ij}}{n_{ij}}$, para cada célula (i,j);

II — Aplicação dos critérios e agrupamentos dos pós-estratos;

III — Obtenção da matriz final;

IV — Cálculo do peso final w_{ij}, para cada célula da matriz final (obtida após o reagrupamento); e

V — Determinação e atribuição do peso inteiro a cada unidade da amostra, de cada célula (i,j).

3.9.3 — Iterative Ratio Estimation Groups

Este processo também utiliza a pós-estratificação e a estimativa de razão; foi concebido de modo a reduzir o erro de amostragem e os efeitos das tendenciosidades ocorridas na seleção da amostra, melhorar a consistência das estimativas, ser econômico na sua execução, etc. Naturalmente, em se usando a pós-estratificação e a estimativa de razão, os princípios gerais desses recursos técnicos, alguns deles já citados também são explorados neste processo.

O método de estimação é realizado através da estimativa de razão desenvolvida em vários estágios sucessivos, geralmente 2 ou 3, numa dada seqüência, sendo posteriormente repetidos novamente, isto é, iterados.

Inicialmente, são definidos os grupos das variáveis de estratificação, um para cada estágio de estimação. Esses grupos e variáveis, dispostos em tabelas cruzadas, dão origem à matriz de ponderação, de forma que as células dos cruzamentos da matriz constituem os pós-estratos envolvidos no processo de estimação. Os grupos correspondem às colunas, ou às linhas, ou às partes dessa célula.

A fim de controlar a ocorrência de baixa freqüência de unidades de amostragem numa célula qualquer, ou então, o surgimento de um peso superior a um máximo pré-fixado, em cada grupo são estabelecidos critérios restritivos, os quais são aplicados a cada grupo e em cada estágio de estimação, visando a aumentar a confiabilidade do processo de estimação. Se, para algum pós-estrato, o critério restritivo não é atendido, procede-se ao agrupamento deste com outro, no mesmo estágio, de acordo com regras estabelecidas *a priori*, até que as condições restritivas sejam aceitas ou se esgotem.

Inicialmente é atribuído a cada unidade de amostragem um peso inicial que é, aproximadamente, igual ao inverso da fração de amostragem. A soma desses pesos, em termos de célula (i, j, k), representa a primeira estimativa gerada pelo processo, para aquela célula; somando-se esses resultados por linhas e por colunas, são obtidas as estimativas dos totais marginais, para cada linha i e para cada coluna j da matriz.

Em seguida, no primeiro estágio de estimação, realiza-se o primeiro ajustamento do peso inicial, multiplicando-o pela razão entre o total marginal do universo e o respectivo total marginal estimado na primeira vez, por grupo, do estágio I.

No estágio II de estimação, os pesos ajustados no estágio I, isto é, na operação anterior, são novamente ajustados, isto é, pela segunda vez, pela razão entre os totais marginais do universo e as estimativas dos totais marginais obtidos com o peso ajustado pela primeira vez, em cada grupo do estágio II.

Se o processo é desenvolvido em dois estágios, então as operações de ajustamento do peso são repetidas na mesma seqüência. Havendo um terceiro estágio, a operação continuaria como se verá adiante.

No terceiro estágio da estimação, os pesos ajustados no estágio II são ajustados novamente, ou seja, pela terceira vez, através da razão entre os totais marginais do universo e as respectivas estimativas obtidas com os pesos ajustados pela segunda vez, em cada grupo do estágio III.

Após esses três estágios de ajustamentos, a seqüência de operações descrita anteriormente é repetida.

Os pesos resultantes da segunda iteração, para o último estágio, constituem os pesos finais do processo. Entretanto, para evitar complicações no arredondamento dos dados tabulados, são utilizados, para obtenção das estimativas finais, pesos inteiros imediatamente próximos ao peso final encontrado em cada célula.

No Censo de População dos Estados Unidos referente a 1970, nas estimativas das características da população foram usados três estágios de estimação abrangendo $(18 \times 24 \times 21) = 864$ células. Neste Censo, os critérios restritivos foram: $N > 85$, para a amostra de 15%, estabelecidos de modo a se ter $n \geqslant 3$ e $w \leqslant 20$, em cada grupo.

Os passos seguidos no cálculo do peso w, até a atribuição do peso de cada unidade da amostra, em cada célula, quando se têm

três estágios, são os seguintes, conforme meu entender:

I — Cálculo de $w_{ijr}^{(1)} = \frac{N}{n}$, para cada célula (i,j,k)

Aplicação das regras restritivas;

II — Cálculo de $N_{ijr}^{(1)} = w_{ijr}^{(1)} \cdot n_{ijr}$, para cada célula;

III — Cálculo de $N_{i..}^{(1)} = \Sigma_j \Sigma_1 N_{ijk}^{(1)}$, para cada linha i, do estágio I;

IV — Cálculo de $\frac{N_{i..}}{N_{i..}^{(1)}}$ para cada linha (grupo) do estágio I.

V — Cálculo de $w_{ijk}^{(2)} = w_{ijr}^{(1)} \cdot \frac{N_{i..}}{N_{i..}^{(k)}}$

primeiro ajustamento, ou

peso ajustado pelo estágio I;

VI — Cálculo de $N_{ijk}^{(2)} = w_{ijk}^2 \, n_{ijk}$, para cada célula;

VII — Cálculo de $N_{.j.}^{(2)} = \Sigma_i \Sigma_k N_{ijk}^{(2)}$, para cada coluna;

VIII — Cálculo de $\frac{N_{.j.}}{N_{.j.}^{(2)}}$;

IX — Cálculo de $w_{ijr}^{(3)} = w_{ijr}^{(2)} \cdot \frac{N_{.j.}}{N_{.j.}^{(2)}}$

segundo ajustamento, ou peso ajustado pelo estágio I;

X — Cálculo de $N_{ijk}^{(3)} = w_{ijk}^{(3)} \, n_{ijk}$, para cada relação ao estágio 3;

XI — Cálculo de $N_{..1}^{(3)} = \Sigma_i \Sigma_j N_{ijk}^{(3)}$, para a 3ª dimensão.

XII — Cálculo de $\frac{N_{..k}}{N_{..k}^{(3)}}$

XIII — Cálculo de $w_{ijk}^{(4)} = w_{ijk}^{(3)} \cdot \frac{N_{..k}}{N_{..k}^{(3)}}$,

terceiro ajustamento, ou peso ajustado pelo estágio III.

Após obter $w_{ijk}^{(4)}$ o processo é repetido novamente a partir do passo II.

onde:

i — número de grupos do estágio I, linhas i = 1, 2, ...18

j — número de grupos do estágio II, colunas j = 1, 2, ...24

k — número de grupos do estágio III, parte de uma célula (i,j); k = (1,2)

3.9.4 — Raking Ratio Estimators Procedure

Do mesmo modo que nos dois processos anteriores, este também faz uso da estimativa de razão com pós-estratificação. O emprego desses dois recursos técnicos mantém os objetivos já citados, ou seja, a produção de estimativas consistentes, com erros de amostragem menores e com o controle das tendenciosidades cometidas durante a seleção da amostra.

O processo de estimação é realizado através da estimativa de razão desenvolvida em duas dimensões ou estágios, de modo sucessivo e em seqüência, sendo posteriormente repetido ou iterado. Maiores detalhes sobre essa metodologia, além dos apresentados abaixo, encontram-se em (15, 6, 3, 4 e 5).

Os grupos das variáveis de estratificação, um para cada dimensão, são definidos visando a reduzir o erro de amostragem, explorando a correlação existente entre as características sob controle e sobretudo assegurar a consistência para os totais marginais entre as estimativas obtidas pela amostra e os respectivos resultados da tabulação dos dados do universo, para as características escolhidas e consideradas mais importantes. Com esses dois grupos de variáveis, dispostos em tabelas cruzadas, constrói-se a matriz utilizada na ponderação, onde cada célula da matriz constitui um pós-estrato.

Neste processo, também são utilizados critérios restritivos para controlar o número mínimo de unidades da amostragem em cada grupo e o respectivo peso inicial, resultante de fração de amostragem ocorrida na coleta, por coluna e por linha da matriz. Os critérios restritivos são aplicados, procedendo-se a agrupamentos de linhas ou de colunas que não os satisfazem, de acordo com as regras de agrupamento estabelecidas *a priori*, até que as condições sejam

aceitas ou o processo de agrupamento chegue ao fim.

De início, são contadas as unidades da amostra e do universo, obtendo-se, então, a matriz da amostra e a matriz do universo. Com base nos totais marginais dessas duas matrizes, procede-se à aplicação dos critérios restritivos, dando seqüência ao desenvolvimento do processo.

No Censo de População do Canadá, referente ao ano de 1976, nas estimativas das características das pessoas, foi utilizada uma matriz com (30 × 34) = 1.020 células. Os critérios restritivos foram $N < 35$ e $W > 12$, aplicados de modo que se qualquer total da população, por linha ou coluna, foi menor que 35, ou então se o peso inicial obtido pela razão entre a contagem da população no universo e a contagem na amostra, para cada linha ou coluna, foi maior do que 12, ou então se o total marginal de contagem na amostra foi zero, procedia-se ao agrupamento das linhas ou colunas onde isso ocorreu. A fração de amostragem desse Censo foi f = 1/3.

O Brasil aplicou este processo no Censo Demográfico de 1980. As características da população foram estimadas usando-se uma matriz com (15 × 48) = 720 células, sendo adotados os seguintes critérios de restrição $n_i \geqslant 5$ ou $n_{.j} \geqslant 5$ e pesos $1 \leqslant \frac{N_{.i}}{n_{.i}} \leqslant 8$ e $1 \leqslant \frac{N_{.j}}{n_{.j}} \leqslant 8$. A fração de amostragem desse Censo foi f = 1/4.

Note-se que $w_{i.} = \frac{N_{i.}}{M_{i.}}$ e $w_{.j} = \frac{N_{.j}}{M_{.j}}$.

A seqüência de cálculo adotada neste processo, aproveitando a descrição contida na metodologia do Censo Demográfico de 1980, é a seguinte:

"Considerando como ponto de partida a coluna, as etapas para o cálculo do peso são apresentadas a seguir:

I) inicialmente são obtidos os totais populacionais por célula N_{ij};

II) o mesmo é feito para os valores da amostra, obtendo-se os totais amostrais, n_{ij};

III) são calculados os totais marginais $N_{.j}$ e $N_{i.}$, para as colunas e linhas da matriz do universo e o total marginal da coluna $n_{.j}$, da matriz da amostra;

IV) é atribuído a cada unidade da amostra um peso inicial $\frac{N_{.j}}{n_{.j}}$, da coluna correspondente ao valor da unidade da amostra. Este estágio assegura consistência para as características das colunas da matriz;

V) utilizando os pesos iniciais atribuídos é efetuado o cálculo da estimativa para cada linha da matriz do universo;

VI) para cada célula da matriz de ponderação é ajustado o peso inicial de cada unidade amostral, na respectiva linha da matriz, por um fator que assegura a consistência da linha com os totais da população; e

VII) são repetidos os dois últimos estágios para as colunas, usando o peso ajustado, até que o processo iterativo se estabilize.

No caso em que se tome como ponto de partida a linha, basta substituir linha por coluna e vice-versa, nos passos anteriormente descritos.

O algoritmo para o cálculo de peso w_{ij} para cada célula (i,j) da matriz de ponderação é descrito a seguir, onde todos os termos se referem às matrizes após os reagrupamentos citados, quando for o caso.

I — Calcula-se $w_{ij}^{(1)} = \frac{N_{.j}}{n_{.j}} \rightarrow$ para cada célula (i,j);

II — Calcula-se $N_{i.}^{(1)} = \sum_{j=1}^{j} w_{ij}^{(1)} n_{ij} \rightarrow$ para cada linha i;

III — Calcula-se $w_{ij}^{(2)} = w_{ij}^{(1)} = \frac{N_{i.}}{N_{i.}^{(1)}} \rightarrow$ para cada célula (i,j);

IV — Calcula-se $N_{.j}^{(2)} = \sum_{i=1}^{I} w_{ij}^{(2)} n_{ij} \rightarrow$ para cada linha j;

V — Calcula-se $w_{ij}^{(3)} = w_{ij}^{(2)} = \frac{N_{i.}}{N_{.j}^{(2)}} \rightarrow$ para cada célula (i,j);

VI — Calcula-se $N_{i.}^{(3)} = \sum_{j=1}^{J} w_{ij}^{(3)} n_{ij} \rightarrow$ para cada linha i; e

VII — Calcula-se $w_{ij}^{(4)} = w_{ij}^{(3)} = \frac{N_{i.}}{N_{i.}^{(3)}} \rightarrow$ para cada célula (i,j);

onde:

$$N_{i.} = \sum_{j=1}^{J} N_{ij} \qquad N_{.j} = \sum_{i=1}^{I} N_{ij}$$

$$n_{i.} = \sum_{j=1}^{J} n_{ij} \qquad n_{.j} = \sum_{i=1}^{I} n_{ij}$$

onde:

$N_{ij}\rightarrow$ é o número de pessoas, na população, na i-ésima linha e j-ésima coluna da matriz;

$n_{ij}\rightarrow$ é o número de pessoas, na amostra, na i-ésima linha e j-ésima coluna da matriz;

$I\rightarrow$ é o número de linhas da matriz; e

$J\rightarrow$ é o número de colunas da matriz.

A fim de evitar a produção de resultados inconsistentes em uma mesma tabulação são utilizados pesos inteiros para obtenção das estimativas. Assim, se o peso $w_{ij}^{(4)}$ não for inteiro, procede-se do seguinte modo: toma-se a estimativa do total de pessoas na célula (i,j), isto é, $\tilde{N}_{ij} = w_{ij}^{(4)} n_{ij}$; designa-se por $\tilde{N}_{ij}$ o arredondamento de $\tilde{N}_{ij}$ e por p_{ii} a parte inteira da razão entre $\tilde{N}_{ij}$ e n_{ij}; então $r_{ij} = \tilde{N}_{ij} - p_{ij}n_{ij}$ é o número de pessoas na célula (i,j) que recebem o peso inteiro $p_{ij} + 1$, enquanto as restantes $n_{ij} - r_{ij}$ recebem o peso p_{ij}.''

3.9.5 — Avaliação dos processos examinados

O processo de estimação adotado pelo Canadá, no Censo de 1971 e posteriores, é semelhante ou igual ao processo seguido pelos Estados Unidos, a partir do Censo de 1970, embora com denominações diferentes, pois ambos empregam o *Raking Ratio Estimatiors Procedure*.

Por outro lado, as estimativas resultam da soma dos pesos atribuídos às unidades da amostra. Como esses pesos são afetados pelas tendenciosidades porventura ocorridas na coleta, naturalmente o processo de estimação que melhor neutraliza ou atenua essas tendenciosidades possibilita melhores resultados. Os pesos calculados a partir da contagem das unidades do universo e da amostra, por célula, atendem mais adequadamente a esse objetivo; além disso, as variações de amostragem que a amostra pode conter são incorporadas às estimativas de modo a torná-las consistentes com o universo.

O método dos grupos de controle utiliza uma única fase de cálculo do peso, a nível de célula, dispensando qualquer espécie de ajustamento. Dessa forma, ele possibilita o pleno uso das informações do universo e da amostra disponíveis por cruzamento da matriz de ponderação, obtendo assim, a melhor consistência entre os dados do universo e as respectivas estimativas das características que definem a matriz de ponderação, a nível de células e das marginais dessa matriz.

O processo RREP objetiva não só produzir estimativas consistentes com o universo, como também atenuar as tendenciosidades da coleta, porém, a partir dos totais marginais da matriz de ponderação, para as variáveis definidoras de cada estágio de estimação. Conseqüentemente, o peso atribuído às unidades da amostra de uma célula, resultante do cruzamento da matriz, necessita ser ajustado, a fim de se conseguir a consistência entre as estimativas e os respectivos resultados do universo, para os totais marginais de cada estágio de estimação.

De acordo com o algoritmo de cálculo do RREP, nota-se que ele se apóia mais na amostra do que no universo, a partir do passo II, em termos de célula. Dessa forma, ele é influenciado por flutuações de amostragem, geralmente mais fortes em áreas de ponderação de menor tamanho. Com o abandono das informações do universo, por célula, são diminuídos os efeitos da pós-estratificação, conseqüentemente, os resultados das estimativas finais são afetados pelas aludidas flutuações, tanto no sentido quanto na intensidade. Além disso, o processo exige iteração no cálculo dos pesos de modo a conseguir melhor consistência marginal das estimativas. De fato, as estimativas provenientes da utilização dos pesos calculados para cada total marginal só são mais consistentes com as respectivas marginais. Por esse motivo, os pesos resultantes dos dois primeiros ajustamentos (linha e coluna), no caso da utilização de dois estágios de estimação, são novamente reajustados. Dessa forma, a matriz de pesos (final) fica ajustada em torno da fração de

amostragem, conseqüentemente proporcionando menor variância.

O processo de estimação dos grupos de controle adotado no Censo Brasileiro até 1970, utiliza poucas variáveis de controle na sua tradicional concepção, com menor número de células, enquanto o processo RREP emprega um número bem maior de variáveis e de células; por esse motivo atribui-se ao RREP maior flexibilidade na estimação, pois exige menor número de agrupamentos, em função dos seus critérios restritivos serem limitados às marginais.

Os métodos do Grupo de Controle e o RREP se fundamentam em princípios de pós-estratificação e de estimativas de razão. Naturalmente, os critérios restritivos que esses princípios requerem devem ser observados igualmente. Assim sendo, é difícil comparar, ou não se deve comparar métodos de estimação cujos critérios restritivos sejam bastante diferenciados ou até extremados.

Enquanto no grupo de controle cada célula da matriz de ponderação constitui um puro pós-estrato, no processo RREP a célula não é um puro pós-estrato, pois os critérios restritivos não são igualmente aplicados às células, mas sim às marginais. Por uma questão de regra, o método grupo de controle tradicional limita a um mínimo o número de unidades na amostra em cada célula, enquanto o método RREP aceita até uma ou nenhuma unidade na amostra de cada célula, daí proporcionar menor número de agrupamentos.

Os excelentes estudos experimentais, realizados pelo IBGE, por ocasião do Censo Experimental de 1980 (13), revelaram que os resultados provenientes das estimativas elaboradas com base nos métodos grupos de controle e RREP não apresentaram comportamentos divergentes, porém, ligeiramente favoráveis ao RREP. Como praticamente não ocorreu tendenciosidade na coleta desse Censo, a eliminação desta não foi apreciada em relação a outros processos.

Uma experiência interessante seria a utilização da mesma matriz de ponderação, adotada no método RREP para o cálculo de pesos, através dos grupos de controle, porém mantidas as mesmas regras restritivas do RREP em relação aos totais das linhas e colunas do grupo de controle, ao invés das células. Assim as células com até uma unidade na amostra seriam usadas no cálculo dos pesos pelo método dos grupos de controle.

É importante salientar que os critérios restritivos são definidos para controlar valores inadequados, ou seja, pesos elevados, fugindo à normalidade, e freqüências muito reduzidas.

A avaliação dos custos de processamento de cada processo de estimação em relação aos ganhos na eficiência das estimativas (melhor precisão e melhor consistência) deveria também ser considerada ao se avaliarem os benefícios-custos em futuras experiências.

3.10 — Diretrizes para a Construção do Processo de Estimação

Depreende-se do exposto anteriormente que o processo de expansão do Censo visa a produzir estimativas consistentes com o universo, controlar ou atenuar as tendenciosidades de coleta e as flutuações da amostragem, reduzir a variabilidade das estimativas, não só em ternos de agregado de tabulação — Unidades da Federação — como também em relação a áreas menores etc.

Nos processos apresentados fez-se uso da pós-estratificação e foram empregados critérios restritivos componentes da estratégia de expansão. Vários aspectos afetam os resultados, dentre eles destacam-se os seguintes:

a) requisitos do processo de estimação;

b) redução da variabilidade;

c) controle das tendenciosidades e das flutuações de amostragem;

d) uso da estratificação e/ou pós-estratificação;

e) consistência;

f) nível geográfico de estimação — área de ponderação;

g) estratégia de ponderação; e

h) uso de duas ou mais matrizes de ponderação.

3.10.1 — Requisitos do processo de estimação

Em conformidade com a teoria de amostragem, o estimador não deve ser tendencioso, mas consistente e eficiente. O ideal seria um estimador que possuísse simultaneamente essas condições, levando em conta, ainda, o custo de sua obtenção.

Os censos por amostragem geralmente são complexos devido à sua dimensão, aos seus propósitos e ao seu custo. Considerando esses fatores e a necessidade de se produzir um processo de estimação que atenda às condições teóricas, em toda a sua plenitude, torna-se difícil a elaboração do processo ideal. A preocupação em se produzirem estimativas não tendenciosas deve ser constante e de igual forma a consistência e até certo ponto a maior eficiência.

O processo de estimativa de razão com uso da pós-estratificação procura controlar ou atenuar as tendenciosidades de coleta, tornando-o mais consistente. Não obstante, o processo de estimativa de razão contém, em sua concepção, alguma tendenciosidade, a qual é desprezível para amostras grandes além de possuir menor variabilidade que o estimador simples. Os resultados do Censo, na maioria das vezes, referem-se a variáveis que não constam da matriz de ponderação, embora possam estar combinados ou cruzados com alguma variável da matriz. Através dos efeitos de correlação entre essas variáveis e as da matriz de ponderação, e do uso do estimador de razão com pós-estratificação, as estimativas são beneficiadas de um modo geral, com melhor precisão, além de não se restringirem os benefícios apenas às variáveis da matriz de ponderação.

O cálculo do peso ou dos fatores de expansão empregados nas estimativas utiliza duas matrizes de ponderação, definidas conforme a pós-estratificação estabelecida, a matriz do universo que contém a contagem da apuração do universo e a matriz da amostra que contém a contagem da apuração das unidades da amostra.

Os estimadores focalizados no presente documento, RREP, REG ou GC utilizam esses princípios, entretanto com custos diferenciados. É provável que nem sempre o estimador mais eficiente, isto é, com menor variabilidade, proporcione menor custo e simultaneamente seja menos tendencioso e mais consistente na plenitude de seus propósitos.

3.10.2 — Redução da variabilidade

O planejamento da amostra aplicada no Censo deve conceder a possibilidade de serem usados processos distintos de estimação de modo a produzirem resultados com menores erros de amostragem.

A estimativa de razão, satisfeitas as suas condições de aplicabilidade, reduz tais erros, os quais são diminuídos ainda mais quando, em combinação com a estimativa de razão, se emprega a pós-estratificação. Portanto, é importante que a redução da variabilidade seja conseguida para as diversas variáveis pesquisadas no Censo, para a obtenção de estimativas de tamanhos diferentes e para níveis geográficos de tamanhos diversificados.

3.10.3 — Controle das tendenciosidades e das flutuações de amostragem

A estimativa de razão é obtida a partir da divisão do total de pessoas ou de domicílios no universo pelos respectivos totais na amostra. A amostra efetiva flutua em relação ao tamanho esperado por diversas causas, como por exemplo: tendenciosidades introduzidas na coleta; flutuações de amostragem devido à dimensão (quanto ao número de unidades de amostragem) da área de coleta ou região para a qual serão elaboradas as estimativas; e por causas aleatórias. Esses fatores influem diretamente nos resultados, acarretando superestimações ou subestimações conforme o caso, se não forem controlados.

No Censo Demográfico de 1970, ocorreu a tendenciosidade de serem incluídos na amostra domicílios de tamanho menor (número de pessoas) em detrimento dos domicílios de maior tamanho. No Censo Escolar do Espírito Santo realizado em 1977, essa tendenciosidade ocorreu de forma mais forte. Como esse tipo de tendenciosidade é difícil de ser controlado na coleta, especialmente na zona rural e em áreas faveladas, e os seus efeitos afetam as estimativas, foram tomadas precauções no planejamento do Censo Brasileiro de 1980, adotando-se a variável tamanho do domicílio como variável de pós-estratificação, a fim de controlar ou atenuar a aludida tendenciosidade na obtenção das estimativas finais.

Posteriormente, verificou-se que esse defeito de coleta ocorreu fortemente em 1980. Os resultados das tabulações avançadas foram sensivelmente afetados por essa tendenciosidade, pois o processo de expansão seguido não controlou essa defi-

ciência. Porém os resultados definitivos do censo foram controlados como o previsto antecipadamente.

Na Tabela 5 são mostradas as características dos chefes das famílias, por número de componentes. Essa tabela permite visualizar a influência do tamanho da família nos resultados. De fato, os perfis das distribuições de idade, anos de estudo, renda, etc. do chefe variam conforme o número de componentes. O domicílio, na maioria das vezes, corresponde à família, pois 95% dos domicílios do País são constituídos por uma única família.

Qualquer estimativa amostral apresenta flutuação em relação ao parâmetro correspondente do universo. Como no Censo são elaborados estimativas para áreas de tamanhos variáveis, e para municípios com pouca população (menos de 5.000 habitantes), cujos pesos ou fatores de expansão são calculados de acordo com o critério de expansão através do uso de pós-estratificação, então as flutuações de amostragem aumentam sensivelmente, daí ser indispensável o controle dessa variação a fim de não prejudicar os resultados finais.

3.10.4 — Uso da estratificação e pós-estratificação

A estratificação contribui para a diminuição dos erros de amostragem. Os efeitos da estratificação são marcantes quando as variáveis que definem a estratificação encerram fortes correlações com os itens investigados.

Como se torna difícil planejar e executar o censo com uma estratificação *a priori*, geralmente faz-se uso da pós-estratificação, onde os estratos são formados após a coleta, sendo a mesma conduzida sem considerar a estratificação. Os pós-estratos são definidos envolvendo as variáveis fundamentais: as mais importantes e as que servem para controlar as tendenciosidades de cole-

TABELA 5
FAMÍLIAS E PESSOAS RESIDENTES EM DOMICÍLIOS PARTICULARES, POR NÚMERO DE COMPONENTES, SEGUNDO ALGUMAS DAS PRINCIPAIS CARACTERÍSTICAS DO CHEFE DA FAMÍLIA

(continua)

NÚMERO DE ORDEM	CARACTERÍSTICAS DO CHEFE DA FAMÍLIA	PESSOAS	FAMÍLIAS		
			Total	Número de Componentes	
				1 Pessoa	2 Pessoas
1	TOTAL	117 960 301	26 806 748	1 554 972	4 440 200
	SEXO				
2	Homens	104 792 834	22 616 470	826 248	3 217 387
3	Mulheres	13 167 467	4 190 276	728 724	1 222 613
	GRUPOS DE IDADE				
4	15 a 19 anos	718 267	290 362	34 695	150 068
5	20 a 24 anos	6 327 381	2 112 039	114 365	697 863
6	25 a 29 anos	12 812 356	3 515 948	125 706	652 684
7	30 a 34 anos	15 852 763	3 560 510	107 271	359 680
8	35 a 39 anos	16 473 523	3 171 354	95 419	225 371
9	40 a 44 anos	17 047 697	3 036 335	103 040	212 301
10	45 a 49 anos	14 075 403	2 554 264	106 964	221 036
11	50 a 54 anos	11 861 477	2 348 504	125 908	289 350
12	55 a 59 anos	8 330 138	1 864 421	130 483	325 006
13	60 a 64 anos	5 584 107	1 462 327	146 207	351 144
14	65 a 69 anos	4 172 014	1 254 180	170 373	371 571
15	70 anos ou mais	4 484 235	1 586 740	289 708	574 967
16	Idade ignorada	216 940	49 764	4 833	8 959
	ANOS DE ESTUDO				
17	Sem instrução e menos de 1 ano	39 082 965	8 481 995	697 351	1 424 524
18	1 ano	6 323 456	1 309 447	71 495	192 074
19	2 anos	10 184 962	2 115 406	98 209	295 839

ta e as flutuações de amostragem de forma tal que os benefícios da estratificação sejam diretos, para essas variáveis, e os efeitos indiretos da estratificação para outras variáveis sejam apreciáveis.

Entre as variáveis mais importantes e que afetam sensivelmente os resultados, destacam-se as seguintes:

a) idade das pessoas — a idade das pessoas é um dos principais itens investigados. A escolaridade, certas ocupações, mortalidade, renda, migração, etc., estão estreitamente vinculadas à idade;

b) sexo — constitui uma condição natural de estratificação. A mortalidade, a fecundidade, certas ocupações, etc., estão ligadas ao sexo das pessoas; e

c) situação do domicílio, urbano e rural — constitui um excelente meio de estratificação geográfica. As características de domicílio, fecundidade, mortalidade, escolaridade, certas ocupações, etc., dependem da situação dos domicílios.

Além destas variáveis, outras poderiam ser usadas na definição dos pós-estratos. É fundamental escolher as variáveis que sejam as mais importantes e que guardem correlação com outras variáveis, de modo que os efeitos da estratificação e da correlação entre elas sejam somados pelo uso da estimativa de razão.

Como o pós-estrato também é usado para controlar ou atenuar *a posteriori* as tendenciosidades de coleta e as flutuações de amostragem, é conveniente que entre as variáveis de estratificação figurem algumas que também tenham a função de controle dessas flutuações.

A escolha das variáveis e o detalhamento da pós-estratificação do Censo Demográfico de 1980 é justificada abaixo:

a) idade das pessoas em grupos etários qüinqüenais até 49 anos, a fim de melhorar a elaboração de estudos demográficos em face da inexistência de informações correntes sobre a idade;

TABELA 5

FAMÍLIAS E PESSOAS RESIDENTES EM DOMICÍLIOS PARTICULARES, POR NÚMERO DE COMPONENTES, SEGUNDO ALGUMAS DAS PRINCIPAIS CARACTERÍSTICAS DO CHEFE DA FAMÍLIA

(continua)

FAMÍLIAS							NÚMERO DE ORDEM
Número de Componentes							
3 Pessoas	4 Pessoas	5 Pessoas	6 Pessoas	7 a 10 Pessoas	11 a 14 Pessoas	15 Pessoas ou mais	
5 028 241	4 839 945	3 772 972	2 543 195	4 124 242	483 495	19 486	1
4 185 588	4 285 064	3 415 677	2 331 837	3 867 579	468 145	18 945	2
842 453	554 881	357 295	211 358	256 663	15 350	541	3
69 591	20 437	8 199	3 552	3 630	186	4	4
730 200	357 249	134 518	45 843	30 797	1 157	47	5
1 010 644	874 522	491 426	219 043	138 716	2 908	99	6
645 060	878 514	690 053	419 702	445 532	14 479	219	7
371 488	615 048	611 316	455 751	737 362	58 498	1 101	8
311 468	475 184	512 002	421 120	871 627	125 492	4 101	9
302 288	388 320	386 540	319 904	704 679	118 873	5 660	10
351 030	365 225	328 988	257 425	541 076	85 321	4 181	11
336 554	292 966	237 339	173 939	322 795	43 120	2 221	12
297 713	220 202	159 352	103 601	164 845	18 255	1 008	13
265 590	169 831	110 025	65 255	92 157	6 879	499	14
328 270	175 080	97 072	53 418	62 736	5 188	301	15
8 345	7 369	6 142	4 642	8 290	1 139	45	16
1 360 933	1 205 965	1 051 997	813 465	1 686 346	231 914	9 500	17
208 024	203 349	178 312	140 017	277 098	37 510	1 568	18
343 721	351 124	304 429	234 541	432 251	53 421	1 871	19

TABELA 5
FAMÍLIAS E PESSOAS RESIDENTES EM DOMICÍLIOS PARTICULARES, POR NÚMERO DE COMPONENTES, SEGUNDO ALGUMAS DAS PRINCIPAIS CARACTERÍSTICAS DO CHEFE DA FAMÍLIA

(continua)

NÚMERO DE ORDEM	CARACTERÍSTICAS DO CHEFE DA FAMÍLIA	PESSOAS	FAMÍLIAS		
			Total	Número de Componentes	
				1 Pessoa	2 Pessoas
20	3 anos	12 974 597	2 772 152	120 325	390 262
21	4 anos	26 672 361	6 168 830	254 004	971 602
22	5 anos	1 944 593	489 409	21 424	91 380
23	6 anos	1 994 487	515 303	22 505	99 431
24	7 anos	1 918 197	506 189	23 709	102 210
25	8 anos	5 334 182	1 384 359	66 985	266 066
26	9 anos	608 045	169 443	8 727	39 148
27	10 anos	947 409	258 354	13 194	55 442
28	11 anos	4 795 126	1 280 591	70 606	256 534
29	12 anos	343 274	97 885	6 166	22 556
30	13 anos	426 053	120 401	7 535	27 974
31	14 anos	583 868	161 953	11 162	35 332
32	15 anos	1 904 512	497 465	32 359	88 816
33	16 anos	1 240 882	306 533	16 479	51 264
34	17 anos ou mais	584 383	146 755	10 956	25 328
35	Anos de estudo não determinados	96 949	24 278	1 801	4 418
36	Sem declaração	–	–	–	–
	SETOR DE ATIVIDADE (1)				
37	Atividades agropecuárias, de extração vegetal e pesca	36 535 894	7 162 467	266 551	872 307
38	Indústria de transformação	15 362 200	3 538 967	106 004	507 559
39	Indústria da construção	9 728 202	2 032 935	64 493	240 989
40	Outras atividades industriais	2 149 065	455 410	15 899	52 431
41	Comércio de mercadorias	9 058 372	2 036 114	78 942	282 985
42	Transportes e comunicações	5 680 617	1 259 141	28 840	153 774
43	Prestação de serviços	10 655 777	2 631 640	181 572	484 403
44	Atividades sociais	3 544 568	887 102	68 292	168 854
45	Administração pública	5 088 168	1 095 435	37 583	134 015
46	Outras atividades	2 257 058	569 890	36 893	101 513
	POSIÇÃO NA OCUPAÇÃO (1)				
47	Empregado	57 408 560	13 023 830	580 007	1 947 282
48	Autônomo	37 669 148	7 570 482	267 559	914 654
49	Empregador	4 347 851	911 763	20 037	99 566
50	Não remunerado	121 361	36 080	2 565	12 625
51	Sem declaração	513 001	126 946	14 921	24 703
	RENDIMENTO MÉDIO MENSAL (SALÁRIO MÍNIMO)				
52	Até 1/2	10 306 240	2 695 274	359 221	636 789
53	Mais de 1/2 a 1	22 059 407	4 927 147	371 698	877 100
54	Mais de 1 a 2	31 784 758	6 874 391	336 256	1 045 654
55	Mais de 2 a 3	16 330 869	3 585 174	134 907	501 856
56	Mais de 3 a 5	15 014 604	3 388 463	115 857	475 797
57	Mais de 5 a 10	10 283 511	2 400 373	88 297	347 429
58	Mais de 10 a 20	4 932 612	1 155 936	44 463	161 480
59	Mais de 20	2 715 624	603 559	20 985	71 796
60	Sem rendimento	4 066 921	1 042 852	58 982	291 038
61	Sem declaração	465 755	133 579	24 306	31 261

(1) Famílias e pessoas cujos chefes tinham 15 anos ou mais e que trabalharam no ano de referência.

TABELA 5
FAMÍLIAS E PESSOAS RESIDENTES EM DOMICÍLIOS PARTICULARES, POR NÚMERO DE COMPONENTES, SEGUNDO ALGUMAS DAS PRINCIPAIS CARACTERÍSTICAS DO CHEFE DA FAMÍLIA

(conclusão)

FAMÍLIAS							NÚMERO DE ORDEM
Número de componentes							
3 Pessoas	4 Pessoas	5 Pessoas	6 Pessoas	7 a 10 Pessoas	11 a 14 Pessoas	15 Pessoas ou mais	
481 973	497 758	416 419	300 917	504 579	57 657	2 262	20
1 254 645	1 267 363	935 193	590 566	812 538	79 645	3 274	21
115 357	103 379	70 228	39 682	44 086	3 728	145	22
124 448	114 012	73 239	39 418	39 326	2 791	133	23
127 509	109 356	68 990	36 214	35 575	2 530	96	24
330 287	306 311	198 682	105 279	103 386	7 032	331	25
43 078	35 330	21 324	10 195	9 082	534	25	26
65 757	56 418	34 918	16 910	14 861	801	53	27
296 552	288 604	191 772	95 761	77 350	3 296	116	28
24 557	21 792	12 881	5 914	3 895	136	8	29
29 039	26 281	16 485	7 738	5 156	193	–	30
36 501	35 357	24 100	11 621	7 611	261	8	31
98 856	116 123	86 364	43 726	30 385	810	26	32
54 886	66 415	57 979	33 180	25 592	701	37	33
25 567	30 166	25 767	15 830	12 746	373	22	34
4 551	4 842	3 893	2 221	2 379	162	11	35
–	–	–	–	–	–	–	36
1 127 792	1 132 040	1 022 655	797 614	1 686 660	246 660	10 188	37
759 147	788 029	559 232	334 716	440 711	42 098	1 471	38
368 669	386 435	316 715	228 339	384 691	41 316	1 288	39
82 117	91 602	74 356	50 187	80 012	8 379	427	40
395 271	425 667	326 961	208 977	287 202	28 829	1 280	41
255 750	280 695	214 617	133 897	175 149	15 787	632	42
529 599	512 217	375 163	231 328	291 739	24 666	953	43
174 176	169 342	125 975	77 690	94 162	8 206	405	44
191 948	227 124	165 814	124 297	174 901	18 942	811	45
117 707	121 132	89 385	48 894	50 674	3 574	118	46
2 587 897	2 605 366	1 957 064	1 262 038	1 879 367	197 734	7 075	47
1 234 846	1 310 314	1 142 196	849 033	1 619 923	222 482	9 475	48
147 740	193 737	173 953	113 204	146 775	15 829	922	49
8 837	4 860	2 751	1 541	2 609	299	13	50
22 856	20 006	14 909	10 123	17 227	2 113	88	51
487 738	360 495	272 666	193 401	343 278	40 370	1 316	52
858 287	737 751	610 302	458 095	896 937	113 024	3 953	53
1 284 303	1 167 572	936 721	660 955	1 260 053	156 683	6 194	54
710 161	705 565	531 807	355 605	572 243	70 019	3 011	55
678 148	731 319	540 811	334 393	459 638	50 087	2 413	56
470 540	558 973	409 098	234 548	265 117	25 086	1 285	57
207 967	274 573	219 635	122 572	116 355	8 436	455	58
92 167	131 934	127 911	80 006	74 128	4 368	264	59
214 930	153 220	110 989	75 255	123 819	14 075	544	60
24 000	18 543	13 032	8 365	12 674	1 347	51	61

FONTE – IBGE, Censo Demográfico – Brasil – 1980.

b) situação do domicílio por constituir um estrato geográfico natural que influi nos resultados;

c) sexo por ser uma variável importante e imperativa;

d) condição de presença pela influência em certas características próprias, e em particular devido à baixa freqüência de não moradores, especialmente em domicílios coletivos e pelo tipo de população, residente ou presente;

e) chefe do domicílio — chefes da família, dada a importância do estudo de suas características; e

f) tamanho do domicílio — para controlar a tendenciosidade de coleta citada anteriormente.

O uso de muitas variáveis de estratificação associado à dimensão da área de ponderação pulveriza a amostra, em termos de pós-estrato, agravando a flutuação de amostragem. Assim, o controle dessas flutuações é realizado através do cálculo dos pesos, usando-se a matriz de ponderação que é constituída pelos pós-estratos formados pelas variáveis de estratificação. Aumentar muito o número de variáveis pode ser prejudicial.

3.10.5 — Consistência

Os resultados do censo são obtidos pela apuração do universo, para variáveis investigadas em todos os domicílios, por estimativas para as demais variáveis que figuram no boletim da amostra. A fim de evitar resultados diferentes entre as apurações do universo e as respectivas estimativas provenientes da amostra, gerando dúvidas nos usuários dos dados, é imprescindível que tais resultados sejam consistentes, isto é, iguais ou praticamente iguais.

A consistência é conseguida através do cálculo do peso para cada unidade da amostragem (pessoa, domicílio, família, etc.), de modo que ele represente a razão entre o número de unidades na matriz do universo e o número de unidades na matriz da amostra. O modo de se alcançar a melhor consistência possível se dá quando os pesos são calculados a nível de cada pós-estrato, isto é, a nível de célula. Como os resultados apresentados geralmente se referem às distribuições marginais da matriz de ponderação, alguns processos de estimação considerados mais modernos apenas se preocupam com a consistência em relação a essas marginais. É oportuno salientar que o processo ideal é aquele que obtém melhor consistência tanto a nível de célula quanto a nível de marginal.

3.10.6 — Nível geográfico de estimação — área de ponderação

O menor nível geográfico de coleta compreende o setor censitário, que em média possui cerca de 900 pessoas, em 1980. A tendenciosidade de coleta ocorre independentemente em cada setor censitário, pois depende das falhas cometidas pelo recenseador. Assim sendo, o setor censitário poderia constituir-se no menor nível geográfico de estimação, com a vantagem de restringir à própria área suas deficiências de coleta.

Se por esse lado a tendenciosidade fica limitada, por outro, as estimativas tornam-se menos consistentes, em face do número de células da matriz de ponderação que necessariamente exigiria reagrupamentos devido às baixas freqüências de unidades de amostragem. As estimativas globais para áreas maiores obtidas pelas somas das estimativas elaboradas como setores censitários seriam menos eficientes.

A área de ponderação constitui-se no menor nível geográfico para o qual é realizada a ponderação, a fim de serem obtidas estimativas para a menor subdivisão administrativa estabelecida nos planos de divulgação. Assim, a área de ponderação pode ser o município, o distrito, a região administrativa, o subdistrito, etc., desde que contenha um número suficiente de unidades na amostra. Como os municípios variam substancialmente de tamanho, torna-se difícil estabelecer qual seria esse número suficiente, pois os municípios pequenos são eles próprios áreas de ponderação.

Em 1980 havia 666 municípios com menos de 2.000 habitantes. As estimativas elaboradas para esses municípios são naturalmente menos eficientes do que as elaboradas para os municípios maiores. A solução para essa questão seria: não se fazer amostragem nos municípios pequenos, ou então aplicar uma amostra maior.

Como há muitas solicitações de divulgação de resultados para níveis geográficos menores do que o município, especialmente para os municípios mais populosos e mais importantes, seria interessante, por ocasião do planejamento do censo, definir esses níveis geográficos, para representarem as áreas de ponderação, desde que as estimativas não sejam prejudicadas. Isto já foi feito em parte no Censo Demográfico brasileiro de 1980.

3.10.7 — Estratégia de ponderação

No cálculo do peso, surgem situações em que ele não pode ser obtido ou então o seu resultado se distancia muito do inverso da fração de amostragem, isto é $\frac{1}{f}$. A fim de contornar essas situações, a estratégia da ponderação prevê critérios restritivos e regras de reagrupamento de modo tal que o cálculo da ponderação é realizado após satisfazer a estratégia estabelecida.

Os critérios restritivos variam conforme o processo de estimação adotado e em função da fração de amostragem usada, como foi visto anteriormente.

Teoricamente, em cada pós-estrato (célula (i,j)), o peso calculado é igual à razão $w_{ij} = \frac{\tilde{N}_{ij}}{n_{ij}}$, ou seja, é igual ao resultado da divisão do total de pessoas no universo pelo total de pessoas na amostra, e é aproximadamente igual a $\frac{1}{f}$, inverso da fração de amostragem.

O tamanho da área de ponderação, o detalhamento da pós-estratificação, as tendenciosidades de coleta e as flutuações de amostragem fazem com que os valores de w_{ij} variem em torno de $\frac{1}{f}$; geralmente quanto menor for N_{ij} maior a possibilidade de w_{ij} variar e se distanciar de $\frac{1}{f}$. Por esse motivo, são fixados os critérios restritivos.

Os critérios restritivos do Censo Brasileiro de 1980 foram estabelecidos a partir dos estudos contidos em (21), onde foram fixados os valores $n_{ij} \geqslant 5$ e $1 \leqslant w_{ij} \leqslant 8$. Neste Censo, a restrição foi levada para as marginais, conforme o processo RREP. Assim sendo, as restrições foram: $n_{i.} \geqslant 5$ e $n_{.j} \geqslant 5$ e $1 \leqslant w_{i.} \leqslant 8$ e $1 \leqslant w_{.j} \leqslant 8$.

Quando os valores de $n_{i.}$ e $n_{.j}$ ou de $w_{i.}$ e $w_{.j}$ para uma cédula da marginal não satisfizerem o critério restritivo, esta seria agrupada a outra, seguindo as regras de reagrupamento, até serem satisfeitas as condições ou se esgotarem as possibilidades.

As regras de reagrupamento utilizadas (19) geralmente procuram preservar os perfis das distribuições de freqüências das variáveis que definem os estratos, evitando-se, salvo em última circunstância, reagrupamento de estratos diferentes, por exemplo, pessoas moradoras com não-moradoras, porque as características dessas pessoas são diferentes, pois são estratos distintos.

3.10.8 — Uso de duas ou mais matrizes de ponderação

Nos censos demográficos geralmente são elaboradas tabulações referentes às características das pessoas, dos domicílios e das famílias. Para cada um desses conjuntos de unidades, são calculados pesos que possibilitarão as estimativas finais.

O peso a ser adotado para cada pessoa, domicílio e família, pode ser calculado a partir de uma única matriz de ponderação, quando entre as variáveis de estratificação figuram informações referentes aos chefes de domicílio que, em mais de 95% das vezes, é o próprio chefe da família. Isto foi feito no Censo Brasileiro de 1970.

Algumas vezes são utilizadas matrizes de ponderação distintas, isto é, uma para as pessoas, outra para os domicílios e outra para as famílias. Em tais situações é necessário que as variáveis de estratificação de cada uma dessas matrizes sejam relevantes, isto é, realmente identifiquem estratos com características distintas ou então em situações extremas.

O Censo Escolar do Espírito Santo em 1977 utilizou duas matrizes de ponderação: uma para as pessoas e outra para os domicílios e famílias, a fim de controlar as tendenciosidades de coleta. Os resultados obtidos apresentaram alguns conflitos de consistência que infelizmente não foram possíveis de serem contornados devido a problemas de custo e tempo hábil de processamento.

No Censo Demográfico de 1980, foram utilizadas duas matrizes de ponderação, uma para as características de pessoas,

também utilizada para as características de famílias, e outra apenas para as características de domicílios, esta muito mais detalhada, relativamente, do que a matriz para pessoas. Os resultados desse Censo referentes aos domicílios e às famílias apresentam algumas inconsistências devido ao uso de duas matrizes de ponderação.

É fundamental, ao serem utilizadas matrizes distintas de ponderação, que elas possibilitem resultados consistentes a fim de inspirarem maior credibilidade junto aos usuários.

3.11 – Cálculo e Apresentação dos Erros

Nos censos por amostragem ocorrem duas espécies de erros, o erro alheio à amostragem e os de correntes do uso de amostragem. Erros do primeiro tipo são mais difíceis de avaliar e controlar, enquanto os do segundo tipo dependem dos procedimentos de amostragem adotados.

Alguns erros alheios à amostragem podem ser avaliados através de análises demográficas ou procedendo-se comparações com outras fontes externas de dados que inspirem confiança. Entretanto, os erros de cobertura são geralmente avaliados por meio de Pesquisas de Avaliação ou Pós-enumeração, realizadas por amostragem, de modo que as estimativas das taxas de erro são calculadas por processos de amostragem ou então pelo método "Dual System".

Nos programas de trabalho dos censos dos EUA constam, como parte integrante das tarefas censitárias, vários estudos e pesquisas visando a conhecer e medir a intensidade das tendenciosidades do censo.

No Brasil o primeiro estudo objetivo reporta a 1967, por ocasião do censo do município de Presidente Prudente. No Censo de 1970 foram planejados, executados e concluídos todos os trabalhos da Pesquisa de Avaliação, os quais foram apurados manualmente. No censo experimental de Taubaté, realizado em 1979, novamente foi planejada uma Pesquisa de Avaliação, cujos resultados ficaram prontos antes da realização da coleta do Censo de 1980. Por último, foi planejada e executada a Pesquisa de Avaliação de 1980, cujos trabalhos de apuração estão inconclusos não por falta de recursos técnicos, mas sim por falta de decisão superior. Os resultados disponíveis provêm das apurações preliminares que possibilitam avaliar apenas as taxas de omissão da coleta de domicílios e seus moradores, cujos procedimentos e metodologia de cálculo constam de vários documentos.

Outros países também realizam pesquisas de avaliação, pois não há censo sem erro. No entanto, não basta promover pesquisas de avaliação; é necessário e indispensável que as tarefas a elas pertinentes sejam acompanhadas e cobradas até o seu término, como parte integrante do censo. Pior do que avaliar os erros, com algumas limitações, é não fazer nada ou então acreditar que o censo é isento de erro, pura utopia.

Os erros de amostragem são calculados com mais precisão utilizando-se os estimadores correspondentes aos processos de estimação de totais, médias e proporções, adotados nos trabalhos censitários. Assim foi realizado o cálculo inicial dos erros de amostragem referentes ao Censo de 1980.

Além da fórmula própria e inerente ao processo de expansão utilizado no censo, outros processos indiretos de cálculo dos erros podem ser utilizados. Dentre eles, destacam-se (13): a fórmula da variância de uma amostra pós-estratificada; a fórmula da variância da estratificação; o método dos grupos aleatórios; e a fórmula do estimador natural, supondo o esquema de uma amostra aleatória simples.

De acordo com os resultados do estudo experimental do censo de Taubaté (13),os erros de amostragem, expressos através do coeficiente de variação, calculados pelo processo indireto dos grupos aleatórios, geralmente superestimam a variabilidade, devido à utilização da estimativa total, nesses grupos de controle. As estimativas do coeficiente de variação obtidas através da pós-estratificação e da estratificação, são aproximadamente da mesma ordem de grandeza e substancialmente menores do que as estimativas obtidas através do método dos grupos aleatórios, devido ao benefício do uso dos pós-estratos. Para as características que não figuram nos pós-estratos, as dife-

renças entre os coeficientes de variação desses três processos são bem menores.

Em 1970, uma indicação da ordem de grandeza dos coeficientes de variação para qualquer variável foi obtida através do cálculo seguindo o esquema de uma amostra aleatória simples. Em 1980, conforme já foi dito, a estimativa dos coeficientes de variação foi realizada utilizando fórmula própria, inicialmente. Posteriormente, os valores calculados para os municípios foram ajustados através de um modelo de regressão, cujos resultados foram utilizados para fornecer uma indicação aproximada dos coeficientes de variação associados às estimativas de qualquer variável. A comparação entre as ordens de grandeza dos coeficientes de variação adotados em 1970 e 1980, para os diversos tamanhos das estimativas, são desprezíveis, isto é, ambos os processos fornecem a mesma precisão.

Considerando os custos de obtenção das estimativas dos erros de amostragem, segundo os diferentes processos e as suas respectivas precisões, sob um ponto de vista de uma utilização prática, o processo adotado em 1970 é preferível, pois o seu custo é desprezível.

No Censo de População de 1980 dos EUA, os cálculos dos erros de amostragem foram obtidos de modo similar ao adotado no Brasil no Censo de 1970, para propósitos gerais, e para outros são apresentados alguns fatores de ajustamento.

O Censo de População da Argentina em 1980, utilizou na maior parte do País uma amostra de áreas (segmentos). Com o intuito de reduzir os trabalhos de computação necessários para estimar os erros da amostragem, foi utilizado inicialmente o método de Keyfitz (10) para cada área de ponderação. Esse método consiste, em linhas gerais, em calcular a variabilidade entre pares de segmentos sobre todos os pós-estratos, considerando o par de segmentos como os segmentos selecionados de um pseudo-estrato em que a área de ponderação foi subdividida. Posteriomente, foram calculados os efeitos do desenho (Deff) para oito áreas distintas em que o País foi dividido, isto devido ao esquema de amostragem adotado neste censo ser o de amostragem de conglomerados.

Todos os Censos Brasileiros já realizados por amostragem não fazem referências aos erros mais graves, que são os erros alheios à amostragem. Estes, para a dimensão de cada município, estado, etc., são bem superiores aos erros de amostragem, isto com base nas Pesquisas de Avaliação de 1970 e 1980. Assim sendo, a utilização de um processo que indique a ordem de grandeza dos erros de amostragem é bastante satisfatória para a utilização dos resultados dos censos.

A apresentação dos erros de amostragem é feita de modo relativo, ou seja, através do coeficiente de variação, ou então de modo absoluto usando o erro absoluto. Como é impraticável calcular erros de amostragem para todas as variáveis e para todas as células dos cruzamentos de uma tabela, geralmente são estimados os erros para diferentes tamanhos de estimativa e para as variáveis principais, reunindo os resultados em tabelas (20).

Um outro aspecto a considerar e explicitar é o nível de confiança da estimativa do erro obtido, isto é, se 68% ou 95%. Os resultados dos *erros absolutos* do Censo de 1970 foram expressos para um nível de confiança de 95%, enquanto os erros relativos apresentados no Censo de 1980 devem corresponder a 68%. A adoção de um ou outro nível de confiança é válida, o importante é que se diga qual o nível de confiança adotado. Sob o ponto de vista prático, o usual é calcular os erros de amostragem para o nível de confiança de 68%, pois a partir deste se obtêm os erros correspondentes aos demais níveis.

BIBLIOGRAFIA

1 – ANÁLISE preliminar sobre erros da amostragem e alheios à amostragem, ocorridos no Censo Demográfico de 1970. Rio de Janeiro, IBGE, 1975. (Estudo elaborado por Luiz Nery da Costa).

2 – ARORA, H. R.; BRACKSTONE, G. J. An investigation of the properties ofraking ratio estimatiors with simple randon sampling (Census Survey Methods Division), Canada Statistics, 1977.

3 – _____;_____. An investigation of the properties of raking ration estimators with simple randon sampling (Census Survey Methods Division), Canada Statistics, 1978. (16)
4 – _____;_____. Weithting procedures of raking ration estimation (Census Survey Methods Division), Canada Statistics, 1978.
5 – BRACKSTONE, G. J. The reliability and consistency of census data (Census Survey Methods Division), Canada Statistics, 1976.
6 – _____; RAO, J. N. k. Raking Ratio Estimators. Canada Statistics, 1975.
7 – CENSO Demográfico Experimental de 1980. Pesquisa de avaliação da precisão da coleta. Rio de Janeiro, IBGE, agosto/1980.
8 – CENSO Escolar de 1977 do Estado do Espírito Santo. Governo do Estado do Espírito Santo, 1979.
9 – CENSUS of Canada, Selected Observations — 1981 — Dr. Edward Ptyor. Census Manager, Social Statistics Field Statistics Canada.
10 – DISENO de La Muestra del Censo Nacional de Poblacion y Vivienda de 1980 — Argentina: Procedimentos de Selecion, Estimacion, Calculo de Efectos de Diseno y Presentacion de Errores de Muestreo-Estadistica vol. XXXVII — Avance n.° 1, Santiago — Chile — júlio 1985.
11 – ESTUDOS sobre a Avaliação da Precisão da Coleta e de Algumas Tendenciosidades do Censo Demográfico de 1970. Rio de Janeiro, IBGE, 1975. (Inédito).
12 – ESTUDOS sobre a Avaliação da Precisão da Coleta e de Algumas Tendenciosidades do Censo Demográfico de 1970. II Resultados definitivos — Brasil. (Inédito).
13 – EXPANSÃO da Amostra do Censo Experimental de 1980. Rio de Janeiro, IBGE, abril/1982. (Estudos Experimentais).
14 – LOS METODOS de Muestreo y Los Censos. ONU/FAO, 1967.
15 – METODOLOGIA do Censo Demográfico de 1980. Rio de Janeiro, IBGE, 1983.
16 – NATIONAL Report prepared by Canada for the XV Session of Coins.
17 – PROPOSIÇÕES Iniciais para a Investigação de Itens por Amostragem no Censo Demográfico de 1980 — versão preliminar. Rio de Janeiro, IBGE 1979/80. (Documento elaborado por Luiz Nery da Costa).
18 – U. S. BUREAU of the Census. Censuses of Population and Housing — Procedural History — 1960.
19 – _____ .Censuses of Population and Housing — 1979.
20 – _____ .Standards for Discussion and Presentation of Errors in Data. Washington, D. C., 1974. (Technical paper, 32).
21 – USO do Processo de Estimativa de Razão no Censo Demográfico. Rio de Janeiro, IBGE, 1978. Elaborado por Luiz Nery da Costa. (Inédito).

RESUMO

Os Censos Demográficos brasileiros de 1960, 1970 e 1980 foram realizados com a aplicação da amostragem na coleta. Os procedimentos metodológicos usados nesses censos guardam semelhanças com os censos de população dos EUA, Canadá e Argentina.

O preparo das presentes notas metodológicas sobre o planejamento de Censos Demográficos por amostragem e a comparação com as metodologias seguidas por aqueles países visam a subsidiar o entendimento da matéria pelos estatísticos e técnicos que utilizam dados demográficos e que desejam começar ou a ampliar seus conhecimentos nessa atividade.

ABSTRACT

The Brazilian Demographic Census of 1960, 1970 and 1980 were carried out by data sampling; the methodological procedures used were similar to those of the U.S.A, Canada and Argentina.

The preparation of these notes on planning the Brazilian Demographic Census by sampling and the comparison with the methods followed by other countries aim at providing the statisticians and technicians who use such demographic data and wish to either begin or broaden their knowledge in this field of activity with the basic information needed to start their studies.

O USO DA TÁBUA DE MORTALIDADE EM ESTUDOS DE SEGUIMENTO

David Dorigo*

1 – INTRODUÇÃO

Um método muito usado para descrever dados de seguimento é a tábua de mortalidade. Historicamente, após um longo tempo relegado ao esquecimento, este método foi retomado a partir de 1950 (Berkson e Gage; ver referências) e, atualmente, vem sendo extensivamente utilizado, principalmente em experimentos clínicos com dados incompletos. Uma descrição da tábua de mortalidade e seus parâmetros principais encontra-se em Gehan (1969). Para maiores detalhes ver Johnson e Johnson (1980), especialmente, Capítulos 4 e 6. Inicialmente usada para descrever a mortalidade de populações humanas, esta técnica é útil para descrever outros tipos de dados de seguimento. Por exemplo, pode-se descrever a duração eventual da doença de um paciente sofrendo de sífilis, como é o caso do presente estudo, onde por "nascimento" entende-se o instante em que, diagnosticada a doença, ele entra no estudo e por "morte" o instante em que recebe alta, quando é considerado curado. Uma vantagem adicional em usar esta técnica é que, na prática, os pacientes dificilmente passam a participar de um estudo de seguimento num mesmo instante. Além disso, alguns deles, por razões várias, retiram-se antes da data prevista para término do estudo. Contudo, supondo que nenhum tratamento ou características dos pacientes se altere, substancialmente, durante o seguimento, tudo acontece como se todos eles passassem a participar do estudo num mesmo instante. Assim, os dados referentes àqueles que se perderam durante o estudo podem ser usados até a data em que eles se retiraram.

Na Seção 2 é apresentada uma descrição sucinta do experimento; na Seção 3 são analisados os dados e, finalmente, na Seção 4 são feitas algumas conclusões.

2 – O ESTUDO DE SEGUIMENTO

Este estudo contou com a participação de 284 pacientes encaminhados ao Ambulatório de Medicina Preventiva da Faculdade de Medicina de Ribeirão Preto, USP, e provenientes das seguintes áreas: a) Município de Ribeirão Preto, b) Sexta Divisão Regional de Saúde (excluindo o município), c) Estado de São Paulo (excluindo a Sexta DRS) e d) Outros Estados.

* Do Instituto de Matemática da Universidade Federal do Rio de Janeiro.
R. bras. Estat., Rio de Janeiro, *48* (189/190): 65-69, jan./dez., 1987

Uma vez que o paciente era escolhido para participar do estudo, as seguintes informações, entre outras, eram anotadas: idade, sexo, ocupação, procedência, situação quanto à evolução ambiental e número de exames de Wassermann.

3 – DESCRIÇÃO E ANÁLISE DOS DADOS

Dos 284 pacientes que constituíram a amostra, 265 (93,3%) estavam compreendidos na faixa etária 15-45 anos, sendo a idade média de 31 anos, aproximadamente (ver Tabela 1).

TABELA 1

DISTRIBUIÇÃO ETÁRIA DOS PACIENTES

Idade (anos)	Freqüência	%
10 – 15	4	1,4
15 – 25	98	34,5
25 – 35	99	34,9
35 – 45	42	14,8
45 – 55	26	9,2
55 – 65	12	4,2
65 – 75	2	0,7
75 +	1	0,4

Do total, 196 (69%) eram do sexo feminino; 198 (69,7%) eram residentes no Município de Ribeirão Preto e outros 70 (24,6%) da área abrangida pela Sexta DRS, ou seja, apenas 16 pacientes não eram residentes dessa área. Este fato é relevante no sentido de que, ao interpretar os dados, não se deve tirar conclusões definitivas sobre a evolução de sífilis para um grupo, aparentemente, tão homogêneo de indivíduos.

À época do levantamento dos dados, feito 96 meses após o início do seguimento (1975), a situação dos pacientes quanto à evolução do seguimento ambulatorial era a seguinte: (ver Tabela 2).

As três primeiras classes, compreendendo 157 pacientes, foram consideradas para efeito de análise como observações incompletas; a última compreende aqueles considerados curados.

Usando a notação costumeira, tal como descrita em Gehan (1969), as seguintes

TABELA 2

DISTRIBUIÇÃO DOS PACIENTES PELA INCIDÊNCIA de Wassermann

Situação dos pacientes	N.° de pacientes
1) em tratamento.....	43
2) abandono com título de wassermann menor que 5	25
3) abandono com título de Wassermann maior que 5.....	89
4) alta médica ou abandono com 1, 2 ou 3 resultados de Wassermann negativos....	127

quantidades são utilizadas na construção da tabela de mortalidade:

h_i = amplitude do i.° intervalo, tendo sido considerados intervalos de 3 meses;

n_i – n.° de pacientes no início do i.° intervalo;

w_i = n.° de pacientes com sífilis e que se retiraram do estudo durante o i.° intervalo;

$n_i = n_i' - \frac{1}{2} w_i =$ n.° de pacientes efetivamente expostos ao tratamento no i.° intervalo;

d_i = n.° de pacientes curados (incluindo os abandonos com 1, 2 ou 3 Wassermann negativos);

$\hat{q}_i = \frac{d_i}{n_i} =$ proporção de curados (incluindo os abandonos com 1, 2 ou 3 Wassermann negativos) no i.° intervalo;

$\hat{p}_i = 1 - \hat{q}_i$ = proporção de afetados pela sífilis no i.° intervalo;

$\hat{P}_i = \hat{p}_i \hat{P}_{i-1}$ = proporção acumulada de doentes até o i.° intervalo (com $\hat{P}_1 = 1$);

$\hat{\lambda}_i = \frac{2\hat{q}_i}{h_i(1 + \hat{p}_i)}$ = taxa de recuperação por intervalo. É uma estimativa da probabilidade de um paciente curar-se, dado que está afetado no início do intervalo;

$\tilde{t}_i = (t_j - t_i) + \frac{(P_j - \frac{\hat{P}_i}{2})h_j}{\hat{P}_j - \hat{P}_{j+1}}$ = mediana do tempo restante da doença no i.° intervalo ($j \neq i$).

A Tabela 3 abaixo mostra os valores obtidos para essas quantidades, calculadas pelo método atuarial, e a Tabela 3a os desvios padrão das correspondentes estimativas, cujas fórmulas de cálculos se encontram, por exemplo, em Gehan.

TABELA 3
ESTIMATIVAS DOS PRINCIPAIS PARÂMETROS (DADOS DE SÍFILIS)

n_i	w_i	n_i	d_i	$\hat{q}_i$	$\hat{p}_i$	$\hat{P}_i$	$\hat{\lambda}_i$	$\tilde{t}_i$
284	29	269,5	10	0,0371	0,9629	1,0000	0,0095	47,37
245	16	237,0	9	0,0380	0,9620	0,9629	0,0129	46,07
220	12	214,0	9	0,0421	0,9579	0,9263	0,0143	45,10
199	9	194,5	12	0,0617	0,9383	0,8874	0,0212	43,67
178	3	176,5	9	0,0510	0,9490	0,8326	0,0174	42,48
166	11	160,5	11	0,0685	0,9315	0,7902	0,0237	40,40
144	2	143,0	7	0,0490	0,9510	0,7360	0,0167	38,59
135	7	131,5	11	0,0837	0,9163	0,7000	0,0291	36,84
117	4	115,0	2	0,0174	0,9826	0,6414	0,0058	39,56
111	3	109,5	6	0,0548	0,9452	0,6303	0,0188	38,04
102	1	101,5	2	0,0197	0,9803	0,5957	0,0066	39,59
99	6	96,0	3	0,0313	0,9688	0,5840	0,0106	37,04
90	8	86,0	1	0,0116	0,9884	0,5657	0,0039	34,74
81	7	77,5	5	0,0645	0,9355	0,5592	0,0222	31,99
69	3	67,5	2	0,0296	0,9704	0,5231	0,0100	31,03
64	6	61,0	2	0,0328	0,9672	0,5076	0,0111	29,65
56	1	55,5	3	0,0541	0,9459	0,4910	0,0185	29,29
52	4	50,0	4	0,0800	0,9200	0,4644	0,0278	30,17
44	1	43,5	7	0,1609	0,8391	0,4273	0,0583	32,19
36	1	35,5	3	0,0845	0,9155	0,3585	0,0294	37,54
32	3	30,5	0	0,0164	0,9836	0,3282	0,0055	36,00
29	1	28,5	1	0,0351	0,9649	0,3228	0,0119	33,00
27	2	26,0	0	0,0192	0,9808	0,3115	0,0065	30,00
25	3	23,5	3	0,1277	0,8723	0,3055	0,0455	27,00
19	1	18,5	1	0,0541	0,9459	0,2665	0,0185	24,00
17	5	14,5	0	0,0345	0,9655	0,2521	0,0117	21,00
12	1	11,5	0	0,0435	0,9565	0,2434	0,0148	18,00
11	1	10,5	0	0,0476	0,9524	0,2328	0,0163	15,00
10	0	10,0	0	0,0500	0,9500	0,2217	0,0171	12,00
10	0	10,0	0	0,0500	0,9500	0,2107	0,0171	9,00
10	1	9,5	0	0,0500	0,9500	0,2001	0,0171	6,00
9	0	9,0	1	0,1111	0,8889	0,1901	0,0392	3,00
8	5	5,5	3	0,5455	0,4545	0,1690	–	–

TABELA 3a

DESVIO PADRÃO DE $\hat{P}_i$, $\tilde{\lambda}_i$ E $\tilde{t}_i$
(DADOS DE SÍFILIS)

$s(\hat{P}_i)$	$s(\tilde{\lambda}_i)$	$s(\tilde{t}_i)$
0,0000	0,0030	5,49
0,0115	0,0043	3,53
0,0163	0,0048	2,56
0,0201	0,0061	2,57
0,0243	0,0058	1,37
0,0269	0,0071	1,36
0,0296	0,0063	1,34
0,0311	0,0088	3,02
0,0331	0,0041	7,92
0,0335	0,0077	7,98
0,0345	0,0047	2,27
0,0348	0,0061	2,29
0,0353	0,0039	2,35
0,0355	0,0099	2,44
0,0367	0,0071	6,63
0,0372	0,0079	6,77
0,0378	0,0107	-----
0,0387	0,0139	8,89
0,0398	0,0220	8,76
0,0410	0,0170	4,27
0,0411	0,0078	0,00
0,0411	0,0119	0,00
0,0412	0,0092	0,00
0,0413	0,0262	0,00
0,0417	0,0185	0,00
0,0419	0,0165	0,00
0,0422	0,0209	0,00
0,0429	0,0230	0,00
0,0437	0,0242	0,00
0,0442	0,0242	0,00
0,0444	0,0248	0,00
0,0445	0,0391	0,00
0,0443	-----	-----

Note-se na Tabela 3 que a taxa de recuperação, à exceção de alguns pontos, está praticamente estável entre 0,005 e 0,030, sugerindo para os dados o modelo exponencial. A Figura 1 mostra que o ajuste feito para a proporção acumulada de doentes, supondo esse modelo, é bastante razoável.

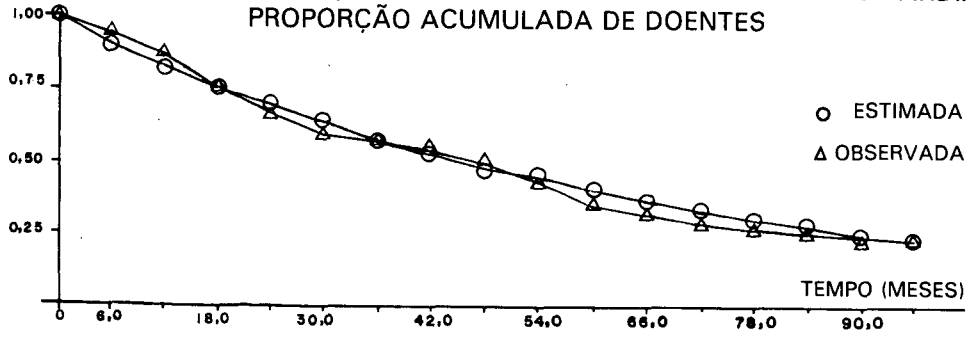
4 – CONCLUSÕES

Esta conclusão é apenas inicial. Uma escolha final de modelo deve ser feita, de preferência, levando-se em conta vários conjuntos de dados semelhantes. Além disso, variáveis explicativas tais como idade, sexo, número de tratamentos completos, número de familiares com Wassermann positivo ajudam a uma melhor inferência a respeito da evolução da sífilis latente tardia numa população. Por outro lado, uma idéia de como a sífilis afeta a vida média de um indivíduo pode ser obtida através da comparação entre a proporção acumulada de sobreviventes numa tábua de mortalidade de uma certa população, cuja idade inicial seja a da média dos pacientes em estudo, e a proporção acumulada tal como calculada na Tabela 1.

Com relação a variáveis explicativas, o modelo clássico de D. Cox (1972) pode ser usado para uma interpretação mais abrangente. O problema de dados censurados, muito comum em estudos de seguimento, não constitui problema, desde que as variáveis de interesse sejam independentes do mecanismo de censura. Caso contrário, a análise dos dados deve ser tratada no contexto de risco competitivo, por meio de tabelas de decremento múltiplo (ver, por exemplo, Johnson e Johnson, Cap. 10).

FIGURA 1

ESTUDO DE RECUPERAÇÃO DE PACIENTES COM SÍFILIS LATENTE TARDIA
PROPORÇÃO ACUMULADA DE DOENTES



BIBLIOGRAFIA

BERKSON, J.; GAGE, R. *Calculation of Survival Rates for Cancer.* Proc. Mayo Clin. v. 25, p. 270-86, 1950.

COX, D. *Regression Models and Life Tables.* J. Royal Stat. Soc., B, v. 34, p. 187-220, 1972.

GEHAN, E. *Estimating Survival Functions From the Life Table.* J. Chron. Dis., v. 21, p. 629-44, 1969.

JOHNSON, R. E.; JOHNSON, N. L. *Survival Models and Data Analysis.* Wiley, New York, 1980.

RESUMO

A tábua de mortalidade, como método para descrever dados de seguimento, foi redescoberta a partir da década de 50. De lá para cá, houve um considerável avanço no conhecimento das técnicas e propriedades estatísticas envolvidas. Os resultados formais, aqui utilizados, não são novidade e o objetivo deste trabalho é divulgar um método que interessa a estatísticos, demógrafos, sociólogos e outros. Neste trabalho são apresentados os resultados de um estudo de seguimento para avaliação do tratamento em 284 pacientes com sífilis latente tardia, tendo como método a tábua de mortalidade.

ABSTRACT

"The life table, as a method for describing follow-up studies, has had a revival since the early fifties. In this article the author presents a description of the cohort life table and applies the usual results to a follow-up study involving 284 patients suffering from syphilis. It is suggested that the remission rate has an exponential form. Some general comments are also made regarding censored observations".

O autor deseja manifestar seus agradecimentos ao Dr. Djalma C. Moreira Filho (UNICAMP) pela cessão dos dados para este trabalho e pelas conversas, então, mantidas, e a Francisco Rocha (UFRJ) pelo auxílio computacional.

MULHER BRASILEIRA: MUITO PRAZER

André Cezar Medici*

1 – INTRODUÇÃO

O conhecimento da saúde feminina não se individualiza por ser o resultado de um punhado de especialidades da medicina, da psicologia ou dos estudos sobre sexualidade e reprodução humanas. A inserção social da mulher na estrutura de classes, na ambiência familiar, no mercado de trabalho, na relação cotidiana com os homens e com as próprias mulheres constitui um conjunto de aspectos imprescindíveis na determinação social dos padrões de saúde feminina.

Tais aspectos, no entanto, não se reproduzem homogeneamente, variando segundo os níveis de desenvolvimento sócio-econômico, cultural e político de cada região. Ressalta-se também que, ao longo dos anos, a mulher tem determinado cada vez mais o seu lugar na sociedade, mediante sua inserção em uma série de lutas gerais ou específicas, que têm acarretado fortes mudanças sociais.

Diante deste quadro, é difícil encontrar um só país onde não seja marcante a presença de mulheres na vanguarda dos movimentos sociais organizados. Estes movimentos têm sido decisivos, não só na formação da opinião pública e nas mudanças das regras de comportamento social, como na dinâmica das políticas públicas dos distintos governos nacionais.

As recomendações das Nações Unidas, sobre a mulher, que se apresentam de forma mais marcante nos anos setenta decorrem da importante compreensão destes movimentos, são também, resultantes das preocupações com o rápido crescimento populacional. Neste particular, a mulher detém papel decisivo no que tange ao valor da consciência de sua condição nos aspectos reprodutivos. Os programas de desenvolvimento e política populacional das Nações Unidas voltaram-se, em especial, para os países do terceiro mundo, onde as questões relativas à reprodução humana estavam fortemente correlacionadas com os níveis e padrões, a serem atingidos, de bem-estar social.

Assim, "como desdobramento desta concepção, desencadeia-se, a partir de 1975, a década da mulher. As metas e objetivos a serem atingidos norteavam-se por três termos: igualdade, desenvolvimento e

* Analista Especializado da Fundação Instituto Brasileiro de Geografia e Estatística e, prof. e pesquisador da Escola Nacional de Saúde Pública da FIOCRUZ.

paz. Analisando-se sob o aspecto atenção especial às questões de emprego, saúde e educação. Essas preocupações deveriam ser explicitadas tanto sob o ponto de vista do planejamento global e da legislação, como dos níveis setoriais, no sentido de promover a participação da mulher na sociedade como agente e beneficiária'' (Macedo, 1985).

Nos últimos anos, o setor público brasileiro passou a conceber planos setoriais onde era explícito o estabelecimento de programas relacionados aos aspectos da saúde, trabalho e assistência social à mulher. No entanto, a falta de diagnósticos atualizados e confiáveis sobre a situação social feminina, aliada às dificuldades de ordem operacional, política e financeira têm impedido que tais programas alcancem, sequer, tímidos resultados. A tudo isso, soma-se a grave situação social de homens e mulheres no Brasil, mergulhados no desemprego, na desnutrição e na ausência dos requisitos mínimos de cidadania social.

Apesar do grande acervo técnico e literário hoje disponível sobre o assunto, grande parte do insucesso dos programas sociais, voltados para a condição feminina, decorre do fato de que a mulher brasileira não vinha, até então, sendo objeto de conhecimento e preocupação das autoridades governamentais.

O objetivo deste artigo é contribuir para a atualização das informações disponíveis sobre a condição feminina, afim de que, diante de suas características demográficas e sócio-econômicas possamos conhecê-las um pouco melhor e traçar programas adequados às reais condições de existência.

2 – A MULHER NA POPULAÇÃO BRASILEIRA

Dos 131,4 milhões de brasileiros, estimados para 1985, cerca de 66,3 milhões eram mulheres. Correspondiam, dessa forma, a 50,4% da população do Brasil naquele ano. Apesar de ser maioria absoluta, a presença feminina no contingente demográfico é diferenciada segundo a estrutura etária; em comparação com os homens, pode-se dizer que é maior nas idades mais avançadas e menor nas idades mais tenras. Assim, enquanto 37,4% dos homens, no ano em foco, tinham menos de 15 anos, apenas 35,4% das mulheres inseriam-se naquele grupo etário. A presença feminina era notadamente maior nos intervalos de 15 a 49 anos e de mais de 50 anos, como pode ser observado na Tabela 1.

Por motivos naturais (biológicos) a proporção de nascimentos do sexo masculino é maior que os do sexo feminino. Os estudiosos de demografia costumam dizer que essa proporcionalidade é de 51% para 49%, respectivamente (Sauvy, 1979). Porém, em condições sociais adversas, a mulher é mais resistente à incidência de doenças nos primeiros anos de vida, o que faz com que a mortalidade infantil masculina seja tenden-

TABELA 1
DISTRIBUIÇÃO DA POPULAÇÃO BRASILEIRA POR SEXO E GRANDES GRUPOS DE IDADE – 1985

GRUPOS DE IDADE	DISTRIBUIÇÃO DA POPULAÇÃO BRASILEIRA POR SEXO					
	Total		Homens		Mulheres	
	Absoluto (1 000 pessoas)	%	Absoluto (1 000 pessoas)	%	Absoluto (1 000 pessoas)	%
TOTAL	131 411,4	100,0	65 139,6	100,0	66 271,8	100,0
0 a 14 anos	47 869,7	36,4	24 388,1	37,4	23 481,6	35,4
15 a 49 anos	66 053,7	50,3	32 418,8	49,8	33 634,9	50,8
50 anos e mais (1) ..	17 488,0	13,3	8 332,7	12,8	9 155,3	13,8

FONTE – IBGE. Estimativas fornecidas pela publicação da PNAD-1985.
(1) Inclui 2,7 mil pessoas sem declaração de idade.

cialmente maior do que a feminina, até um determinado intervalo etário.

Em países subdesenvolvidos, — e até mesmo em países desenvolvidos onde é marcante a estratificação social — o peso da mortalidade infantil masculina é decisivo na composição, por gênero, da população. Dessa forma, o número de mulheres tende, necessariamente, à superar o número de homens. Outro fator fundamental na determinação social da hegemonia quantitativa das mulheres na população é a inserção no mercado de trabalho. Por participarem mais diretamente da construção econômica da sociedade, os homens, nos países de baixo patamar de desenvolvimento social, se expõem, cotidianamente, a maior volume de riscos de acidentes, desgaste físico e pressões psico-sociais decorrentes do trabalho. Assim, a esperança de vida do sexo masculino, especialmente nas classes sociais menos favorecidas, tende a ser menor do que a registrada para o sexo feminino, fazendo com que, em média, as mulheres tenham uma trajetória de vida mais extensa do que a dos homens.

A combinação desses fatores de ordem biológica e social pode ser ilustrada através da participação feminina na composição populacional, na medida em que avança a idade. No Brasil, as mulheres representam 48,8% do conjunto de pessoas de 0 a 4 anos de idade. Esta proporção evolui continuamente até chegar a 55,1% do total de pessoas de 70 anos e mais. O grupo etário onde esta composição é mais equilibrada é o da faixa de 15 a 19 anos (Tabela 2).

Os diferenciais de esperança de vida ao nascer entre homens e mulheres vêm-se tornando cada vez maiores. Tomando-se como exemplo o Estado de São Paulo, observa-se que a média, no período 1939-41, era de 44,3 anos para os homens e de 46,7 anos para as mulheres. Apesar das mulheres apresentarem uma possibilidade de viver mais anos, superior à encontrada para os homens, o diferencial era de apenas 2,4 anos. No biênio 1979-81, a esperança de vida masculina estendeu-se para 63,3 anos e a feminina elevou-se para 70,0 anos, naquele mesmo estado. Desta forma, o diferencial mulheres-homens aumentou para 6,7 anos (Tabela 3).

TABELA 2
PERCENTUAL DE MULHERES NA POPULAÇÃO BRASILEIRA, SEGUNDO GRUPOS DE IDADE — 1985

GRUPOS DE IDADE	PERCENTUAL DE MULHERES NA POPULAÇÃO (%)
TOTAL	50,4
0 a 4 anos	48,8
5 a 9 anos	48,9
10 a 14 anos	49,5
15 a 19 anos	50,2
20 a 24 anos	51,1
25 a 29 anos	51,6
30 a 34 anos	50,8
35 a 39 anos	50,9
40 a 44 anos	51,1
45 a 49 anos	50,7
50 a 54 anos	51,5
55 a 59 anos	51,6
60 a 64 anos	50,9
65 a 69 anos	53,1
70 anos e mais	55,1

FONTE — IBGE. Estimativas fornecidas pela publicação da PNAD-1985.

Embora tenha ocorrido um contínuo aumento do trabalho feminino, a taxa de atividade das mulheres no Brasil ainda se encontra em patamares muito baixos. É possível que a elevação em massa do trabalho feminino traga efeitos diretos na redução dos diferenciais de esperança de vida entre homens e mulheres, a exemplo do que ocorre em alguns países desenvolvidos.[1]

Tomando-se, ainda, o caso de São Paulo, verifica-se que a vida média de homens e mulheres aumentou sensivelmente nos últimos 40 anos. A esperança de vida feminina ao nascer apresentou ganhos de 23,3 anos, nesse período, fato que se deve, em grande medida, à redução da mortalidade infantil e aos progressos da medicina social incorporados às práticas da saúde coletiva, bem como às medidas de saneamento básico, es-

[1] Esta relação entre o trabalho feminino e a redução dos diferenciais de mortalidade entre homens e mulheres, deve ser vista com certa reserva. Em geral, a inserção da mulher no mercado de trabalho não ocorre nos mesmos ramos de atividade e nas mesmas ocupações partilhadas pelos homens. A depender da natureza do trabalho, da intensidade, riscos e insalubridade decorrentes do processo de labor, este diferencial poderia ser mantido ou até aumentar.

TABELA 3
ESPERANÇA DE VIDA, EM ANOS, PARA O ESTADO DE SÃO PAULO — 1939/41 — 1979/81

PERÍODOS DE REFERÊNCIA	ESPERANÇA DE VIDA								
	Esperança de vida ao nascer (e 0)			Esperança de vida aos 5 anos (e 5)			Esperança de vida aos 10 anos (e 10)		
	Homens	Mulheres	∇	Homens	Mulheres	∇	Homens	Mulheres	∇
1939/1941	44,29	46,68	2,39	53,81	55,53	1,72	49,74	51,37	1,63
1949/1951	52,75	55,89	3,14	58,08	60,75	2,67	53,66	56,30	2,64
1959/1961	59,04	63,67	4,63	60,76	65,01	4,27	56,11	60,34	4,23
1969/1971	59,32	65,48	6,16	60,87	66,50	5,63	56,13	61,73	5,60
1979/1981	63,30	70,02	6,72	62,61	68,81	6,20	57,78	63,95	6,17

FONTE — Fundação SEADE.

pecialmente no que diz respeito aos sistemas de abastecimento d'água e esgotamento sanitário[2].

Ao lado de tudo isso, o padrão reprodutivo dos brasileiros tem-se alterado continuamente nos últimos 15 anos. A taxa de fecundidade total que se encontrava, em 1970, no patamar de 5,76 filhos por mulher, reduziu-se para 3,53 filhos por mulher em 1984. Esta medida não é homogênea variando segundo as regiões. As áreas rurais ainda têm comportamentos reprodutivos que levam à taxa de fecundidade total a se encontrar no patamar de 5,32, contra os 3,03 encontrados nas áreas urbanas. Da mesma forma, regiões com menor desenvolvimento sócio-econômico, como o Nordeste, apresentam níveis de fecundidade total de 4,96 filhos por mulher contra os 2,96 e 3,04 encontrados nas regiões de maior desenvolvimento como a Sudeste e a Sul (Tabela 4).

Três fatores podem ser apontados como decisivos na rápida redução dos níveis de fecundidade no Brasil. O primeiro é dado pelo aumento vertiginoso das taxas de urbanização. Por estimativa, em 1985, cerca de 72,6% da população brasileira esteja residindo em áreas urbanas. Destes, a maioria (51,2%) era constituída por mulheres, ocorrendo o inverso nas áreas rurais, onde a hegemonia era masculina (Tabela 5). A urbanização, por si só, carrega fortes efeitos na redução do número de filhos, pois "as antigas famílias rurais, que antes tinham seu tamanho vinculado à disponibilidade de terras e a necessidade absoluta de explorá-las, como forma de aumentar a riqueza, trocam de lugar com as famílias urbanas, inseridas na lógica do assalariamento, onde o nível de vida e a renda são determinados, não por elas, mas pelo sistema econômico. Às metrópoles, constituídas por um número crescente de indivíduos, em função basicamente das migrações, criam efeitos psicológicos nas famílias no sentido de planejar um menor número de filhos" Medici e Aguiar (1984).

O segundo são as mudanças de comportamento decorrentes do próprio processo de urbanização, cabendo destacar o aumento dos níveis de inserção da mulher no mercado de trabalho e as dificuldades inerentes a esta inserção, no que tange ao cuidado, manutenção e educação dos filhos[3]. O aumento da escolaridade e a disseminação de informações sobre práticas anticonceptivas também atuam no número programado de filhos.

O terceiro fator é a própria difusão dos meios anticoncepcionais em todas as camadas da população. Este, talvez, esteja sendo,

[2] Os efeitos dos sistemas de saneamento nos padrões de mortalidade têm sido testados em diversas ocasiões. Estudos de Vetter e Simões, apontam que os diferenciais de esperança de vida ao nascer, nos domicílios que detêm sistemas adequados de saneamento, com relação aos que não têm sistemas dessa natureza, alcançam, em média, 12 anos.

[3] A inserção no mercado de trabalho atua num duplo sentido. Se, por um lado, pode inibir a vontade dos casais, em ter um número elevado de filhos, por outro, pode bloquear a própria participação feminina na atividade econômica pela inexistência de aparelhos sociais (tais como creches, escolas maternais, jardins de infância etc.,) hoje inacessíveis à maioria da população brasileira.

TABELA 4
TAXAS DE FECUNDIDADE TOTAL POR SITUAÇÃO DO DOMICÍLIO, SEGUNDO BRASIL E GRANDES REGIÕES — 1970-1984

ANOS E GRANDES REGIÕES	TAXAS DE FECUNDIDADE TOTAL		
	Total (%)	Situação do Domicílio	
		Urbana (%)	Rural (%)
1970			
Brasil	5,76	4,54	7,72
Norte	8,15	6,62	9,59
Nordeste	7,53	6,44	8,45
Sudeste	4,56	3,83	7,14
Sul	5,42	4,06	6,86
Centro-Oeste	6,42	5,31	7,71
1980			
Brasil	4,35	3,63	6,40
Norte	6,45	5,24	8,04
Nordeste	6,13	4,94	7,66
Sudeste	3,45	3,17	5,46
Sul	3,63	3,20	4,55
Centro-Oeste	4,51	3,97	5,98
1984			
Brasil	3,53	3,03	5,32
Norte	4,04	–	–
Nordeste	4,96	4,00	6,47
Sudeste	2,96	2,70	4,99
Sul	3,04	2,79	3,62
Centro-Oeste	3,38	3,06	4,57

FONTE — IBGE. Censos Demográficos e PNAD-1984, Suplemento Especial de Fecundidade.

TABELA 5
PERCENTUAL DE MULHERES NA POPULAÇÃO BRASILEIRA, SEGUNDO A SITUAÇÃO DE DOMICÍLIO — 1985

SITUAÇÃO DE DOMICÍLIO	PERCENTUAL DE MULHERES NA POPULAÇÃO (%)
TOTAL	50,4
URBANO	51,2
RURAL	48,3

FONTE — IBGE. Estimativas fornecidas pela publicação da PNAD-1985.

no presente momento, o mais forte e mais direto fator responsável pela rápida queda nos níveis de fecundidade no Brasil. Os organismos internacionais de política de população e suas correias nacionais de transmissão; os movimentos organizados de mulheres e, ainda que tímidos, os programas governamentais, especialmente os do Ministério da Saúde e o da Previdência têm sido decisivos, tanto pelo papel pedagógico, como por facilitar acesso, aos meios de planejamento familiar, à grande parte das camadas populacionais antes destituídas de qualquer conhecimento sobre o assunto.

3 — DIFERENCIAIS DE ESCOLARIDADE

À primeira vista, as taxas de alfabetização não diferem substancialmente entre homens e mulheres. Para a população de 5 anos e mais, estas taxas situavam-se em 73,4% no ano de 1985, tanto para homens quanto para as mulheres. À medida em que aumenta a idade, começam a surgir diferenciais nas taxas acumuladas de alfabetização, ficando mais elevadas as taxas masculinas. Para exemplificar, a taxa de alfabetização de 7 anos e mais é 77,8% para homens e 77,4% para mulheres; de 20 anos e mais é 79,2% e 75,8% e de 60 anos e mais é 58,2% e 47,4%, respectivamente. Neste último grupo, como pode ser visto, o diferencial supera dez pontos percentuais (Tabela 6).

Que motivo poderia levar as taxas acumuladas de alfabetização femininas a serem paulatinamente menores do que as masculi-

TABELA 6
TAXAS DE ALFABETIZAÇÃO POR SEXO, SEGUNDO A IDADE — BRASIL — 1985

IDADE EM ANOS	TAXAS DE ALFABETIZAÇÃO POR SEXO		
	Total (%)	Homens (%)	Mulheres (%)
5 anos e mais	73,4	73,4	73,4
7 anos e mais	77,6	77,8	77,4
10 anos e mais	79,8	80,3	79,3
15 anos e mais	79,3	80,4	78,3
20 anos e mais	77,4	79,2	75,8
25 anos e mais	74,9	77,3	73,3
30 anos e mais	71,9	75,0	70,3
40 anos e mais	65,5	69,6	61,7
50 anos e mais	59,6	64,7	55,0
60 anos e mais	52,5	58,2	47,4

FONTE — IBGE. PNAD — 1985.
NOTA — A taxa de alfabetização é definida pela divisão entre o total de pessoas alfabetizadas sobre o total de pessoas numa dada faixa etária.

nas? Certamente, o comportamento histórico do padrão de alfabetização no Brasil, segundo o sexo. A Tabela 7 revela que, nas idades mais jovens (até 29 anos) as mulheres são mais alfabetizadas do que os homens, e nas idades superiores a 30 anos, ocorre o inverso. Este fato, por si só, ilustra a mudança de padrão de comportamento das famílias no que tange à educação formal dos seus filhos. Em épocas passadas (até mesmo de passado recente) as mulheres não eram preparadas para o mercado de trabalho e, nesse sentido, pouco interesse havia, por parte das famílias, em alfabetizá-las. Este comportamento, ao que parece, vem-se modificando e, nesse último caso, a escolaridade formal passa a ser percebida socialmente, não apenas como um requesito para o mercado de trabalho e sim como um valor universal. Dessa forma, as mulheres levam vantagem nos atributos da escolaridade, de vez que os homens, especialmente nas camadas sociais menos favorecidas, ao se inserirem mais cedo no mercado de trabalho passam a dispor de menos oportunidades de alfabetização nas idades mais jovens. Este padrão de comportamento se distribui no mesmo sentido nas regiões urbanas e rurais, embora nas últimas ocorra com mais intensidade, como pode ser observado pelos diferenciais de alfabetização entre homens e mulheres no grupo etário de 7 a 29 anos (Tabela 8). Isto porque é justamente nas regiões rurais que a precocidade do trabalho masculino se verifica com maior intensidade.

TABELA 7
TAXAS DE ALFABETIZAÇÃO ESPECÍFICA POR SEXO, SEGUNDO OS GRUPOS DE IDADE — BRASIL — 1985

GRUPOS DE IDADE	TAXAS DE ALFABETIZAÇÃO POR SEXO	
	Homens (%)	Mulheres (%)
5 a 6 anos	6,9	7,3
7 a 9 anos	52,9	57,4
10 a 14 anos	79,8	85,4
15 a 19 anos	86,2	91,3
20 a 24 anos	87,7	90,5
25 a 29 anos	87,6	88,0
30 a 39 anos	84,0	81,8
40 a 49 anos	76,7	71,8
50 a 59 anos	71,2	63,1
60 anos e mais	58,2	47,4

FONTE — IBGE. PNAD — 1985.

Quando é observada a distribuição dos homens e mulheres segundo os anos de estudo adquiridos, ao longo do processo de escolarização formal, verifica-se que, em 1985, as mulheres são maioria em quase todas as faixas, sendo responsáveis por 50,2% do total de pessoas de 1 a 4 anos de estudo; 50,9% das de 5 a 8 anos e 54,3% das de 9 a 11 anos (Tabela 9).

Isto só não acontece na faixa dos 12 anos e mais, onde as mulheres respondem por

TABELA 8
TAXAS DE ALFABETIZAÇÃO POR SITUAÇÃO DE DOMICÍLIO E SEXO, SEGUNDO GRUPOS DE IDADE SELECIONADOS — BRASIL — 1985

GRUPOS DE IDADE	SITUAÇÃO DE DOMICÍLIO E SEXO			
	Urbana		Rural	
	Homens (%)	Mulheres (%)	Homens (%)	Mulheres (%)
7 anos e mais	85,6	83,6	57,6	58,9
7 a 29 anos	87,6	89,8	61,3	68,0
30 anos e mais.....	83,1	76,0	52,0	45,4

FONTE — IBGE. PNAD — 1985.

TABELA 9
DISTRIBUIÇÃO DAS PESSOAS DE 10 ANOS E MAIS POR SEXO, SEGUNDO ANOS DE ESTUDO — BRASIL — 1985

ANOS DE ESTUDO	DISTRIBUIÇÃO DAS PESSOAS POR SEXO		
	Total (1 000 pessoas)	Homens (%)	Mulheres (%)
TOTAL (1)	98 034,9	49,0	51,0
Sem instrução e menos de 1 ano	20 456,7	48,4	51,6
1 a 4 anos.....	41 766,3	49,8	50,2
5 a 8 anos.....	21 021,6	49,1	50,9
9 a 11 anos.....	10 091,5	45,7	54,3
12 anos ou mais.....	4 698,8	52,3	47,7

FONTE — IBGE. PNAD-1985.
(1) Exclui as pessoas sem declaração de anos de estudo.

47,7% do total de pessoas. Esse último grupo corresponde à instrução superior onde, pelos motivos históricos já mencionados, as mulheres teriam desvantagens comparativas em função de padrões pretéritos de comportamento das famílias, no que tange à escolaridade feminina. Existem evidências, no entanto, de que as mulheres, hoje em dia, compõem a maioria dos ingressantes nos cursos universitários, inclusive em profissões antes tidas como tradicionalmente masculinas, como medicina, por exemplo.

Vale dizer, no entanto, que do total de mulheres de 10 anos e mais existentes no Brasil em 1985, 21,1% não tinham instrução ou tinham menos de 1 ano de estudo. A maioria das mulheres (42%) tinha entre 1 e 4 anos de estudo e apenas 4,5% detinha 12 ou mais anos de estudo (Tabela 10). Por estimativa 16,1 milhões, cerca de 24,3% do total de mulheres do País, freqüentavam escola naquele ano. Destas, a maior parte, 77,9% cursava o 1.º grau; 10,5% o 2.º grau; e 4,3% a Universidade (Tabela 11).

Em termos comparativos, a escolaridade feminina no Brasil não é, nos dias de hoje, substancialmente distinta da masculina. Pode-se dizer que, nos últimos 30 anos, as mulheres não só alcançaram, até mesmo suplantaram, as taxas de alfabetização dos homens nos grupos mais jovens. Quanto aos níveis de escolaridade, observa-se que as mulheres se diferenciam por disporem de maiores contingentes entre 9 a 11 anos de estudo e menores com 12 e mais anos de estudo do que os homens, embora, ao que parece, este quadro esteja sendo superado, de vez que parcela crescente dos universitários pertence ao sexo feminino.

TABELA 10
DISTRIBUIÇÃO PERCENTUAL DOS HOMENS E MULHERES, SEGUNDO OS ANOS DE ESTUDO — BRASIL — 1985

ANOS DE ESTUDO	DISTRIBUIÇÃO PERCENTUAL			
	Homens		Mulheres	
	Absoluto (1 000 pessoas)	%	Absoluto (1 000 pessoas)	%
TOTAL (1)	48 090,9	100,0	49 984,1	100,0
Sem instrução e menos de 1 ano	9 901,9	20,6	10 554,8	21,1
1 a 4 anos	20 794,5	43,2	21 011,8	42,0
5 a 8 anos	10 326,7	21,5	10 694,9	21,4
9 a 11 anos	4 608,7	9,6	5 482,9	11,0
12 anos ou mais	2 459,1	5,1	2 239,7	4,5

FONTE — IBGE. PNAD — 1985.
(1) Exclui as pessoas sem declaração de anos de estudo.

TABELA 11
MULHERES ESTUDANTES DE CINCO ANOS OU MAIS, SEGUNDO O GRAU QUE FREQÜENTAM — BRASIL — 1985

GRAU FREQÜENTADO	MULHERES ESTUDANTES DE CINCO ANOS OU MAIS	
	Absoluto (1 000 pessoas)	%
TOTAL[1]	16 140,7	100,0
PRÉ-ESCOLAR	1 174,4	7,3
1.º GRAU	12 572,3	77,9
2.º GRAU	1 699,2	10,5
SUPERIOR	694,8	4,3

FONTE — IBGE. PNAD — 1985.
(1) Exclui as pessoas sem declaração de grau freqüentado.

4 — A MULHER E A ATIVIDADE ECONÔMICA

Uma das mais rápidas transformações na conjuntura do mercado de trabalho dos últimos vinte anos diz respeito ao aumento da participação da mulher. Entre 1976 e 1985, a taxa de atividade feminina evoluiu de 28,7% para 36,9%. Nesse período, o contingente de mulheres economicamente ativas cresceu numa taxa geométrica anual de 5,6%, enquanto masculino cresceu, apenas 2,9%.

Mesmo com essa performance, as taxas de atividades masculinas ainda são mais de duas vezes superiores às femininas, como pode ser observado na Tabela 12. O aumento da inserção da mulher no mercado de trabalho decorre de modificações gerais, tais como mudanças nos padrões de comportamento social refletidas na maior aceitação da mulher num número crescente de ocupações e postos de trabalho, como, também, de fatores conjunturais como rebaixamento dos salários ocorrido ao longo do período 1964-84. Neste último caso, houve um estímulo ao emprego feminino, quando da estratégia de recomposição dos orçamentos familiares, deteriorados pelo arrocho salarial dos anos sessenta e setenta. Tal estímulo não se estendeu somente aos cônjuges mulheres, mas também aos filhos, e representa boa parte do emprego feminino nos postos de trabalho de baixa qualificação.

No entanto, o aumento da escolaridade feminina tem permitido maiores oportunidades de ingresso da mulher em ocupações e postos de trabalho especializados, influenciando o incremento da taxa de atividade feminina não só nas idades jovens (a partir dos 18 anos) como nas mais maduras. Assim, a taxa de atividade feminina específica para o grupo etário 20-24 anos passou de 41,5% para 50,1% entre 1976 e 1985, enquanto que a respectiva taxa masculina evoluiu apenas de 89,7% para 92,5%, no mesmo período (Tabela 13). Esta faixa de

TABELA 12
TAXAS DE ATIVIDADE ESPECÍFICA POR SEXO, SEGUNDO OS ANOS DE REFERÊNCIA – BRASIL 1976-1985

ANOS DE REFERÊNCIA	TAXAS DE ATIVIDADE ESPECÍFICA POR SEXO		
	Total (%)	Homens (%)	Mulheres (%)
1976	50,68	73,53	28,66
1977	52,91	74,02	32,56
1978	53,28	74,40	32,90
1979	53,62	74,53	33,45
1981	53,42	74,63	32,87
1983	54,84	74,80	35,57
1985	56,08	76,03	36,89

FONTE – IBGE. Metodologia da PNAD na década de 1970 e PNADs – 1981, 1983 e 1985.
NOTA – A taxa de atividade específica por sexo é definida como razão entre a população economicamente ativa de 10 anos e mais, dividida pela população total de 10 anos e mais, no sexo considerado.

TABELA 13
TAXAS DE ATIVIDADE ESPECÍFICA POR SEXO, SEGUNDO OS GRUPOS DE IDADE – BRASIL 1976-1985

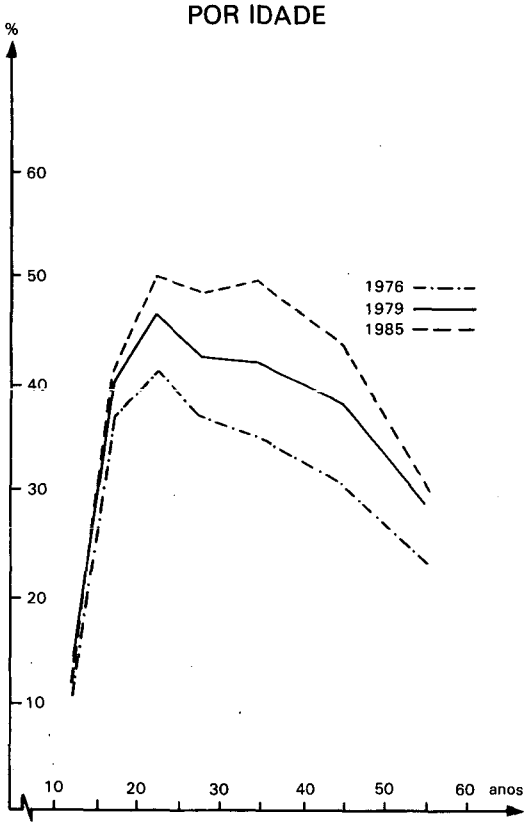
GRUPOS DE IDADE	TAXAS DE ATIVIDADE POR SEXO					
	1976		1979		1985	
	Homens (%)	Mulheres (%)	Homens (%)	Mulheres (%)	Homens (%)	Mulheres (%)
10 a 14 anos	25,68	11,43	27,05	14,23	26,52	12,24
15 a 19 anos	68,84	37,02	70,36	40,23	73,33	41,74
20 a 24 anos	89,73	41,47	91,39	46,48	92,50	50,13
25 a 29 anos	96,34	36,83	96,48	42,37	97,16	48,46
30 a 39 anos	96,43	34,71	97,18	41,60	97,44	49,71
40 a 49 anos	93,39	30,35	93,95	38,00	93,95	43,53
50 a 59 anos	84,10	22,81	83,27	28,35	80,79	30,30
60 anos ou mais	49,31	9,74	47,71	10,59	45,24	10,42

FONTE – IBGE. PNADs – 1976, 1979 e 1985.

idade corresponde ao maior percentual de mulheres em toda a trajetória da vida ativa feminina. Mas o maior incremento, em termos de contingente, foi o verificado nas faixas etárias seguintes, como pode ser visto no gráfico 1.

Este gráfico mostra que, em 1985, a distribuição etária da atividade feminina é bimodal. A primeira moda é a faixa de 20 a 24 anos, período em que, em média, as mulheres se inserem com maior freqüência no mercado de trabalho. Dos 25 aos 29 anos, existia um refluxo provocado pelo período de guarda e educação dos filhos. Dos 30 aos 39 anos ocorre a segunda moda, facilitada pelo crescimento dos filhos e seu ingresso mais autônomo na escola ou no mercado de trabalho. Vale dizer que as próprias empresas dificultam a contratação de mulheres nas faixas etárias contidas entre o casamento e o nascimento dos primeiros filhos, deprimindo os níveis de atividade econômica feminina, em particular dos 25 aos 29 anos de idade.

GRÁFICO 1
BRASIL 1976-1985
TAXA DE ATIVIDADE FEMININA
POR IDADE



Os dados da Pesquisa Nacional por Amostra de Domicílios indicam, ainda, que a escolaridade é determinante na elevação dos níveis de participação econômica da mulher. Em 1985, a taxa de atividade feminina foi de 62,06% para as mulheres com 9 ou mais anos de estudo e de apenas 36,55% para as com 5 a 8 anos de estudo. O brusco diferencial de atividade acima registrado, entre estas duas faixas de escolaridade não é observado nas faixas etárias masculinas correspondentes conforme registro na tabela 14 do anexo. Ao mesmo tempo, a taxa de atividade masculina é mais de duas vezes superior à feminina em todos os intervalos de anos de estudo, com exceção da faixa de 9 anos e mais, onde a relação é de apenas 1,4.

Estes dados indicam que, quanto maior a escolaridade alcançada pelas mulheres, maiores suas oportunidades concretas de inserção na força de trabalho. Desta forma, 26,0% das mulheres economicamente ativas detêm 9 anos de estudo ou mais, contra apenas 16,9% dos homens (Tabela 15).

A População Economicamente Ativa — — PEA feminina no Brasil, era composta por 17,7 milhões de mulheres ocupadas e por 0,7 milhão a procura de trabalho, em 1985. A taxa feminina de desemprego aberto[4] foi

TABELA 14
TAXAS DE ATIVIDADE ESPECÍFICA POR SEXO, E SITUAÇÃO DO DOMICÍLIO, SEGUNDO OS ANOS DE ESTUDO — BRASIL — 1985

ANOS DE ESTUDO	TAXA DE ATIVIDADE ESPECÍFICA POR SEXO					
	Total		Urbano		Rural	
	Homens	Mulheres	Homens	Mulheres	Homens	Mulheres
TOTAL	76,03	36,89	73,06	37,48	83,86	35,05
Sem instrução e menos de 1 ano	74,40	28,69	65,57	26,03	82,29	32,05
1 e 2 anos	65,09	27,76	57,03	26,47	78,29	30,14
3 e 4 anos	76,17	33,69	71,62	32,01	88,53	38,53
5 e 8 anos	77,02	36,55	75,61	36,36	86,88	38,05
9 anos ou mais	87,54	62,06	87,37	64,78	90,85	58,99

FONTE — IBGE. PNAD — 1985.

[4] A taxa de desocupação, erroneamente denominada taxa de desemprego aberto, é dada pela relação entre as pessoas que, não tendo trabalhado, procuraram efetivamente trabalho num dado intervalo de tempo sobre a população economicamente ativa. A população economicamente ativa é constituída pela soma das pessoas ocupadas com aquelas que, ao estarem desocupadas, procuraram efetivamente trabalho num dado período de referência. Uma discussão sobre estes conceitos pode ser vista em Medici e Aguiar — 1985.

TABELA 15
DISTRIBUIÇÃO DAS PESSOAS ATIVAS, POR SEXO, SEGUNDO OS ANOS DE ESTUDO — BRASIL — 1985

ANOS DE ESTUDO	DISTRIBUIÇÃO POR SEXO			
	Homens ativos		Mulheres ativas	
	Absoluto	%	Absoluto	%
TOTAL	36 554 841	100,0	18 408 877	100,0
Sem instrução e menos de 1 ano	7 367 235	20,2	3 027 742	16,4
1 e 2 anos	4 651 682	12,7	1 871 159	10,2
3 e 4 anos	10 395 313	28,4	4 807 751	26,2
5 a 8 anos	7 953 739	21,8	3 909 462	21,2
9 anos ou mais	6 186 872	16,9	4 792 763	26,0

FONTE — IBGE. PNAD — 1985.

de 3,78%, aliás menor observada para a PEA feminina na década de 80. Em confronto com o desemprego masculino, nota-se que, em termos relativos, não há grandes diferenças nos níveis de desocupação entre os sexos. Cabe observar, no entanto que, com exceção de 1976 e 1983, as taxas de desocupação femininas têm sido sempre maiores do que as masculinas (Tabela 16).

Os níveis de desemprego feminino, no Brasil, nos últimos dez anos (1976-85) apresentam uma absoluta estabilidade no período 1977-79. A crise econômica no início dos anos oitenta aumenta este percentual abruptamente, chegando, em 1983, a seu nível mais elevado (4,79%). A rápida recuperação econômica, a partir da Nova República, reduziu o desemprego das mulheres para 3,78%. Os dados da Pesquisa Mensal de Emprego do IBGE indicam, também fortes reduções nas taxas de desocupação em 1986, adicionados a uma rápida recuperação do poder aquisitivo do salário, tanto dos homens como das mulheres.

TABELA 16
TAXAS DE DESOCUPAÇÃO ESPECÍFICA POR SEXO, SEGUNDO OS ANOS DE REFERÊNCIA — BRASIL — 1976-1985

ANOS DE REFERÊNCIA	TAXAS DE DESOCUPAÇÃO		
	Total (%)	Masculina (%)	Feminina (%)
1976	1,84	1,85	1,81
1977	2,32	2,12	2,76
1978	2,36	2,13	2,76
1979	2,76	2,74	2,76
1981	4,26	4,21	4,38
1983	4,86	4,89	4,79
1985	3,35	3,17	3,78

FONTE — IBGE. Metodologia da PNAD e PNADs — 1981, 1983 e 1985.

Ao que tudo indica, as mulheres no Brasil vêm galgando, rapidamente, um novo delineamento de sua inserção na atividade econômica, marcado por uma participação cada vez maior no mercado de trabalho. Os avanços na escolaridade feminina, por sua vez, têm induzido a um crescimento do emprego de mulheres em ocupações de maior qualificação profissional. Isso não significa, no entanto, que os níveis de atividade feminina no Brasil estejam próximos aos desejáveis. Nas nações desenvolvidas ou nos países socialistas, a inserção da mulher no mercado de trabalho é largamente superior à verificada no Brasil, onde ainda há um longo caminho a percorrer na busca de uma identidade profissional feminina, externa àquela vivenciada no ambiente doméstico.

5 — CARACTERIZAÇÃO DO EMPREGO FEMININO

A força de trabalho no Brasil vem passando, paulatinamente, por um processo de terciarização. Chama-se terciarização da

força de trabalho à inserção cada vez maior de pessoas ocupadas no setor terciário da economia, representado pelo comércio e pelos serviços das mais diversas naturezas. No Brasil, o emprego, no setor terciário, cresceu de 37,8% para 49,4% dos ocupados entre 1970 e 1985. Porém, a terciarização da economia em países subdesenvolvidos apresenta certas particularidades. "É fato notório, nos dias de hoje, que os países desenvolvidos detêm imensos contingentes de sua força de trabalho nas chamadas atividades terciárias, tanto no setor público como no privado. O setor serviços, em particular, foi um dos que mais cresceram nos últimos 50 anos, haja vista a proliferação das atividades sociais como educação, saúde, cultura e lazer e das atividades "racionalizadoras" do aparelho produtivo e social, como a administração e gerência, informação, comunicações e seus desdobramentos contemporâneos como a informática e a telemática. Todas essas atividades passaram a absorver grandes fatias de massa de lucro global dos setores produtivos criando, em paralelo, emprego e renda para um grande número de trabalhadores e auxiliando, direta ou indiretamente, no escoamento da produção de bens materiais, conjuntamente com a elevação do nível de consumo social e bem-estar destas populações.

Nos países subdesenvolvidos e em vias de desenvolvimento, a expansão do setor terciário também é fenômeno presente, porém não necessariamente em função do mesmo movimento que ocorreu nos países centrais. A aplicação de tecnologias modernas, juntamente com o atraso secular das formas de produção no campo, geravam um desemprego relativo de mão-de-obra e uma dificuldade estrutural de absorção do excedente de força de trabalho nas atividades que necessitam de maior conhecimento técnico. O rápido aumento das taxas de urbanização, acoplado aos movimentos de concentração de terras e aumento de produtividade no campo foram os ingredientes necessários para detonar um processo de terciarização da economia. A tônica foi dada pelo subemprego em atividades de baixa produtividade e de baixa remuneração, onde o impacto nos níveis de renda e de consumo social foi, sem sombra de dúvida, marginal. Isto não significa dizer que não houve o crescimento de um terciário que utilize tecnologias modernas e preste serviços técnicos. Este também existe e, mesmo não sendo hegemônico na totalidade do processo de terciarização, foi determinado conjuntamente com o crescimento das camadas médias urbanas e das atividades terciárias incentivadas pelo Estado" Medici (1985).

No Brasil, o processo de terciarização tem atingido mais fortemente a força de trabalho feminina. Pode-se dizer que, em 1985, 69,2% das mulheres ocupavam-se em atividades terciárias, enquanto que, apenas 39,4% dos homens inseriam-se nesse setor. Como corolário, as mulheres participam menos do mercado de trabalho nos setores primário e secundário. No setor primário, correspondente às atividades agropecuárias e extrativas, inseriam-se apenas 18,4% das mulheres contra 33,6% dos homens. Nas atividades do setor secundário (atividades industriais e de construção civil) o percentual de ocupados era ainda menor; 12,4% de mulheres contra 27,0% de homens (Tabela 17).

Com um detalhamento maior, pode-se dizer que os setores que absorvem com mais intensidade o trabalho feminino são a prestação de serviços (32,2%); o agrícola (18,4%); o social (17,1%); a indústria de transformação (11,6%) e o comércio de mercadorias (10,4%). Destes cinco ramos de atividade, dois caracterizam-se pela absoluta hegemonia das mulheres em seu desempenho: prestação de serviços e o setor social. Nestes, os homens participam apenas com 8,9% e 3,1%. Outros setores, como a indústria de transformação e o comércio, têm sido caracterizados, ao longo dos últimos anos, por forte crescimento da participação feminina, embora continuem com predominância masculina. A inserção feminina na indústria não ocorre homogeneamente em todos os ramos. Aqueles que dependem de atividades, em que o emprego da força física é marcante, são ocupados em sua quase totalidade por homens. Ocorre o inverso nas atividades industriais onde a habilidade manual e a paciência são requisitos para o trabalho. Assim, nos ramos tradicionais têm-se observado uma maior inserção feminina nas indústrias de fiação e tecela-

TABELA 17
DISTRIBUIÇÃO PERCENTUAL DAS PESSOAS OCUPADAS POR SEXO, SEGUNDO OS SETORES DE ATIVIDADE — BRASIL — 1985

SETORES DE ATIVIDADE	DISTRIBUIÇÃO PERCENTUAL POR SEXO					
	Total		Homens		Mulheres	
	Absoluto (1 000 pessoas)	%	Absoluto (1 000 pessoas)	%	Absoluto (1 000 pessoas)	%
TOTAL	53 236,9	100,0	35 462,9	100,0	17 774,0	100,0
Primário	15 190,4	28,5	11 913,9	33,6	3 276,5	18,4
Secundário	11 784,0	22,1	9 574,1	27,0	2 208,9	12,4
Terciário	26 262,5	49,4	13 974,9	39,4	12 288,6	69,2

FONTE — IBGE. PNAD — 1985.

gem desde os tempos mais remotos do crescimento industrial no Brasil (Tabela 18).

Recentemente, a inserção feminina na indústria tem sido mais intensa em alguns ramos como as linhas de montagem de componentes eletrônicos, mecânica fina, etc. Vale dizer que grande parte da indústria desses componentes, localizada em Manaus, utiliza a mão-de-obra feminina, embora pagando salários bem inferiores às taxas de remuneração vigentes nesses setores.

Em termos gerais, as mulheres respondiam, em 1985, por 33,3% do total de ocupados. No entanto, em dois ramos de atividade as mulheres são maioria absoluta. Cerca de 65% dos ocupados no setor de prestação de serviços são mulheres, o mesmo ocorrendo nas atividades sociais, onde a participação feminina é de 73% (Tabela 19). A maioria dos empregos no ramo prestação de serviços corresponde aos chamados serviços pessoais, destacando-se os

TABELA 18
DISTRIBUIÇÃO DAS PESSOAS OCUPADAS POR SEXO, SEGUNDO OS RAMOS DE ATIVIDADE — BRASIL — 1985

RAMOS DE ATIVIDADE	DISTRIBUIÇÃO POR SEXO			
	Homens		Mulheres	
	Absoluto (1 000 pessoas)	%	Absoluto (1 000 pessoas)	%
TOTAL	35 462,9	100,0	17 774,1	100,0
Agrícola	11 913,9	33,6	3 276,5	18,4
Indústria de Transformação	5 779,9	16,3	2 067,4	11,6
Construção Civil	3 041,6	8,6	55,8	0,3
Outras Atividades Industriais	753,6	2,1	85,7	0,5
Comércio de Mercadorias	3 959,6	11,2	1 855,0	10,4
Prestação de serviços	3 142,5	8,9	5 711,6	32,1
Serviços Auxiliares	1 029,9	2,9	403,6	2,3
Transporte e Comunicações	1 766,7	5,0	149,3	0,8
Social	1 118,6	3,1	3 032,4	17,1
Administração Pública	1 754,4	4,9	592,4	3,3
Outras	1 202,2	3,4	544,4	3,1

FONTE — IBGE. PNAD — 1985.

TABELA 19
DISTRIBUIÇÃO PERCENTUAL DAS PESSOAS OCUPADAS POR SEXO, SEGUNDO RAMOS DE ATIVIDADE — BRASIL — 1985

RAMOS DE ATIVIDADE	DISTRIBUIÇÃO PERCENTUAL POR SEXO		
	Total	Homens	Mulheres
TOTAL	100,0	66,7	33,3
Agrícola	100,0	78,4	21,6
Indústria de Transformação	100,0	73,7	26,3
Construção Civil	100,0	98,2	1,8
Outras Atividades Industriais	100,0	89,8	10,2
Comércio de Mercadorias	100,0	68,2	31,8
Prestação de Serviços	100,0	35,5	64,5
Serviços Auxiliares	100,0	71,8	28,2
Transportes e Comunicações	100,0	92,2	7,8
Social	100,0	26,9	73,1
Administração Pública	100,0	74,5	25,5
Outras Atividades	100,0	68,8	31,2

FONTE — IBGE. PNAD — 1985.

empregados domésticos, faxineiros etc. Já nas atividades sociais, as ocupações mais freqüentes são aquelas ligadas aos setores de saúde e educação. Assim, os ramos de atividade onde é hegemônica a presença feminina, constituem-se, praticamente, em tarefas desempenhadas pelas mulheres em épocas mais remotas no seio das famílias, ou seja, os afazeres domésticos e as tarefas de cuidado com educação e saúde dos filhos. A maior inserção da mulher nestes segmentos da atividade econômica, corresponde, em certas medidas à mercantilização desses serviços, os quais encontram-se cada vez mais disponíveis no mercado e cada vez menos na esfera do autoconsumo. Vale dizer, ainda, que excluídos os ramos de prestação de serviços e o social, a participação feminina no conjunto dos ocupados cai de 33% para 22%.

Estes dados indicam que ainda há um grande espaço a ser ocupado pelas mulheres nos ramos de atividade ligados à indústria e aos serviços, pois a inserção feminina nas atividades agrícolas, apesar de ser elevada, tem sido historicamente subestimada. A participação de 21,6% de mulheres, nos ocupados agrícolas, registrada em 1985 (Tabela 19), reflete, em grande medida, o movimento de assalariamento no campo, deixa porém de levar em conta o trabalho da mulher nas pequenas unidades familiares de produção agrícola, o qual, gera-

mente é diluído nas pesquisas domiciliares pelo registro equivocado em afazeres domésticos. Esse fenômeno foi marcante no Censo Demográfico de 1970, onde o sub-registro de trabalho feminino no campo foi bastante elevado. Mudanças conceituais e alterações na forma de captação trouxeram uma sensível redução desse sub-registro nas pesquisas subseqüentes.

6 — FORMALIDADE X INFORMALIDADE: UMA ANALISE DA POSIÇÃO OCUPACIONAL DA MULHER NO MERCADO DE TRABALHO

A distribuição das mulheres ocupadas em 1985 revela que o maior contingente do trabalho feminino é composto por empregados assalariados. Cerca de 67% das mulheres encontravam-se nessa posição, percentua esse, ligeiramente maior do que o verificado para os homens (64%). As mulheres também apresentavam, naquele ano, contingente de trabalhadoras autônomas e de empregadores, inferior ao registrado para a parcela masculina. A Tabela 20 mostra que a participação de autônomos era de 24% para os homens e 19% para as mulheres e a dos empregadores, de 4,2% e 1,1%, respectivamente.

TABELA 20
DISTRIBUIÇÃO DAS PESSOAS OCUPADAS POR SEXO, SEGUNDO A POSIÇÃO NA OCUPAÇÃO — BRASIL — 1985

POSIÇÃO NA OCUPAÇÃO	DISTRIBUIÇÃO DAS PESSOAS POR SEXO					
	Total		Homens		Mulheres	
	Absoluto (1 000 pessoas)	%	Absoluto (1 000 pessoas)	%	Absoluto (1 000 pessoas)	%
TOTAL	53 236,9	100,0	35 462,9	100,0	17 774,0	100,0
Empregados	34 387,7	64,6	22 544,5	63,6	11 843,2	66,6
Autônomos	12 112,2	22,7	8 642,1	24,4	3 470,1	19,5
Empregadores	1 701,4	3,2	1 501,2	4,2	200,2	1,2
Não Remunerados	5 035,6	9,5	2 775,1	7,8	2 260,5	12,7

FONTE — IBGE. PNAD — 1985.

O emprego feminino ainda pode ser caracterizado pela presença marcante de trabalhadores sem remuneração. Cerca de 13% das mulheres eram, em 1985, trabalhadoras não remuneradas, enquanto que entre os homens esse percentual não ultrapassava os 8%.

Embora o assalariamento feminino seja maior, em termos relativos, que o encontrado para o efetivo masculino, as mulheres empregadas se encontram em situação mais precária que os homens. Das mulheres assalariadas, 48,9% não tinham carteira de trabalho assinada pelo empregador, enquanto apenas 40,6% dos empregados do sexo masculino estavam nessa situação.

Uma taxa aproximada de precariedade do mercado de trabalho poderia ser obtida pela soma dos empregados sem carteira assinada; dos autônomos e dos trabalhadores sem remuneração. A inclusão dos autônomos nessa *proxy* de precariedade justificar-se-ia dado que a maioria desses trabalhadores é constituída por biscateiros, tarefeiros e empregados domésticos, sem ocupação regular e com baixo nível de qualificação profissional.

A taxa de precariedade ficaria em torno de 64,9% para as mulheres ocupadas e de 58,4% para os homens ocupados. Estes dados, no entanto, devem ser utilizados com reservas, dada a existência de empregados com altos rendimentos, e sem carteira assinada, ou, ainda, a presença de autônomos em posições favoráveis no mercado de trabalho, como é o caso dos profissionais liberais. Portanto, o citado indicador refere-se, mais fortemente, à avaliação de situações distintas do assalariamento formal.

Retornando à análise das mulheres na condição de assalariadas, observa-se que, no setor secundário, as formalidades para o emprego feminino, indicadas pelo número de carteiras de trabalho assinadas, são maiores do que as verificadas para os homens. A Tabela 21 mostra que 80% das mulheres ocupadas nas atividades industriais ou assemelhadas portavam carteiras, enquanto apenas 77% dos homens encontravam-se nessa condição. Nos outros dois setores, a participação do emprego formal no assalariamento feminino era bem inferior à dos empregados do sexo masculino.

Um outro indicador que pode aferir a inserção da mulher no mercado de trabalho formal é dado pela taxa de contribuição para a Previdência Social. Como pode ser visto na Tabela 22, desde 1981, existe uma tendência inequívoca à queda dos níveis de contribuição para a previdência, no Brasil. Apenas 47,3% dos trabalhadores brasileiros eram contribuintes do Sistema Nacional de Previdência e Assistência Social, em 1985. Tal constatação apresenta duplo efeito. Por um lado, mostra que grande parte dos trabalhadores no Brasil estão destituídos dos benefícios e da assistência médica garantida pelos serviços previdenciários. Por outro lado, indica que existe grande caminho a percorrer no sentido de universalizar o trabalho regular e garantir maior mar-

TABELA 21
DISTRIBUIÇÃO PERCENTUAL DOS EMPREGADOS POR CONDIÇÃO DE POSSE DE CARTEIRA DE TRABALHO ASSINADA E SEXO, SEGUNDO O SETOR DE ATIVIDADE — BRASIL — 1985

SETOR DE ATIVIDADE	DISTRIBUIÇÃO DOS EMPREGADOS POR SEXO E POSSE DE CARTEIRA DE TRABALHO ASSINADA								
	Total			Homens			Mulheres		
	Total (%)	Possuíam (%)	Não possuíam (%)	Total (%)	Possuíam (%)	Não possuíam (%)	Total (%)	Possuíam (%)	Não possuíam (%)
TOTAL	100,0	56,5	43,5	100,0	59,4	40,6	100,0	51,1	48,9
Primário	100,0	17,3	82,7	100,0	17,8	82,2	100,0	13,7	86,3
Secundário	100,0	77,1	22,9	100,0	76,3	23,7	100,0	80,2	19,8
Terciário	100,0	58,0	42,0	100,0	62,5	37,5	100,0	48,5	51,5

FONTE — IBGE. PNAD — 1985.

gem de custeio ao Sistema de Previdência e à Assistência Social no Brasil, dadas as atuais bases financeiras de sustentação[5]. A crise que se configura no início dos anos oitenta, mostra que a sustentação financeira do SINPAS torna-se mais vulnerável justamente nos períodos em que a demanda por benefícios previdenciários e assistência médica aumenta. Verifica-se que a taxa de contribuição à previdência caiu de 49,9% em 1981, para 47,3% em 1985 (Tabela 22).

A análise da citada taxa, segundo o sexo, evidencia que as mulheres se encontram em situação mais precária do que os homens, em termos de registro, junto às entidades previdenciárias. Apenas 44% das mulheres ocupadas contribuíram para a previdência em 1985, enquanto que, entre os homens, tal contribuição chegava a 49%. De qualquer forma, os níveis de contribuição não só de homens, como de mulheres para a previdência no Brasil, encontram-se muito aquém do desejado.

Uma rápida visualização dos dados expostos na Tabela 23 mostra que, em quase todos os setores de atividade, o registro de contribuição a entidades previdenciárias é menor para as mulheres com relação aos homens ocupados. A exceção se encontra

TABELA 22
EVOLUÇÃO DA TAXA DE CONTRIBUIÇÃO À PREVIDÊNCIA POR SEXO, SEGUNDO OS ANOS DE REFERÊNCIA — BRASIL — 1976-1985

ANOS DE REFERÊNCIA	EVOLUÇÃO DA TAXA DE CONTRIBUIÇÃO POR SEXO		
	Total	Homens	Mulheres
1978	48,3	50,8	42,5
1979	49,1	51,4	44,4
1981	49,9	51,7	46,0
1983	47,7	49,8	43,3
1985	47,3	49,1	43,7

FONTE — IBGE. PNADs — 1976 a 1985.
NOTA-A taxa de contribuição à previdência é definida pela razão entre as pessoas ocupadas que contribuem à previdência e o total das pessoas ocupadas.

[5] O principal problema é que estas bases não se sustentam, dada a impossibilidade de universalizar o acesso aos serviços previdenciários pela via da contribuição sobre a folha de salários. A este respeito vide Medici e Silva — 1986.

TABELA 23
TAXA DE CONTRIBUIÇÃO À PREVIDÊNCIA POR SEXO, SEGUNDO OS RAMOS DE ATIVIDADE — BRASIL — 1985

RAMOS DE ATIVIDADE	TAXA DE CONTRIBUIÇÃO POR SEXO		
	Total	Homens	Mulheres
Agrícola	7,1	8,6	1,8
Indústria de Transformação	78,6	81,8	69,4
Indústria de Construção	49,5	49,0	78,1
Outras Atividades Industriais	77,1	76,1	86,0
Comércio de Mercadorias	59,2	61,1	55,2
Prestação de Serviço	32,1	54,7	19,7
Serviços Auxiliares	73,9	75,1	70,9
Transportes e Comunicações	76,7	75,5	90,9
Social	85,0	85,7	84,8
Administração Pública	85,1	82,7	92,3
Outras Atividades	80,4	77,8	86,3

FONTE — IBGE. PNAD — 1985.

nos ramos de construção civil (onde o contingente feminino é pequeno, e geralmente encontra-se nos escritórios de administração das empresas, nas outras atividades industriais, nos transportes e comunicação e na administração pública. A diferença mais gritante no que tange à taxa de contribuição previdenciária é a que pode ser verificada no ramo de prestação de serviços. Neste ramo, apenas 19,7% das mulheres ocupadas são contribuintes de sistemas previdenciários; proporção significativamente menor do que os 54,7% observados para os homens. É justamente nesse ramo que se ocupavam 5,7 milhões de mulheres brasileiras, quase 1/3 de toda a força de trabalho feminina do País, nas mais desqualificadas atividades, desde empregos domésticos até biscates dos mais variados. Os dados indicam que cerca de 4,2 milhões de mulheres ocupadas no citado ramo não contribuem para instituições de previdência.

O crescimento da ocupação feminina, particularmente nas camadas sociais mais empobrecidas, esteve associado, nestes últimos anos, a vários fenômenos como a urbanização, o assalariamento e a queda do poder aquisitivo das remunerações do trabalho. As mulheres passaram a deter importante papel na composição dos orçamentos familiares. Tal fato, associado à ausência de políticas sociais voltadas para a criação de creches e escolas, bem como à difusão dos meios e informações sobre contracepção devem ter fortes efeitos na redução recente dos níveis de fecundidade.

Como conseqüência desses fatores, a mulher ainda ingressa no mercado de trabalho em condições desfavoráveis, haja vista a grande desvantagem feminina na disputa dos postos de trabalho no mercado formal. Assim, as mulheres, em sua grande maioria, estão inseridas em ocupações onde a posse da carteira de trabalho assinada ou a contribuição para a Previdência Social é bem menor do que a verificada no conjunto dos trabalhadores masculinos.

7 — OS RENDIMENTOS DA MULHER TRABALHADORA

Os indicadores disponíveis sobre rendimentos no Brasil, captados em pesquisas domiciliares, não permitem construir séries históricas muito longas. Isto porque, somente a partir do Censo Demográfico de 1960, iniciou-se a coleta regular, com base em informações censitárias de dados dessa natureza. Todas as fontes existentes desde então, destacando-se Censos Demográficos e Pesquisas Nacionais por Amostra de Domicílios (PNADs), evidenciam-se que:

a) os rendimentos médios do trabalho masculino têm sido sempre superiores aos verificados para as mulheres trabalhadoras Medici (1982);

b) em conseqüência, na distribuição de homens e mulheres por classes de rendimento, observa-se sensível diferença, marcada pela maior freqüência relativa de mulheres, nas faixas de rendimento mais baixas, ocorrendo o inverso nas faixas mais elevadas, onde é maior o número relativo de homens;

c) por apresentar rendimentos mais baixos, inclusive nos limites superiores da distribuição, a renda do trabalho feminino é menos concentrada do que a relativa ao trabalho masculino, o que fica demonstrado nas medidas de distribuição existente.

Cada um desses três pontos merece uma análise pormenorizada. Entre 1970 e 1977, a relação entre o rendimento médio dos ocupados do sexo masculino e o do feminino aumentou de 1,65 para 2,09, indicando forte crescimento da renda dos homens com relação à das mulheres trabalhadoras. Vale lembrar que, apesar da elevação do trabalho feminino naquele período, a repressão aos movimentos sociais de cunho sindical impediu uma evolução dos salários reais, principalmente nas ocupações de baixa e média remuneração, que constituíam a quase totalidade do emprego feminino, no período. Presume-se que esse tenha sido um dos motivos que levaram a uma elevação no diferencial dos rendimentos médios entre homens e mulheres.

A partir de 1977, esses diferenciais começam a se reduzir, indicando uma elevação do rendimento médio feminino superior à verificada para os homens. Assim, a razão entre os dois rendimentos decresce de 2,09 em 1977 para 1,86 em 1981 (Tabela 28). De qualquer forma o rendimento médio masculino é cerca de 86% superior ao encontrado para as mulheres, de onde se deduz que ainda há um grande caminho a percorrer no sentido de elevar a remuneração do trabalho feminino.

Vários são os motivos que podem explicar o diferencial dos rendimentos entre homens e mulheres. O primeiro diz respeito à duração da jornada média de trabalho. Em 1985, cerca de 10,6% dos homens trabalhavam menos de 40 horas por semana, enquanto 35,1% das mulheres tinham essa jornada. O fato de haver um maior contingente de mulheres trabalhando *em tempo parcial* influencia, naturalmente, o nível de renda auferido, tornando-o mais baixo que o percebido pelos trabalhadores do sexo masculino. Portanto, cerca de 56% dos homens trabalhavam de 40 até 48 horas por semana, contra 44% das mulheres, o mesmo acontecendo na faixa de 49 horas e mais, onde os percentuais eram de 34% contra 20%, respectivamente (Tabela 26).

Assim, a existência de distintas jornadas semanais de trabalho pode explicar boa parte do diferencial de rendimento entre homens e mulheres ocupados. Resta saber quais os motivos que levam as mulheres a trabalharem menos do que os homens. Com relação a isso, há uma série de considerações. Questões já amplamente discutidas, como a dupla jornada de trabalho da mulher — em sua ocupação cotidiana e em afazeres domésticos — bem como a necessidade de cuidar dos filhos num país onde as empresas não cumprem a legislação relativa ao fornecimento de creches e os governos não suprem satisfatoriamente tais carências, fazem com que grande parte das mulheres procurem ocupações de tempo parcial, como forma de conciliar sua vocação profissional, à sobrevivência econômica ao desenvolvimento da família. Nesse sentido, a própria estratificação social dos empregos ocupados por homens e mulheres reflete uma divisão sexual do trabalho, dada por condições sociais alheias à própria vontade feminina. Existem exceções, especialmente, quando ligadas aos empregos de nível superior. Em muitos destes empregos não existem diferenciais de remuneração quando se trata de postos de trabalho semelhantes, com a mesma jornada de trabalho para homens e mulheres e igual faixa etária. O grande problema da baixa remuneração feminina reside, portanto, na diferença quanto à natureza das ocupações e à magnitude da jornada de trabalho. Na medida em que as mulheres passam a concorrer com os homens (e isto já vem ocorrendo no Brasil) para ocupar os empregos hoje tidos como *tipicamente* masculinos, poderá ocorrer uma sensível redução nos diferenciais de salário entre os sexos.

O segundo ponto diz respeito à distribuição de homens e mulheres ocupados por faixa de renda. Quanto a isso, a primeira constatação é a existência de grande número de mulheres que trabalham sem rendi-

mento no Brasil. Em 1976, 17,7% dos ocupados femininos não auferiam qualquer rendimento. A literatura disponível encontra nos empregos domésticos e no trabalho familiar agrícola a maioria desse contingente. Não resta dúvida, no entanto, que o aumento da urbanização, a mercantilização da economia, marcada por uma efetiva monetarização das relações de troca e o assalariamento, enquanto, em contrapartida, pela ausência crescente de formas não monetárias de remuneração e sobrevivência, são fatores decisivos na redução do número de trabalhadores não remunerados no Brasil, Medici (1984). Assim, a participação dos ocupados sem remuneração, no total de trabalhadores femininos, reduziu-se para 13,0% em 1985. No entanto, apenas 8,0% dos homens encontravam-se nessa condição, naquele mesmo ano.

Além desta situação de precariedade, a sub-remuneração é outro problema que tem afetado mais o emprego feminino do que o masculino. Assim, enquanto 63,8% das mulheres não auferiam rendimentos ou detinham renda de até 1 salário mínimo em 1976, somente 38,3% dos homens inseriam-se nessa faixa. Em 1981, a participação dos ocupados com renda de até 1 salário mínimo reduziu-se para 56,1% e 32,7% para mulheres e homens, respectivamente. Os fatores apontados anteriormente, ou seja, a urbanização, a mercantilização e o assalariamento, valem para explicar essa queda. Mas a redução do contingente de sub-remunerados pode estar indicando, também, uma perda qualitativa relacionada às formas não-monetárias de sobrevivência superior aos benefícios materiais decorrentes da monetarização, e não necessariamente uma elevação na qualidade de vida, como apontam as análises tradicionais. Sociedades como a brasileira, marcadas por uma grande desigualdade social e regional no seu processo de desenvolvimento e por uma forte modernização capitalista recente, trazem, muitas vezes, perdas irreparáveis para as condições de vida daqueles que *involuntariamente* são obrigados a trocar sua condição de subsistência pela de assalariados.

Porém, entre 1981 e 1985, a participação dos ocupados com renda de até um salário mínimo elevou-se, novamente, de 56,1% para 58,6%, no caso das mulheres, e de 32,7% para 33,8%, no dos homens, mesmo levando-se em conta que o salário mínimo teve uma queda real de 16% ao longo do período. A crise econômica do início dos anos oitenta teve forte efeito no aumento dos níveis de desemprego e subemprego, bem como, na queda dos salários reais. Em 1985, já se apresenta o início da recuperação econômica, embora os dados ainda reflitam efeitos da crise na distribuição das pessoas, por faixas de renda.

A Tabela 24 indica como o nível de renda feminino é muito mais baixo do que o masculino. Nas faixas de remuneração mais altas, em 1985, cerca de 16,5% dos homens ocupados ganhavam mais de 5 salários mínimos, enquanto apenas 7,2% das mulheres incluíam-se nessa faixa.

Padronizando-se o rendimento dos ocupados, por classes de horas trabalhadas, observa-se que, mesmo para jornadas de trabalho com a mesma duração, as mulheres encontram-se em posição desvantajosa com relação aos rendimentos auferidos pelos homens. Assim, 46,6% das mulheres, que trabalhavam entre 40 e 48 horas semanais, recebiam até um salário mínimo contra 32,0% dos homens, em 1985. Esse diferencial fica ainda mais gritante, quando se considera a classe de 49 horas e mais por semana; onde 66,2% das mulheres recebiam até 1 salário mínimo, contra 27,7% dos homens no mesmo ano (Tabela 25). Os empregos domésticos são casos típicos de convivência de longas jornadas de trabalho, associadas à baixa remuneração. Boa parte dos empregos, sem proteção trabalhista, também se encontram nessa condição. Como já foi visto, cerca de 51,5% dos empregados femininos não tinham carteira de trabalho assinada em 1985, enquanto que apenas 37,5% dos homens estavam nessa condição.

O terceiro ponto é decorrência da análise feita nos aspectos anteriores. Se as mulheres não têm grandes diferenciais de rendimentos, dado que apenas 2,2% delas recebiam mais de 10 salários mínimos em 1985 (Tabela 24), naturalmente sua renda tende a ser menos concentrada. A Tabela 27 mostra

TABELA 24

DISTRIBUIÇÃO PERCENTUAL DAS PESSOAS OCUPADAS POR ANOS DE REFERÊNCIA E SEXO, SEGUNDO CLASSES DE RENDIMENTO MÉDIO MENSAL — BRASIL — 1976-1985

CLASSES DE RENDIMENTO MÉDIO MENSAL	DISTRIBUIÇÃO POR ANOS DE REFERÊNCIA E SEXO					
	1976			1981		
	Total	Homens	Mulheres	Total	Homens	Mulheres
TOTAL (1)	100,00	100,00	100,00	100,00	100,00	100,00
Sem rendimento	12,37	10,21	17,71	10,08	8,30	13,98
Até 1/2 salário mínimo	11,20	6,49	22,86	13,18	8,81	22,81
Mais de 1/2 a 1 salário mínimo	22,21	21,60	23,73	16,77	15,62	19,32
Mais de 1 a 2 salários mínimos	25,79	27,66	21,17	25,72	26,90	23,12
Mais de 2 a 5 salários mínimos	18,14	21,31	10,29	23,30	26,94	15,30
Mais de 5 a 10 salários mínimos	6,23	7,50	3,07	6,98	8,26	4,16
Mais de 10 salários mínimos	4,06	5,23	1,17	3,97	5,17	1,31

CLASSES DE RENDIMENTO MÉDIO MENSAL	DISTRIBUIÇÃO POR ANOS DE REFERÊNCIA E SEXO		
	1985		
	Total	Homens	Mulheres
TOTAL (1)	100,00	100,00	100,00
Sem rendimento	9,67	7,99	13,04
Até 1/2 salário mínimo	12,59	7,14	23,45
Mais de 1/2 a 1 salário mínimo	19,82	18,66	22,12
Mais de 1 a 2 salários mínimos	22,59	24,34	19,10
Mais de 2 a 5 salários mínimos	21,96	25,39	15,12
Mais de 5 a 10 salários mínimos	8,32	9,99	4,98
Mais de 10 salários mínimos	5,05	6,49	2,19

FONTE — PNADs — 1976, 1981 e 1985.
(1) Exclui as pessoas sem declaração de rendimentos.

TABELA 25

DISTRIBUIÇÃO PERCENTUAL DAS PESSOAS OCUPADAS NO BRASIL POR SEXO E HORAS SEMANAIS TRABALHADAS, SEGUNDO CLASSES DE RENDIMENTO MÉDIO MENSAL — 1985

CLASSES DE RENDIMENTO MÉDIO MENSAL	DISTRIBUIÇÃO POR CLASSES DE HORAS TRABALHADAS E SEXO					
	Até 39 horas		De 40 a 48 horas		49 horas ou mais	
	Homens	Mulheres	Homens	Mulheres	Homens	Mulheres
TOTAL (1)	100,00	100,00	100,00	100,00	100,00	100,00
Até 1 salário mínimo (2)	62,73	69,35	32,01	46,62	27,69	66,24
Mais de 1 a 2 salários mínimos	12,54	13,72	24,88	23,94	27,15	17,85
Mais de 2 a 5 salários mínimos	14,70	12,28	26,26	19,63	27,31	10,21
Mais de 5 a 10 salários mínimos	6,05	3,53	10,31	6,72	10,69	3,68
Mais de 10 salários mínimos	3,98	1,12	6,54	3,09	7,16	2,02

FONTE — IBGE. PNAD — 1985.
(1) Exclui os sem declaração de rendimentos.
(2) Inclui os sem rendimento.

TABELA 26
DISTRIBUIÇÃO PERCENTUAL DAS PESSOAS OCUPADAS NO BRASIL POR SEXO, SEGUNDO CLASSES DE HORAS SEMANAIS TRABALHADAS — 1985

CLASSES DE HORAS SEMANAIS TRABALHADAS	DISTRIBUIÇÃO POR SEXO		
	Total	Homens	Mulheres
TOTAL (1)	100,00	100,00	100,00
Até 39 horas	18,78	10,60	35,10
De 40 a 48 horas	51,85	55,56	44,45
49 horas e mais	29,37	33,84	20,45

FONTE — IBGE. PNAD — 1985.
(1) Exclui os sem declaração de horas trabalhadas.

TABELA 27
DISTRIBUIÇÃO PERCENTUAL DOS RENDIMENTOS DA PEA, COM RENDA MAIOR QUE ZERO, POR SEXO, SEGUNDO DECIS E PERCENTIS — BRASIL — 1979-1981

DISTRIBUIÇÃO DA PEA EM DECIS E PERCENTIS	DISTRIBUIÇÃO DOS RENDIMENTOS DA PEA POR SEXO					
	Total		Homens		Mulheres	
	1979 (%)	1981 (%)	1979 (%)	1981 (%)	1979 (%)	1981 (%)
10^-	1,00	1,03	1,24	1,08	1,53	1,42
10	2,12	1,88	2,24	2,17	1,68	1,57
10	2,72	2,62	2,77	2,85	2,89	2,70
10	3,67	3,82	3,81	3,78	4,38	4,02
10	4,60	4,56	4,31	4,27	4,92	4,97
10	5,31	5,27	5,32	5,59	6,90	6,74
10	7,21	7,53	7,19	7,20	8,19	7,91
10	10,12	10,12	10,23	10,15	10,46	10,74
10	15,90	15,52	16,01	15,44	16,45	16,35
10^+	47,35	47,65	46,86	47,47	42,60	43,58
5^+	34,42	35,17	33,76	34,86	29,48	30,49
1^+	14,10	15,40	13,79	15,22	11,06	12,45
GINI	0,5747	0,5783	0,5674	0,5716	0,5313	0,5459

FONTE — IBGE. Tabulações Especiais das PNADs — 1979 e 1981.

que o coeficiente de Gini[6] referente à renda das mulheres ocupadas era de 0,5459 em 1981 enquanto que o relativo aos homens era pouco maior — 0,5716 — indicando maior concentração. Os dados mais recentes indicam, no entanto, que a renda feminina apresenta ritmo de concentração mais intenso do que o verificado para os homens. De qualquer forma, vale destacar que, em 1981, o decil mais rico das mulheres ocupadas recebia 43,6% da renda total feminina proveniente do trabalho, enquanto o decil mais pobre recebia apenas 1,4%.

Todas essas constatações mostram que a remuneração do trabalho feminino longe está de ser adequada e, em comparação

[6] O Índice de Gini limite inferior, é definido pela fórmula:

$$G = 1 - \sum_{i=1}^{n} (Y_i + Y_{i-1})(X_i - X_{i-1})$$ onde

G = Índice de Gini.
Y_i = Percentual acumulado da renda no estrato i.
X_i = Percentual acumulado da população no estrato i.
n = Número de classes de renda.

Tal medida tem sido o indicador freqüentemente utilizado para medir desníveis de renda entre estratos de população e renda. Varia de zero a um, sendo zero a máxima igualdade e 1 a máxima desigualdade.

TABELA 28
EVOLUÇÃO DO VALOR MÉDIO DOS RENDIMENTOS DAS PESSOAS OCUPADAS SEGUNDO O SEXO — BRASIL — 1977-1981

SEXO	VALOR MÉDIO DOS RENDIMENTOS (cruzeiros de 1981)		
	1977	1979	1981
TOTAL	24 941	24 983	24 495
HOMENS	29 361	29 365	28 424
MULHERES	14 029	14 663	15 292
RAZÃO HOMEM/ /MULHER	2,09	2,00	1,86

FONTE — IBGE. Tabulações Especiais das PNADs — 1977, 1979 e 1981.

com à masculina encontra-se em posição inferior. A Tabela 29, relativa aos empregados formais, ou seja, àqueles que têm qualquer tipo de vínculo trabalhista ou previdenciário com o empregador, mostra que, para a faixa de renda de até 3 salários mínimos, as mulheres detêm escolaridade sensivelmente maior que os homens, tanto na indústria, como nos setores ligados ao comércio e aos serviços. Assim, não é pela falta de atributos de escolaridade, competência ou habilidade que os salários femininos são menores do que os verificados para os homens. Além dos motivos aqui mencionados, existem muitos casos de discriminação salarial da mulher, pelo fato de ser mulher. Portanto, a igualdade de oportunidades e de remuneração do trabalho para homens e mulheres é, antes de tudo, uma questão de cidadania.

8 — A MULHER E SUA POSIÇÃO NA FAMÍLIA

As pesquisas domiciliares do IBGE costumam definir como chefe da família a "pessoa (homem ou mulher) responsável pelo domicílio ou pela família[7]". Esta definição, por ser bastante vaga, encontra-se carregada de subjetivismo. Na medida em que é o entrevistado que declara quem é o chefe da família, fica difícil saber qual o critério que definiu essa relação; se foi o nível de rendimentos, a responsabilidade pelo provimento das condições de existência familiar ou qualquer outro atributo. Numa sociedade *tradicional*, como a brasileira, vários são os motivos que poderiam ser dados para invalidar esse conceito e sua forma de captação.

TABELA 29
PERCENTAGEM DE EMPREGADOS COM SALÁRIO INFERIOR A TRÊS SALÁRIOS MÍNIMOS, POR SETORES DE ATIVIDADE ECONÔMICA E SEXO, SEGUNDO GRAUS DE INSTRUÇÃO — BRASIL — 1976

GRAUS DE INSTRUÇÃO	SETORES DE ATIVIDADE E SEXO					
	Indústria		Comércio		Serviços	
	Mulheres (%)	Homens (%)	Mulheres (%)	Homens (%)	Mulheres (%)	Homens (%)
1.º grau, 4.ª série incompleta	93,79	81,58	94,51	89,34	95,17	84,55
1.º grau, 4.ª série completa	92,96	70,70	90,93	84,12	93,13	79,63
1.º grau, 8.ª série completa	78,85	51,49	86,94	71,70	79,11	59,56
2.º grau completo	56,91	26,52	80,05	50,08	65,87	35,42
Superior completo	27,84	4,82	48,69	18,70	37,15	18,42

FONTE — Relação Anual de Informações Sociais — 1976, Ministério do Trabalho. In: MACEDO, Concessa Vaz de. *O trabalho feminino e suas possíveis implicações sobre o crescimento demográfico.* CEPPD, 1985.

[7] Conforme o manual do recenseador do Censo Demográfico de 1980.

No entanto, pesquisas domiciliares anteriores, como o Censo Demográfico de 1970 e as PNADs realizadas ao longo da década passada, tinham conceitos ainda mais rígidos, na medida em que consideravam como chefe, em princípio, o marido. "A mulher só era considerada chefe, quando não fosse casada ou vivesse em união consensual estável. Por esta razão, a categoria cônjuge (...) era constituída, exclusivamente, por mulheres" Silva (1982).

A análise do papel da mulher na família é de grande importância, pois "as entradas e saídas da mulher no mercado de trabalho, o número de horas trabalhadas e a ocupação exercida são definidas, não só pela estrutura desse mercado, mas também em função da composição e organização da família. No que se refere a este último aspecto, destacam-se, entre outros fatores, o número e a idade dos filhos e a presença de outras pessoas, parentes ou não, que possam auxiliar a mulher ou até mesmo substituí-la na realização das tarefas domésticas" Silva (1982).

Os dados da PNAD-1985 revelam que, da totalidade de mulheres brasileiras, 8,8% eram chefes de família; 36,6% eram cônjuges; 48,2% eram filhas; 5,5% eram outros parentes e 1,5% não tinham parentesco com o chefe. Esta distribuição é bastante distinta da masculina onde cerca de 40,4% dos homens são chefes e os cônjuges não passam de 0,1% (Tabela 31). Assim, percebe-se que a maioria esmagadora das famílias, onde a mulher é considerada chefe, são aquelas cujos núcleos foram partidos por viuvez, separação, desquite ou divórcio. Para exemplificar, em 1980, 72,1% das mulheres separadas, divorciadas e desquitadas eram chefes de família, o mesmo ocorrendo com 65,5% das viúvas e 8,4% das solteiras. No entanto, somente 0,4% das mulheres casadas eram consideradas chefes de família no mesmo ano Goldani (1984).

Analisando-se ainda as informações de 1985, percebe-se que cerca de 1 milhão de mulheres não tinham qualquer grau de parentesco com a família, número esse cerca de 3 vezes superior ao encontrado para os homens na mesma condição. Presume-se que nesse contingente estão as empregadas domésticas, profissão tipicamente feminina, marcada pela resistência no emprego.

Mesmo com todas estas restrições quanto aos conceitos e à subjetividade das respostas, a Tabela 30 indica que a participação feminina entre os chefes de família, aumentou de 13,0% para 18,2% entre 1970 e 1985. Estes dados só não esclarecem se, o que está ocorrendo, é uma maior consciência na declaração de quem realmente é o chefe, no caso das famílias com união do casal, ou se tem aumentado o número de mulheres separadas, desquitadas, divorciadas, viúvas ou de mães solteiras, como chefes de família.

Um fato objetivo que pode ser destacado é a diminuição relativa do número de chefes de famílias com responsabilidade quase integral no orçamento familiar, na década de setenta. "Essa redução, embora ocorrendo em todas as faixas de rendimento, foi mais significativa nas classes mais baixas, onde são sentidas mais fortemente as pressões econômicas, levando a que maior número de pessoas colaborem no orçamento doméstico, de modo a manter os níveis de subsistência existentes" Silva (1982).

TABELA 30
DISTRIBUIÇÃO PERCENTUAL DAS FAMÍLIAS POR SEXO DO CHEFE, SEGUNDO OS ANOS DE REFERÊNCIA — BRASIL — 1970-1985

ANOS DE REFERÊNCIA	SEXO DO CHEFE		
	Total (%)	Homens (%)	Mulheres (%)
1970	100,0	87,0	13,0
1977	100,0	84,6	15,4
1985	100,0	81,8	18,2

FONTE — IBGE. Amostra de 1% do Censo Demográfico de 1970 e PNADs — 1977 e 1985.

TABELA 31
DISTRIBUIÇÃO DAS PESSOAS RESIDENTES EM DOMICÍLIOS PARTICULARES, POR SEXO, SEGUNDO A POSIÇÃO OCUPADA NA FAMÍLIA — BRASIL — 1985

POSIÇÃO NA FAMÍLIA	DISTRIBUIÇÃO DAS PESSOAS POR SEXO			
	Homens		Mulheres	
	Absoluto	%	Absoluto	%
TOTAL	65 008 756	100,0	66 218 596	100,0
Chefes	26 259 325	40,4	5 837 101	8,8
Cônjuges	66 079	0,1	23 806 664	36,0
Filhos	35 759 504	55,0	31 920 511	48,2
Outros parentes	2 571 164	4,0	3 613 156	5,5
Sem parentesco	352 684	0,5	1 041 164	1,5

FONTE — IBGE. PNAD — 1985.

Portanto, tudo indica que a própria realidade social, no dia-a-dia dos orçamentos domésticos, por exemplo, vem derrubando o mito do homem como chefe da família. Embora a realidade ande sempre mais depressa do que a consciência, é necessário que esta última avance no sentido de mostrar as vantagens da co-gestão familiar e os benefícios de assumir a divisão de responsabilidades.

9 — A MULHER NEGRA

Os últimos dados disponíveis nas pesquisas domiciliares sobre a distribuição da população brasileira por cor, mostram que, em 1982, 56,9% das mulheres eram brancas, 7,2% negras, 35,3% pardas e 0,6% amarelas (Tabela 32)[8].

As análises que têm sido feitas sobre diferenciais sócio-econômico, segundo a cor, revelam, no entanto, que não existem grandes distinções entre negros e pardos, no que diz respeito às suas condições materiais de existência, embora estes estejam em situação bastante desvantajosa com relação aos brancos e amarelos. Grande parte destas seqüelas decorrem da baixa capacidade da sociedade em integrar negros e pardos, desde a época do término da escravidão, em fins do século passado.

TABELA 32
DISTRIBUIÇÃO PERCENTUAL DA POPULAÇÃO BRASILEIRA POR SEXO, SEGUNDO A COR — BRASIL — 1982

COR	DISTRIBUIÇÃO POR SEXO	
	Homens (%)	Mulheres %
TOTAL	100,0	100,0
Brancos	56,0	56,9
Pretos	7,4	7,2
Pardos	35,9	35,3
Amarelos	0,7	0,6

FONTE — IBGE. PNAD — 1982.

Assim, a situação da mulher negra, especialmente no que diz respeito à sua integração no mercado de trabalho, revela-se bastante precária, pois, além de carregar todas as dificuldades de inserção econômica *inerentes à condição feminina*, tem de atravessar todas as barreiras decorrentes de sua origem social e do preconceito.

Tais dificuldades podem ser expressas, em primeira instância, pela posição que ocupam no mercado de trabalho. Enquanto 41,6% das mulheres brancas empregadas não possuíam carteira de trabalho assinada, cerca de 57,3% e 55,8% do contingente feminino negro e pardo, respectivamente, encontravam-se nessa situação, em 1982.

[8] A declaração de cor nas pesquisas domiciliares do IBGE, como a PNAD-82, é feita de forma que o entrevistado escolha sua cor dentro das quatro opções dadas pelo questionário fechado, ou seja, branca, negra, parda ou amarela.

O diferencial, entre empregadas negras sem carteira assinada e de homens brancos na mesma condição (33,4%), chegava a 25 pontos percentuais (Tabela 33).

A situação torna-se ainda mais gritante quando observamos os diferenciais de rendimento para empregados e trabalhadores autônomos, segundo a cor. Neste caso, cerca de 25% dos homens brancos auferiam renda de até 1 salário mínimo, em 1982, enquanto que para as mulheres negras esse percentual alcançava os 71% (Tabela 34).

Diferenciais dessa natureza não precisam de mais palavras para serem qualificados. Revelam, no entanto, que a questão social da mulher atinge proporções dramáticas, quando se refere a determinados segmentos populacionais historicamente excluídos das benesses do desenvolvimento econômico e dos poucos direitos de cidadania social até então existentes.

CONSIDERAÇÕES FINAIS

Grande parte das informações descritas neste trabalho mostram que a dimensão social dos problemas que atingem a mulher brasileira, não nos qualificam a caracterizar a questão feminina como uma questão menor. A abrangência e a transversalidade dos problemas sociais vividos pelas mulheres fazem com que a questão feminina deva ser vista como plural no conjunto das necessidades mais amplas e imediatas do contexto social brasileiro.

TABELA 33
PARTICIPAÇÃO PERCENTUAL DOS EMPREGADOS SEM CARTEIRA DE TRABALHO ASSINADA NO TOTAL DE EMPREGADOS POR SEXO, SEGUNDO A COR — BRASIL — 1982

COR	PARTICIPAÇÃO POR SEXO		
	Total (%)	Homens (%)	Mulheres (%)
TOTAL	48,8	40,5	47,4
Brancos	36,2	33,4	41,6
Pretos	52,1	49,4	57,3
Pardos	52,1	50,5	55,8

FONTE — IBGE. PNAD — 1982.

TABELA 34
DISTRIBUIÇÃO PERCENTUAL DOS OCUPADOS POR COR E SEXO, SEGUNDO CLASSES DE RENDIMENTO MÉDIO MENSAL — BRASIL — 1982

CLASSES DE RENDIMENTO MÉDIO MENSAL	DISTRIBUIÇÃO DOS OCUPADOS POR COR E SEXO					
	Branco		Preto		Pardo	
	Homens	Mulheres	Homens	Mulheres	Homens	Mulheres
TOTAL	100 100,0	100,0	100,0	100,0	100,0	100,0
Até 1 Salário Mínimo	25,0	45,0	47,4	71,1	46,1	70,0
Mais de 1 Salário Mínimo	75,0	55,0	52,6	28,9	53,9	30,0

FONTE — PNAD — 1982.
NOTA — Considerou-se apenas os empregados e os trabalhadores por conta própria. Estão excluídos, assim os empregadores e as pessoas sem declaração de posição, bem como os membros da família que trabalham sem remuneração.

Todas as situações aqui levantadas revelam que, no plano dos direitos sociais, a luta das mulheres é a luta pela prestação de assistência integral a sua saúde; pela definição da maternidade como relevante função social e pela colaboração do estado na manutenção das crianças; pela obrigatoriedade do ensino gratuito pré-escolar e do primeiro grau, garantindo educação indiferenciada entre os sexos; pela obrigatoriedade de criação e manutenção de creches pelas empresas para os filhos de seus empregados. Se observadas com atenção, todas essas causas são justas e universais. Não precisariam que fossem, apenas, bandeiras de movimentos de mulheres. No plano desses direitos, a luta das mulheres não passa da *luta pela defesa dos direitos à vida, à maternidade, ao crescimento dos filhos e à sobrevivência em condições sociais mais dignas.*

No plano das questões relativas ao trabalho, a luta feminina tem-se desenvolvido nas reivindicações por igualdade de direitos entre trabalhadores urbanos e rurais, pelo abono da falta ao serviço, de homens e mulheres, por motivos de doença de filhos menores ou dependentes, pela integração da empregada doméstica como assalariada, com a mesma plenitude de direitos das outras profissões, por direitos iguais aos dos homens na disputa por postos de trabalho; enfim, questões que também não se limitam aos problemas de labor feminino, mas que implicam na necessidade de desenvolver novas formas de funcionamento do mercado de trabalho.

Por fim, no plano dos direitos humanos, as mulheres defendem a garantia de sua integridade física, moral e psicológica e um tratamento mais igualitário e responsável.

Questões como essas e muitas outras não são difíceis de serem solucionadas, embora apontem caminhos para problemas mais profundos, como a grande desigualdade na distribuição da riqueza hoje reinante no Brasil. Um relacionamento mais generoso da sociedade, do estado e dos próprios homens com as mulheres, não impede e até avança para que todos continuem reivindicando uma sociedade mais justa e equânime. Assim, a luta continua, só que com muito mais prazer.

BIBLIOGRAFIA

ALTMAN, Ana Maria Goldani. Informação de família no Censo Demográfico. In: SEMINÁRIO METODOLÓGICO DOS CENSOS DEMOGRÁFICOS, 3., 1984 — Ouro Preto. *Censos, Consensos, Contra-Sensos.* São Paulo, ABEP, s.d.

CANESQUI, Ana Maria. A Saúde da mulher em debate: *Revista Saúde em Debate.* Rio de Janeiro, (15/16): 24-9, fev. 1984.

CARDOSO, Lilibeth Maria; MEDICI, A. C. Educação. In: PERFIL estatístico de crianças e mães no Brasil; características sócio-demográficas, 1970-1977. Rio de Janeiro, IBGE/UNICEF, 1982.

MACEDO, Concessa Vaz de. *O Trabalho feminino e suas possíveis implicações sobre o crescimento demográfico.* Brasília, CEPPD, 1984.

MEDICI, André Cezar; AGUIAR, Marco Antonio de Souza. Planejamento familiar X planejamento governamental. *Revista Saúde em Debate.* Rio de Janeiro (15/16): 20-23, fev. 1984.

_____; _____. *Como entender o mercado de trabalho.* Rio de Janeiro, IBASE, 1985.

_____. Estrutura e dinâmica da força de trabalho médica no Brasil na década de 70. *Revista Brasileira de Administração Pública, 19*(2); 32-74 abr./jun. 1985.

_____. *Considerações sobre a distribuição dos rendimentos da população economicamente ativa no Brasil em 1980.* Rio de Janeiro, IBGE 1982. mimeo.

_____. Notas interpretativas sobre a variável renda nos censos demográficos brasileiros. In: SEMINÁRIO METODOLÓGICO DOS CENSOS DEMOGRÁFICOS, 3., 1984, Ouro Preto. *Censos, Consensos, Contra-sensos.* São Paulo, ABEP, s.d.

_____. Trabalho e rendimento. In: PERFIL estatístico de crianças e mães no Brasil; características sócio-demográficas, 1970-1977. Rio de Janeiro, IBGE/UNICEF, 1982.

OLIVEIRA, Luiz Antonio Pinto de; SIMÕES, Celso Cardoso da Silva. Características sócio-demográficas: In: PERFIL estatístico de crianças e mães no Brasil; características sócio-demográficas, 1970-1977. Rio de Janeiro, IBGE/UNICEF, 1982.

PAIVA, Paulo de Tarso Almeida. Os dados de força de trabalho nos censos demográficos. In: SEMINÁRIO METODOLÓGICO DOS CENSOS DEMOGRÁFICOS. 3; 1984, Ouro Preto. *Censos, Consensos, Contra-sensos. São Paulo, ABEP, s.d.*

_____. *A mulher no mercado de trabalho urbano. In: ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS, 2., 1980, São Paulo. Anais...* São Paulo, ABEP, 1980, v. 2.

SALM, Cláudio Leopoldo. *Escola e trabalho.* São Paulo, Brasiliense, 1982.

SAUVY, Alfred. *Elementos de demografia.* Rio de Janeiro, Zahar, 1979.

SILVA, Rosa Maria Ribeiro da. *Famílias.* In: PERFIL estatístico de crianças e mães no Brasil; características sócio-demográficas, 1970-1977. Rio de Janeiro, IBGE/UNICEF, 1982.

VETTER, David Michael; SIMÕES, Celso Cardoso da Silva. Acesso à infra-estrutura de saneamento básico e mortalidade. *Boletim Demográfico,* Rio de Janeiro, *10*(4):4-29, abr./jun. 1980.

RESUMO

Analisamos as principais mudanças verificadas nos indicadores econômicos e sociais sobre a situação feminina no Brasil durante os últimos 10 anos. Vários aspectos são investigados, tais como características demográficas (idade, nupcialidade, inserção na família), educação, mercado de trabalho e distribuição de renda. Toda a informação estatística é cotejada com a informação sobre a situação masculina.

ABSTRACT

We have analysed the main changes of economic and social situation indicators of women in Brazil over the last ten years. We have investigated several aspects such as demographic characteristics (age, nuptiallity, familly position), education, labour market and income distribution. All the statistical data were compared with the corresponding data for men.

OBSERVAÇÕES SOBRE O CÁLCULO DO ÍNDICE DE GINI (E SUA DECOMPOSIÇÃO) COM DADOS INDIVIDUAIS DA RENDA

José W. Rossi*

1 — INTRODUÇÃO

O índice de Gini é, sem dúvida, a medida mais comumente usada para aferir o grau de concentração na distribuição da renda pessoal. A sua popularidade decorre, em grande parte, da facilidade de interpretação. De fato, o índice é obtido a partir da chamada curva de Lorenz da distribuição, a qual tem, nas abscissas, as frações acumuladas das pessoas (classificadas das mais pobres para as mais ricas) e, nas ordenadas, as correspondentes parcelas da renda recebida por tais pessoas. No diagrama que serve de base para a curva de Lorenz, a linha de 45° indica, é claro, a perfeita distribuição da renda. O índice de Gini é dado então pela razão da "área entre a curva de Lorenz e a linha de 45°" e a "área sob a linha de 45°".

Nestas notas apresentamos as fórmulas alternativas usadas no cálculo do índice de Gini para dados individuais da renda, discutindo alguns aspectos de uma formulação recentemente proposta por Shalit (1985).

2 — O ÍNDICE DE GINI E SUA DECOMPOSIÇÃO

Há várias fórmulas alternativas para o cálculo do índice de Gini com dados individuais da renda. Para definir essas medidas, suponha-se uma população com n indivíduos; denote o indivíduo por i (i = 1, 2, ..., n). Sejam as rendas dos indivíduos ordenadas, como segue: $y_1 \leqslant y_2 \leqslant \ldots \leqslant y_n$, cuja renda média é $\bar{x}$. Seja ainda F_i a parcela acumulada dos indivíduos ordenados dos mais pobres para os mais ricos, e Q_i a correspondente parcela acumulada da renda, com $F_o = Q_o = 0$; isto é, $F_i = \frac{i}{n}$ e $Q_i = \frac{1}{n\bar{x}} \sum_{k=1}^{i} y_k$, para i = o, 1, ..., n. As seguintes fórmulas alternativas para o índice de Gini podem ser então estabelecidas (ver Anand – 1983):

$$G_1 = 1 - \sum_{i=o}^{n-1} (F_{i+1} - F_i)(Q_{i+1} + Q_i), \qquad (1)$$

$$G_2 = \sum_{i=1}^{n-1} (F_i Q_{i+1} - F_{i+1} Q_i), \qquad (2)$$

*Do IPEA — Instituto de Planejamento Econômico e Social — e da UFRJ.

$$G_3 = \frac{1}{2n^2\bar{x}} \sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j| =$$

$$= \frac{1}{n^2\bar{x}} \sum_{i=1}^{n} \sum_{j \leq i} (y_i - y_j), \quad (3)$$

$$G_4 = \frac{n+1}{n} - \frac{2}{n^2\bar{x}} \sum_{i=1}^{n} (n+1-i)y_i, \quad (4)$$

$$G_5 = \frac{2}{n\bar{x}} \text{Cov}(y_i, i) \quad (5)$$

Com base na fórmula (5), Shalit (1985) propôs que se estime primeiramente a regressão

$$x = a + bi + u, \quad (6)$$

para obter-se então

$$\hat{b} = \text{Cov}(x,i)/V(i), \quad (7)$$

onde V significa variância. Como a variância dos n primeiros números naturais é $V(i) = (n^2 - 1)/12$, Shalit sugere que o índice de Gini seja calculado a partir de

$$G = \frac{(n^2-1)}{6n} \cdot \frac{\hat{b}}{\bar{x}} \quad (8)$$

Esta última fórmula só é válida, porém, se não houver qualquer coincidência nos valores das rendas dos indivíduos. Freqüentemente, contudo, este não é o caso, particularmente no Brasil, onde se sabe que uma boa parcela da população recebe o salário mínimo, e às vezes não tem renda de qualquer espécie. Nestas circunstâncias, a sugestão é que, em casos de empate (conforme proposto por Pyatt et alii – 1980), a todos os indíviduos (com a mesma renda) seja dada a média dos *rankings* que se obteria caso houvesse uma pequena diferença entre estes. Em seguida, calcular-se-ia o índice de Gini de acordo com a fórmula (5), onde **i** não mais seria o conjunto dos números naturais. É claro que em tais circunstâncias não se poderá usar a fórmula (8), e se desejar proceder como em (6) e (7), o índice de Gini seria, então:

$$G = \frac{2\hat{b}V(i)}{n\bar{x}}, \quad (9)$$

notando-se, uma vez mais, que **i** não mais seria o conjunto dos números naturais até n.

Quanto à decomposição do índice de Gini, segundo os seus fatores componentes (ou fontes de renda), dois procedimentos alternativos podem ser adotados, conforme passamos a mostrar. Suponha-se, inicialmente, que a renda total do indivíduo seja decomposta em k componentes (por exemplo, renda do trabalho, renda de aluguel, pensão, etc.). Desta forma, seguindo o procedimento já exposto, o índice de Gini para a renda do fator k seria calculado por

$$G_k = \frac{2}{n\bar{x}_k} \text{Cov}(x_k, i') \quad (10)$$

onde **i'** segue agora a ordenação da renda do fator **k** (isto é, $(x_{k1} \leq x_{k2} \leq \ldots \leq x_{kn})$, que difere, pois, de **i**, que segue a ordenação da renda total; $\bar{x}_k$ é a renda média do fator **k**, e **n** é o número total de indivíduos, que é, obviamente, idêntico àquele da fórmula (5). É útil definir, também aqui, o chamado índice de concentração do fator **k**, que é dado por (Pyatt et alii – 1980):

$$C_k = \frac{2}{n\bar{x}_k} \text{Cov}(x_k, i) \quad (11)$$

Isto é, a covariância é agora calculada entre a renda do fator **k** e o *ranking* do indivíduo, segundo sua renda total.

Com essas considerações, o índice de Gini para a distribuição da renda total seria (Pyatt et alii — 1980):

$$G = \sum_k \alpha_k C_k, \quad (12)$$

ou alternativamente

$$G = \sum_k \alpha_k \Theta_k G_k, \quad (13)$$

onde α_k é a renda do fator **k** como proporção da renda total, e $\theta_k = C_k/G_k$.[1]. Isto é, G não é uma média ponderada dos vários G_k.

1 – Ressalte-se aqui que Shalit (1985) decompõe o índice de Gini como $G = \sum_k (\alpha_k)^2 \theta_k G_k$, o que é, todavia, incorreto. A razão para este erro é que Shalit define a renda total com $x = \Sigma \alpha_k x_k$, ao invés da forma correta $x = \Sigma x_k$. Para considerações adicionais sobre a decomposição do índice de Gini, segundo as distintas fontes de renda, ver Rossi (1982) ou (1983).

Ilustramos, a seguir, o uso das fórmulas (5) e de (10) a (12) com uma aplicação a dados de renda de uma amostra aleatória de 18.240 famílias selecionadas do Censo Demográfico Brasileiro de 1980. A renda total da família foi devidamente decomposta, segundo os sete fatores-componentes mostrados na Tabela 1. Note-se, primeiramente, que o índice de Gini com valor próximo a 0,99 para as rendas dos fatores *aposentadoria, pensão e, propriedade e capital* significa, é claro, que poucas das 18.240 famílias percebem renda destes fatores. Outra observação é que o uso da fórmula (12) ou da (13) resulta, efetivamente, no mesmo índice de Gini para a renda total que aquele obtido para todas as rendas[2].

Um último ponto que merece ser notado é que os índices de concentração (C_k) são aqui sempre menores que os correspondentes índices de Gini (G_k), o que não constitui mera coincidência. Observe-se que $\theta_k = C_k/G_k = Cov(x_k,i)/Cov(x_k,i')$. Como a covariância da renda de um fator com o próprio *ranking* deste fator é necessariamente maior do que a covariância desta mesma renda com o *ranking* da renda total (ou *ranking* de qualquer outra renda), segue-se que o índice de concentração de um dado fator será, no máximo, igual ao índice de Gini deste mesmo fator.

Para concluir, estas notas têm o propósito, apenas, de discutir alguns aspectos relacionados ao uso de certas fórmulas no cálculo do índice de Gini com dados desagregados. Assim, a aplicação aqui efetuada serve, somente, para ilustrar as idéias apresentadas. É claro que as técnicas discutidas poderiam ser utilizadas para analisar, por exemplo, o comportamento intertemporal da distribuição de renda com os dados anuais da PNAD. Esta questão extrapola, porém, os nossos objetivos aqui, merecendo, de fato, pesquisas à parte.

TABELA I

ÍNDICE DE GINI E ÍNDICE DE CONCENTRAÇÃO COM BASE EM DADOS DE 18.240 FAMÍLIAS DO CENSO DEMOGRÁFICO BRASILEIRO DE 1980

Renda dos Fatores-Componentes	Índices de Desigualdade	
	Gini (G_k)	Concentração (C_k)
Ocupação Principal	0,4064	0,3386
Venda de Produtos	0,9429	0,6828
Aposentadoria	0,9852	0,4502
Aluguel	0,8353	0,3587
Pensão	0,9978	0,2837
Propriedade e Capital	0,9991	0,3586
Outras Fontes	0,5037	0,4498
Todas as Rendas	0,4014	0,4014
Média Ponderada		0,4014*

* — De acordo com a fórmula (12), cujo valor coincide com aquele calculado pela fórmula (5).

BIBLIOGRAFIA

ANAND, S. *Inequality and Poverty in Malaysia: measurement and decomposition,* Oxford University Press, 1983.

PYATT, G. ; CHEN, C. ; FEI, J, The Distribution of Income by Factor Components. *Quarterly Journal of Economics.* v. XCV, n.° 3, nov. 1980, p. 451-73.

ROSSI, J. W. Índices de Desigualdade de Renda e Medidas de Concentração Industrial: aplicação a casos brasileiros. Zahar, 1982.

_____. Decomposição Funcional do Índice de Gini com Dados de Renda do Brasil. *Revista Brasileira de Economia,* Rio de Janeiro, v. 37, n.° 3, jul./set. 1983, p. 337-48.

SHALIT, H. Calculating the Gini Index of Ineguality for Individual Data. *Oxford Bulletin of Economics and Statistics.* v. 47, n.° 2, may 1985, p. 185-89.

2 – Por razões óbvias o índice de concentração para todas as rendas é igual ao índice de Gini dessas mesmas rendas.

RESUMO

Nestas notas apresentamos várias fórmulas alternativas para o cálculo do índice de Gini com dados individuais da renda, dando atenção especial a uma formulação recentemente proposta na literatura. Considerações são, também, efetuadas quanto à decomposição do índice segundo as várias fontes de recebimento de renda. Uma aplicação com dados da renda do Censo de 1980 é então efetuada para ilustrar os conceitos apresentados.

ABSTRACT

In this article we present several alternative expressions for the computation of Gini's index for individual income data, paying special attention to an approach recently proposed in literature. Consideration has also been made as to the index decomposition according to the many sources of income. An application to data of 1980 Census is then made to illustrate ideas presented.

INSTRUÇÕES BÁSICAS PARA PREPARO DOS ORIGINAIS

Os originais entregues para publicação devem obedecer as seguintes normas:

1 — Texto datilografado em papel branco tamanho ofício, em um só lado, em espaço duplo, com margem de 3 cm, sem rasuras ou emendas que dificultem sua leitura e compreensão.
As laudas deverão ser numeradas seguidamente, comportando até 72 batidas por linha e com 30 linhas por páginas.
Obs.: texto oriundo de autores do IBGE será datilografado em lauda-padrão fornecida pelas Diretorias.
Devem ser remetidas 02(duas) vias do trabalho;

2 — A primeira página do original (folha de rosto) deve conter título, nome completo do(s) autor(es), qualificação profissional, com indicação das atividades exercidas, dos órgãos a que estão vinculados, do endereço para correspondência, bem como colaboradores, agradecimentos e auxílios recebidos;

3 — O título deve ser conciso, específico e descritivo, registrando as palavras-chave que representem o conteúdo do artigo;

4 — Os artigos devem ser acompanhados de um resumo informativo, de modo a expressar seus pontos relevantes, datilografados em espaço duplo e folha separada, em português e inglês, contendo, aproximadamente, 200 palavras;

5 — As notas explicativas devem ser numeradas numa seqüência única e datilografadas em folhas separadas, com indicação dos números respectivos;

6 — As tabelas, inseridas nos textos, devem ser apresentadas em folhas separadas e precedidas de títulos que permitam perfeita identificação dos dados, com registro dos correspondentes números de ordem, nos locais de inserção;

7 — No caso de listagens e tabelas extensas, e de outros elementos de suporte, podem ser empregados apêndices;

8 — As fórmulas matemáticas devem ser apresentadas com clareza, para evitar problemas de interpretação;

9 — Não devem ser utilizadas reproduções de ilustrações elaboradas através do sistema *"plotter"*;

10 — As fotografias devem ser nítidas, em preto e branco, contrastadas, de preferência em tamanho 6 × 9 cm, nunca superior a 12 × 18 cm; os gráficos desenhados a nanquim, em papel branco ou vegetal; os dados e dizeres que acompanham os desenhos, em letra de forma; as legendas das ilus-

trações, datilografadas em folhas separadas e numeradas de acordo com a figura respectiva, com indicação no texto, pelo número de ordem, dos locais de inserção das figuras e, ainda, menção da fonte e permissão para reprodução, quando já houverem sido publicadas;

11 — O formato de impressão máximo de encartes estabelecido para os documentos cartográficos é de 50 × 55 cm. Sempre que haja redução ou ampliação do documento cartográfico original, deverá constar deste apenas a escala gráfica.

O desenho original deve ser feito em material estável. No caso de documentação cartográfica de precisão ou, quando a densidade de informações contidas num mapa ou cartograma dificulte a sua leitura, será excepcionalmente permitida a impressão em cores. Em caso contrário, os valores cor serão substituídos por hachuras, retículas ou símbolos gráficos compatíveis com a escala.

Os documentos cartográficos devem ser precedidos de títulos que permitam perfeita identificação e em suas legendas devem constar: classificação, nomes ou siglas das Unidades da Federação representadas, ano da publicação, escala, projeção (exceto nos cartogramas) e as convenções cartográficas menos conhecidas.

A documentação cartográfica utilizada, com o nome ou sigla da fonte e outros elementos complementares compatíveis à escala, devem ser descritos de modo sucinto. No caso de mapas e cartogramas deve existir flexibilidade na disposição dos títulos, legendas e outras referências, utilizando-se os espaços vazios oferecidos pelo próprio desenho. Deve ser estabelecida uma graduação de importância adotando-se diferentes tamanhos de tipos nos dizeres da legenda.

A moldura, em torno do desenho de um mapa ou cartograma, deve garantir uma margem no papel. Para as cartas pertencentes ao mapeamento sistemático, devem ser obedecidas às normas e especificações inerentes a cada carta, de acordo com a escala e classificação (contatar com o órgão responsável por esse mapeamento ou com a Comissão de Cartografia).

As cartas, mapas ou cartogramas, inseridos ou anexados, devem ser referenciados no texto por um número de ordem correspondente.

As legendas e outras referências devem estar destacadas do desenho e afastadas das margens. No caso de cartas do mapeamento sistemático, ver as normas e especificações de cada tipo de escala.

É aconselhável que para a elaboração de uma base precisa sejam utilizados os documentos cartográficos realizados pelo IBGE ou por outros órgãos integrantes do Sistema Cartográfico Nacional;

12 — As citações bibliográficas no texto devem ser feitas de acordo com o Projeto ABNT 14.01.01.005 — Apresentação de citações em documentos;

13 — As referências bibliográficas devem ser numeradas em seqüência única e apresentadas em folhas separadas com indicação dos números respectivos. Devem ser redigidas segundo a norma brasileira respectiva (ABNT — NBR — 6023 Referências Bibliográficas), contendo indicação por extenso dos títulos dos periódicos, quando se tratarem de referências de artigos. A exatidão e adequação das referências a trabalhos consultados e mencionados no texto são de responsabilidade do autor;

14 — Quando houver necessidade de dividir o trabalho em capítulos, seções e partes, esses devem ser numerados, progressivamente, de acordo com as normas da ABNT (NB – 69).