

MI/277.13

E12

doc. 6.1

IBGE/DIRETORIA DE PESQUISAS E INQUÉRITOS

DEPARTAMENTO DE POPULAÇÃO

DIVISÃO DE PESQUISAS

APLICAÇÃO DA AMOSTRAGEM

NOS

CENSOS DEMOGRÁFICOS

Tema 4

Luiz Nery da Costa

APRESENTAÇÃO

Vários países utilizam a técnica de amostragem nos trabalhos censitários, dentre eles destacam-se os E.U.A., o Canadá e a Argentina, cujos censos de população guardam semelhanças com os censos demográficos realizados pelo Brasil, durante o período 1960/80.

Tendo participado desses três censos, e considerando a escassez de literatura específica sobre o assunto, resolvi reunir neste documento, notas técnicas e metodológicas sobre a matéria, inclusive examinando a experiência dos países citados, a fim de facilitar o entendimento e o debate da mesma.

Ao preparar este documento, iniciado em 1986, por iniciativa própria, algumas observações foram emitidas, as quais são de inteira responsabilidade do autor e refletem seu pensamento nesta área.

SUMÁRIO

1.	INTRODUÇÃO	02
2.	APLICAÇÃO DA AMOSTRAGEM NA COLETA DO CENSO	02
2.1.	- Razões do Uso da Amostragem	02
2.2.	- Planejamento de Amostragem	03
2.2.1	- Definição dos itens da investigação	04
2.2.2	- Plano de divulgação dos resultados	05
2.2.3	- Uso de dois ou mais questionários	06
2.2.4	- Critérios para inclusão dos quesitos nos questionários	07
2.2.4.1	- Critérios para a enumeração completa	07
2.2.4.2	- Critérios para a investigação por amostragem	08
2.2.5	- Precisão do Censo	10
2.2.5.1	- Ordem de grandeza do erro de amostragem	11
2.2.6	- Unidade Básica de Amostragem	14
2.2.6.1	- Opção pelo tipo de unidade de amostragem	15
2.2.7	- Tamanho da amostra	18
2.2.7.1	- Reflexões para os futuros censos	19
2.2.8	- Processo de Seleção da Amostra	21
2.2.9	- Processos de Estimacão	24
2.2.9.1	- Estimativa de razão	26
2.2.9.2	- "Ratio Estimation Groups"	28
2.2.9.3	- "Iterative Ratio Estimators Groups"	30
2.2.9.4	- "Raking Ratio Estimators Procedure"	34
2.2.9.5	- Apreciação dos processos examinados	37
2.2.10	- Diretrizes para a Construção do Processo de Estimacão	39
2.2.10.1	- Requisitos do processo de estimacão	40
2.2.10.2	- Redução da variabilidade	41

2.2.10.3 - Controle das tendenciosidades e das flutuações de amostragem	41
2.2.10.4 - Uso da estratificação e pós-estratificação	42
2.2.10.5 - Consistência	44
2.2.10.6 - Nível geográfico de estimação - área de ponderação	45
2.2.10.7 - Estratégia de ponderação	46
2.2.10.8 - Uso de duas ou mais matrizes de ponderação	47
2.2.11 - Cálculo e apresentação dos erros	48
2.2.11.1 - Erros alheios e amostragem	48
2.2.11.2 - Erros de amostragem	49

ANEXO	52
REFERÊNCIAS	54

APLICAÇÃO DA AMOSTRAGEM NOS CENSOS DEMOGRÁFICOS

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. Praticamente neste período foram editados os principais livros exclusivamente de amostragem, e pontificaram os estudos científicos da eficácia dessa técnica.

Devido às crescentes necessidades de informações estatísticas para o planejamento, com redução ou limitação no custo da realização de um Censo e ainda a obtenção de resultados com mais rapidez, a adoção da amostragem, nas tarefas pertinentes a um Censo, vem sendo ampliada e 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, 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 vem sendo cada vez mais sofisticados.

A literatura existente sobre a utilização da técnica de amostragem nos Censos é relativamente 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, aperfeiçoá-los.

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

2. APLICAÇÃO DA AMOSTRAGEM NA COLETA DO CENSO

2.1. 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 de um modelo de questionário, preenchido para cada domicílio.

No Censo Demográfico de 1960 utilizou-se, pela primeira vez, a técnica de amostragem na coleta das informações. Desde então o uso da amostragem nos trabalhos censitários vem sendo empregado em maior escala.

A redução nos custos da coleta e de outras fases dos trabalhos censitários, a possibilidade de obtenção dos resultados com maior rapidez e precisão, por se tratar de menor massa de dados, constituem as principais razões do uso da amostragem na coleta do Censo (1).

Sem dúvida, o apoio técnico propiciado pela U.S.A.I.D, ao IBGE contribuiu de modo expressivo para a aplicação da amostragem no Censo de 1960.

É oportuno salientar que em outros países, no passado, alguns censos foram realizados por amostragem devido a impossibilidade de uma enumeração completa decorrente da falta de recursos, carência de pessoal, condições materiais deficientes, etc. (2). A situação contemporânea é diferente, pois esses fatores impeditivos não são predominantes, apenas procura-se otimizá-los dentro de uma concepção tecnológica mais avançada.

2.2. PLANEJAMENTO DE AMOSTRAGEM

A adoção do uso da amostragem na coleta do Censo requer que sejam apreciados diversos pontos que influem nos resultados finais, dentre eles destacam-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 do Censo
- . Unidade básica de amostragem
- . Tamanho da amostra
- . Processo de seleção da amostra
- . Processos de estimação
- . Diretrizes para a construção do processo de estimação
- . Cálculo e apresentação dos erros

A seguir são elaboradas considerações a respeito de cada ponto assinalado anteriormente, onde a ordenação dessas considerações não necessita ser seguida tal como se apresenta.

2.2.1. Definição dos itens da investigação

A formulação dos itens a serem investigados nos Censos visam a atender às necessidades nacionais e às recomendações e princípios apontados pelas Nações Unidas, a fim de possibilitar estudos comparativos nacionais e internacionais 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 conhecer três grandes conjuntos de características: de pessoas, de famílias e de domicílios ou habitação.

Inúmeros tópicos são estudados, dentre eles destacam-se as investigações das características sobre: condição 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. Em 1960 foram indagados 24 quesitos de pessoas, em 1970 a investigação cresceu para 32 perguntas e, finalmente, atingiu a 57 itens em 1980.

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.

Maiores detalhes sobre a investigação censitária de 1980 são encontrados no volume "Metodologia do Censo Demográfico de 1980" (3).

2.2.2. Plano de 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 e pelo volume de tabelas cruzadas.

Implicitamente contidos ou ligados aos objetivos do Censo, a maioria das tabelas é concebida para apresentar resultados a nível das Unidades da Federação, outras a nível de Município e algumas a nível de Distrito.

Por outro lado, as tabelas elaboradas muitas vezes apresentam cruzamentos de diversos itens pesquisados, algumas delas com detalhamentos bastante acentuados.

A dificuldade de investigação de certo item, associada ao nível de detalhamento geográfico, e ainda ao grau de cruzamento previsto no plano tabular, deve ser adequadamente ponderada na fase do planejamento da operação censitária, em função da qualidade e precisão dos resultados, principalmente quando obtidos a partir da amostra.

Algumas tabelas são elaboradas, cruzando informações, para subsidiarem estudos enfocados a nível nacional ou quando muito a nível regio

nal. Não há razão de concebê-las a nível de detalhamento menor, pois não possibilitarão decisões político-administrativas a nível de áreas menores do que as Unidades da Federação.

Entretanto, muitas outras tabulações devem ser concebidas a nível de Município para auxiliarem no planejamento regional e muito poucas a nível de 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 ser atendido com o uso da amostragem.

2.2.3. 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 itens e todos os demais, são aplicados apenas nas unidades selecionadas para a amostra.

Dessa forma a investigação de poucos quesitos 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, 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 à amostras distintas.

Quando o elenco de quesitos da investigação censitária for grande ou envolvendo muita complexidade, sendo necessária 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 frequentes quando o número de quesitos do questionário é grande.

2.2.4. 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 pessoas 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 quesitos componentes da investigação censitária, o estabelecimento do nível de detalhamento geográfico dos resultados e do grau de desagregação das características investigadas, com o uso de muitas tabulações cruzadas, necessários a atender aos objetivos do Censo, devem, então, ser definidos os critérios de inclusão dos quesitos nos questionários a serem utilizados na Coleta do Censo.

Por ocasião do planejamento dos Censos Demográficos de 1960 a 1980, o assunto foi devidamente apreciado e levando-se em consideração a experiência do "Bureau of the Census" (4) e (5), pode-se delinear regras gerais para a enumeração completa dos quesitos e para a investigação por amostragem.

2.2.4.1. Critérios para a enumeração completa

As características que devem figurar obrigatoriamente em todos os questionários do Censo, portanto, constituem os itens básicos, e devem atender algumas condições, principalmente as que se destinam, por exemplo:

- 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 servem 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, as variáveis: sexo, idade, condição de presença, condição de presença no domicílio, situação e tamanho do domicílio (número de pessoas recenseadas) foram utilizadas para esse fim.
- 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.

2.2.4.2. Critérios para a investigação por amostragem

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 devem considerar:

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

No Censo 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 um código de 7 dígitos.

b) Características sujeitas a grandes variabilidades de respostas.

A investigação das características sobre ocupação, rendimento e nível de instrução apresentaram em 1970 os maiores índices de inconsistência (6). Estas variabilidades, alheias ao processo de amostragem, apresentam erros bastante superiores àqueles decorrentes da amostragem.

c) Características difíceis de investigação

Os resultados do Censo de 1980 sobre a mortalidade das pessoas moradoras no domicílio, e 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, a investigação dessas ocorrências foi insatisfatória, revelando, possivelmente, dificuldades na sua investigação.

A escolha dos temas a serem selecionados para a investigação por amostragem devem considerar: a dificuldade da codificação; a complexidade da crítica; a variabilidade das respostas, principalmente em função dos erros alheios à amostragem, a extensão da seqüência dos quesitos pertinentes à caracterização do assunto estudado, como é o caso do Migração, Fecundidade, Ocupação, Rendimento, etc. Os quesitos portadores de uma dessas observações devem ser investigados por amostragem, exceto, quando por algum motivo sejam indispensáveis ao processo de expansão, como é o caso da investigação da idade.

2.2.5. 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 e, isto é possível com a utilização da amostragem na coleta. As Pesquisas de Avaliação dos Censos de 1970 e 1980 propiciaram estimativas preliminares das taxas de omissão da população da ordem de 7% e 3% respectivamente.

Enquanto os erros alheios à amostragem, quase sempre são incontornáveis, 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 a 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, tais fatores determinam os erros de amostragem.

Portanto, não é demais esclarecer que os erros de amostragem podem ser reduzidos, rigorosamente previstos e controláveis, o que não ocorre em relação às tendenciosidades, que podem ser deliberadas e imprevisíveis no planejamento, com erros maiores, geralmente sem medidas objetivas e algumas vezes desconhecidas as suas causas e ocorrências, aqui no Brasil, não obstante a existência de alguns estudos já realizados (6) e (7).

2.2.5.1 Ordem de grandeza do erro de amostragem

A estimação de qualquer parâmetro θ (total, média, proporção, etc.) é realizada com uma margem de flutuação, dentro de intervalo de confiança, obtido a partir do erro de amostragem. Esse erro, previsto por ocasião do planejamento do Censo, pode ser absoluto ou relativo, definidos por:

$$E = Z \sqrt{V^2 \{\theta\}} \quad \text{ou} \quad E_r = Z \sqrt{V_r^2 \{\theta\}} \quad (1)$$

onde

E = erro absoluto previsto

E_r = erro relativo previsto

Z = múltiplo do desvio padrão, inerente ao coeficiente de confiança

$V^2 \{\theta\}$ = variância do parâmetro θ

$V_r^2 \{\theta\}$ = variância relativa do parâmetro θ

Os valores de Z , provenientes da tabela $N(0;1)$, mais usuais são:

Z	Áreas da Normal %	Coeficiente do $(1-\alpha)$
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 tamanho n , são selecionadas cem vezes, resultam 100 intervalos de confiança, dos quais 95% deles devem conter o parâmetro θ .

Substituindo-se em (1) o valor de $V^2\{\theta\}$, pela sua forma definidora, chega-se a expressão que possibilita o cálculo do tamanho n da amostra.

O tamanho da amostra depende, principalmente, da variância 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 $V^2\{\theta\}$ varia conforme o modelo de amostragem seguido e o processo de estimação.

No Censo são pesquisadas inúmeras variáveis e 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 simples, pois, aproximadamente as aplicações da amostragem na coleta do Censo são equivalentes a esse modelo.

A seguir são apresentadas a ordem de grandeza dos erros absolutos e relativos de amostragem esperados, concernentes a diversos tamanhos de universos e da amostra (1), no presente caso utilizamos a fração de amostragem $f = \frac{n}{N}$, como alternativas de n .

Erros absolutos de amostragem na estimação de alguns totais por diferentes tamanhos de amostra, segundo alguns tamanhos de universo:

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%.

Erros relativos de amostragem na estimação de alguns totais, por diferentes tamanhos de amostra segundo alguns tamanhos de universos:

Tamanho do Universo	Tamanho 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%.

A apreciação dos valores inseridos nessas duas tabelas 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 em 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.

As diversas alternativas de tamanho de amostras, com as diferentes frações de amostragem, foram apreciadas por ocasião do planejamento do Censo de 1980, maiores detalhes sobre o assunto podem ser obtidos em (1) e (5).

2.2.6. Unidade Básica de Amostragem

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

Os Estados Unidos da América empregou, pela primeira vez, a amostragem na coleta de dados do Censo de População de 1940; desde então os Censos de População e Habitação passaram a utilizar cada vez mais a amostragem nos trabalhos censitários. Neste Censo 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 (2).

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 (4).

O Canadá, desde o Censo de População de 1971 até o último, em 1981, tem aplicado a amostragem como principal método de coleta censitária. A unidade básica de amostragem adotada foi o domicílio (8) e (9).

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.

Todos os domicílios existentes nos segmentos censitários, selecionados aleatoriamente para a amostra de áreas, preencheram 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 de 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 na coleta, em 1960, 1970 e 1980, foi utilizado o domicílio como unidade de amostragem (em domicílios particulares).

2.2.6.1. Opção pelo tipo de unidade de amostragem

O assunto merece cuidadosa apreciação sobre 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 dos Censos de cada País.

Quando a investigação censitária objetiva conhecer as características das pessoas, dos domicílios e das famílias, tudo torna-se mais simples de se obter em se adotando o domicílio como unidade de amostragem, incluindo na amostra todas as pessoas nele recenseadas, ao invés da pessoa. 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 1950. Destaque-se, ainda, que a correlação intraclasse do sexo das pessoas 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 (11). 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 podem ser adotados em Países que disponham de condições para tal, como é o caso dos Estados Unidos, Canadá, etc.

Nos Países menos desenvolvidos, com menor recursos técnicos e materiais, é uma temeridade a auto-enumeração, com distribuição dos questionários previamente pelo Correio. 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 (3). Apesar disso, não é demais acres

centar alguns aspectos adicionais. A adoção de uma unidade conglomerada visa, principalmente:

- a) maior facilidade operacional, com aplicação de um único modelo de questionário;
- b) maior rapidez na coleta, talvez em um ou poucos dias;
- c) melhor qualidade das informações coletadas.

Além disso, a precisão das informações coletadas no Censo dependem 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 questionários, 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 aumenta 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, sem influências políticas; treinamento mais adequado, controle e acompanhamento da coleta com maior rigidez, etc.

No Censo Experimental Demográfico de 1980, 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 foram substancialmente melhores do que os resultados do Censo de 1970, para a zona urbana do Estado de São Paulo (13).

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 intra-classe em 0,3 a 0,4 para a renda.

2.2.7. Tamanho da Amostra

Teoricamente o tamanho da amostra é obtido a partir da expressão (1) indicada em 2.2.5.1., explicitando-se o valor de n , contido em $\sqrt{2} \{ \theta \}$

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 2.2.5.1. 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 e de tamanhos de universos.

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, como porcentagens do tamanho da amostra.

Os Estados Unidos utilizaram a fração de amostragem de 1/20 no Censo de População de 1940, posteriormente às frações de amostragens básicas foram 1/5, 1/4, 1/5, respectivamente nos Censos de 1950, 1960, 1970. Em 1980, nas áreas referentes a menor divisão civil, com menos de 2500 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, para a menor carga de investigação junto ao público (8).

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 amostragem menores e esquemas de levantamento dos dados diferenciados, com o intuito de reduzir os custos de operação censitária, maiores detalhes estão em (1) e (3).

É oportuno registrar a aplicação de amostragem na realização do Censo Escolar de 1977, pelo Governo do Estado do Espírito Santo. Este levantamento na realidade, constituiu 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 (14).

2.2.7.1. 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 têm havido uma tendência de aumentar o número de quesitos na investigação por amostragem, sobrecarregando o tamanho do questionário, com graves consequências 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 a 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 a 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, a nível de cada classe ou categoria, conforme é esperado e se depreende das tabelas contidas em 2.2.5.1.

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 é inferior 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 (1) e (3), 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 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 Micro áreas foram efetivados. No Censo de 1980 ocorreram várias tendenciosidades na seleção da amostra e na cobertura,

não obstante o processo já viesse sendo repetido por três vezes seguidas.

2.2.8. 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 correspondente 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 e, estabelecidas de forma a quebrar ou interromper possíveis seqüências cíclicas do universo, de modo a obter bastante variações nas séries sistemáticas de seleção, evitando-se 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 atribuído, aleatoriamente, uma das quatro letras, os domicílios assinalados com a letra atribuída a ele foram selecionados para a amostra (4).

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 atribuído a cada recenseador, um início aleatório, como ponto de partida da seleção sistemática, de 1 em 5 domicílio. (5).

O Canadá também utilizou um processo de seleção sistemático na seleção dos domicílios para amostra nos Censos de População e Habitação de 1971, 1976 (9).

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

O Brasil também utilizou um processo de seleção sistemático 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 dos Estados Unidos, usado em 1960.

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 em 1980, com maior intensidade, ocorreu uma grave tendenciosidade na coleta, principalmente nas áreas menos urbanizadas. Essa tendenciosidade constitui na inclusão de domicílios com menor número de pessoas nas linhas de amostra.

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, verificados 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 às 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 eliminou essa falha de coleta, em consequência, 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, 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 eliminar 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; além disso, 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, mesmo que intencionalmente, 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.

2.2.9. 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 de 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 das estimativas a nível do universo, de todos os itens da investigação censitária, principalmente daqueles investigados por amostragem, separadamente, ou de forma cruzada.

O conhecimento das características básicas de todas as unidades da população, possibilita o uso do processo de estimativa de razão com pós-estratificação, aumentando a eficiência do processo, tornando mais consistente as estimativas das características básicas com os respectivos valores da apuração do universo, trazendo maior credibilidade aos resultados. (15)

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 a 24,7%.

Face aos recursos de informações disponíveis e as particularidades de a amostra efetivamente selecionada, o processo de estimação deve ser concebido de modo a neutralizar a variação ocorrida na fração da amostragem, bem como, controlar 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 atendem ao seguinte: controlar ou neutralizar as tendenciosidades ocorridas na seleção da amostra; redução na variabilidade das estimativas; aumento da consistência entre os resultados obtidos na apuração do universo e as estimativas provenientes da amostra; permita o cálculo de estimativas razoavelmente precisas do erro de amostragem e o custo de processamento seja econômico. Esses pontos, considerados pelos Estados Unidos na escolha do processo, seguido em 1970, hoje são 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, relata-se, a seguir, quais os processos de estimação já seguidos pelos Estados Unidos, Canadá, Argentina e Brasil.

Os Estados Unidos, no Censo de 1950, utilizou 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 foram utilizados o processo iterativo de estimativas 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, semelhante ao dos Estados Unidos, para aquele ano, nas tabulações ini-

cialmente estabelecidas e efetivadas para 11 unidades da federação. Posteriormente, para concluir a apuração do Censo, foi utilizada 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 estimativas de razão por grupos com pos-estratificação 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.

2.2.9.1. Estimativa de Razão

De modo geral, a estimativa do total da característica X_h obtida pelo uso da estimativa da razão, é expressa por:

$$\hat{X}_h = \sum_i \frac{\hat{X}_i}{\hat{Y}_i} \cdot Y_i \quad (1)$$

onde,

\hat{Y}_h - estimativa de razão de característica X_h ;

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

\hat{Y}_i - estimativa simples de Y_h , no i -ésimo grupo baseado nas informações do universo e,

Y_i - apuração completa de Y_h , para todas as unidades do universo, no i -ésimo grupo.

Como, em (1) as estimativas simples resultam dos totais da amostra de X_h e Y_h multiplicados pelo inverso da fração de amostragem, em cada grupo i , então (1) pode ser assim representado:

$$\hat{X}_h = \sum \frac{\bar{Y}_{hj}}{\bar{y}_{hj}} \cdot x_{hi} \quad (2)$$

onde,

x_{hi} - total de característica x_h , na amostra, para o i -ésimo grupo

y_{hi} - corresponde ao total de unidades da amostra que no universo pode possuir a característica y_h para o grupo i -ésimo.

Y_{hi} - total de unidades que, no universo, possui característica h , para o i -ésimo grupo.

Note-se que $\frac{Y_{hi}}{y_{hi}} = \frac{N_{hi}}{n_{hi}}$, portanto dá origem ao peso

so a ser atribuído a cada unidade de amostragem.

Genericamente, o estimador do total de certa característica Y_h , em cada grupo ou célula, é definido por:

$$\hat{Y}_h = \sum w_{hj} \cdot y_{hj} \quad (3)$$

onde,

\hat{Y}_h - estimador da característica Y_{hj} .

w_{hj} - peso atribuído a j -ésima unidade da amostra.

y_{hj} - valor da característica h , na j -ésima unidade da amostra.

Observa-se em (3) que o total estimado Y_h corresponde a 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.) Y_h corresponde à soma dos produtos do peso de cada unidade da amostra que possua a característica h pelo seu respectivo valor.

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

- "Ratio Estimation Groups" - conhecido no Brasil como Estimativas de Razão por Grupos de Controle;
- "Iterative Ratio Estimation Groups" e,
- Raking Ratio Estimators Procedure (RREP), denominado no Brasil, em 1980, como Processo Iterativo por Totais Marginais (PIETOM).

2.2.9.2. "Ratio Estimation Groups" - O processo Estimativo da Razão Por Grupos de Controle objetiva aumentar a eficiência da amostragem através do recurso da pós-estratificação, uso da estimativa e razão, em cada pós-estrato.

A pós-estratificação possibilita introduzir os efeitos da estratificação, enquanto a estimativa da 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 que, através delas se procura explorar os ganhos da estratificação, por um lado, e aquelas que dão origem a 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 onde as suas células constituem os pós-estratos ou grupos de controle; no Censo Demográfico de 1970 essa matriz possuía 46 grupos, ver anexo.

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

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

Nos Censos Demográficos de 1960 e de 1970, do Brasil, 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 $N < 50$ e $w > 16$.

A fim de evitar a produção de resultados inconsistentes, a partir da amostra, são utilizados nas estimativas pesos inteiros imediatamente próximos a 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 restritivos,
- III - Obtenção da matriz final,
- IV - Cálculo do peso final w_{ij} , para cada célula da matriz final (obtida após o reagrupamento).
- V - Determinação e atribuição do peso inteiro a cada unidade da amostra, de cada célula (i,j) .

2.2.9.3 "Iterative Ratio Estimators 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 da razão, os princípios gerais desses recursos técnicos, alguns deles já citados, são explorados neste processo.

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

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 a 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 frequência de unidades da amostragem na amostra, 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 aumentar a confiabilidade do processo de estimação.

Se, para algum grupo, o critério restritivo não é atendido, procede-se ao agrupamento deste grupo 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, a nível de célula (i, j, l) , 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, obtido através do produto da razão entre o total marginal do universo e o respectivo total marginal estimado na primeira vez, por grupo, do estágio I, pelo peso inicial de cada célula:

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 de 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 segue 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 de estágio III.

Após esses três estágios de ajustamentos, a seqüência de operação 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 nas 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 $n \geq 3$ e $w \leq 20$, 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 tem três estágios, são os seguintes, conforme meu entender:

I - Cálculo de $w_{ijl}^{(1)} = \frac{N}{n}$, para cada célula (i,j,l)

Aplicação das regras restritivas;

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

III - Cálculo de $N_{i..}^{(1)} = \sum_j \sum_l N_{ijl}^{(1)}$, 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 $w_{ijl}^{(2)} = w_{ijl}^{(1)} \cdot \frac{N_{i..}}{N_{i..}^{(1)}}$, primeiro ajustamento, ou

peso ajustado pelo estágio I;

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

VII - Cálculo de $N_{.j.}^{(2)} = \sum_i \sum_l N_{ijl}^{(2)}$, para cada coluna;

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

IX - Cálculo $w_{ijl}^{(3)} = w_{ijl}^{(2)} \cdot \frac{N_{.j.}}{N_{.j.}^{(2)}}$, segundo ajustamento, ou pe

so ajustado pelo estágio II;

X - Cálculo de $N_{ijl}^{(3)} = w_{ijl}^{(3)} \cdot n_{ijl}$, em relação ao estágio 3;

XI - Cálculo de $N_{..l}^{(3)} = \sum_i \sum_j N_{ijl}^{(3)}$, para a 3ª dimensão.

XII - Cálculo de $\frac{N_{..l}}{N_{..l}^{(3)}}$;

XIII- Cálculo de $w_{ijl}^{(4)} = w_{ijl}^{(3)} \cdot \frac{N_{..l}}{N_{..l}^{(3)}}$, terceiro ajustamento ou peso ajustado pelo estágio III.

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

onde,

i - número de grupos do estágio I, linhas $i = 1, 2, \dots, 18$

j - número de grupos do estágio II, colunas $j = 1, 2, \dots, 24$

l - número de grupos do estágio III, parte de uma célula (i, j) ;

$l = (1, 2)$

2.2.9.4 - "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 tendências, 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 (3), (15), (16), (17) e (18).

Os grupos das variáveis de estratificação, um para cada dimensão, são definidos visando 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 a seguir ao desenvolvimento do processo.

No Censo de População do Canadá, referente ao ano de 1976, nas estimativas das características das pessoas, foi utilizado uma matriz com $(30 \times 34) = 1\ 020$ células. Os critérios restritivos foram $N = 35$ e $W = 12$, aplicados de modo que se qualquer total da população, no universo, por linha ou coluna foi menor que 35, ou então 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 nessas situações. 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 $(15 \times 48) = 720$ células, sendo adotados os seguintes valores de restrição $n_{i.}$ ou $n_{.j} \geq 5$ e pesos $1 \leq \frac{N_{i.}}{n_{i.}} \leq 8$ e $1 \leq \frac{N_{.j}}{n_{.j}} < 8$. A fração de amostragem desse censo foi $f = 1/4$.

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 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;
- 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 $w_{ij}^{(1)} = \frac{N_{.j}}{n_{.j}}$ → para cada célula (i,j)

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

- III - Calcula-se $w_{ij}^{(2)} = w_{ij}^{(1)} \frac{N_{i.}}{N_{i.}^{(1)}}$ → para cada célula (i,j)
- IV - Calcula-se $N_{.j}^{(2)} = \sum_{i=1}^I w_{ij}^{(2)} n_{ij}$ → para cada coluna j
- V - Calcula-se $w_{ij}^{(3)} = w_{ij}^{(2)} \frac{N_{.j}}{N_{.j}^{(2)}}$ → para cada célula (i,j)
- VI - Calcula-se $N_{i.}^{(3)} = \sum_{j=1}^J w_{ij}^{(3)} n_{ij}$ → para cada linha i
- VII - Calcula-se $w_{ij}^{(4)} = w_{ij}^{(3)} \frac{N_{i.}}{N_{i.}^{(3)}}$ → para cada célula (i,j)

onde ,

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

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

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

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

I → é o número de linhas da matriz;

J → é o número de colunas da matriz.

A fim de evitar a produção de resultados inconsistentes em uma mesma tabulação são utilizados para obtenção das estimativas pesos inteiros. 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_{ij} 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} ."

2.2.9.5 Apreciação 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 Estimation Procedure".

Em geral as estimativas resultam da soma dos pesos atribuídos às unidades da amostra. Como esses pesos são afetados pelas tendenciosidades, por ventura 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, 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 apoia mais na amostra do que no universo, a partir do passo II, a nível 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 são mais consistentes com a respectiva marginal. 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 de fração de amostragem, conseqüentemente proporciona menor variância.

O processo do grupo de controle 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, por força dos critérios restritivos.

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 unidade ou é nenhuma, na amostra, em cada célula, daí proporcionar menor número de agrupamentos.

Os excelentes estudos experimentais, realizados pelo IBGE, por ocasião do censo experimental de 1980 (19), revelaram que os resultados pro-

venientes das estimativas elaboradas com base nos métodos grupos de controle e "RREP" não apresentaram comportamentos divergentes, porém, ligeiramente favorável 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", no cálculo dos pesos, através dos grupos de controle, porém mantidas as mesmas regras restritivas do "RREP" às linhas e colunas do grupo de controle, ao invés da célula, assim as células com até uma unidade na amostra seriam usadas na expansão.

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

A avaliação dos custos de processamento de cada processo de estimação em relação aos ganhos na eficiência das estimativas (menor erro e melhor consistência) seria um caminho natural de se avaliar os benefícios custos em futuras experiências.

2.2.10 - Diretrizes Para a Construção do Processo de Estimação

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

Nos processos abordados fez-se uso da 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;
- h) uso de duas ou mais matrizes de ponderação.

2.2.10.1 - Requisitos do Processo de Estimação

Em conformidade com a teoria de amostragem o estimador deve não ser tendencioso, 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 a sua dimensão, os seus propósitos e o 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 produzir 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 e tornando-o mais consistente, não obstante, o processo de estimativa de razão contenha, em sua concepção, alguma tendenciosidade, a qual é tolerável, além de diminuir a variabilidade. Os resultados do censo em sua maioria das vezes referem-se às variáveis que não constam da matriz de ponderação embora possam estar combinados com alguma variável da matriz. Através dos efeitos de correlação entre as variáveis e o uso do estimador de razão com pós-estratificação as estimativas são beneficiadas de um modo geral além de não se restringir apenas as variáveis da matriz de ponderação.

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.

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.

2.2.10.2 - Redução da Variabilidade

O planejamento da amostra aplicada no censo deve conceber a possibilidade de serem usados processos distintos de estimação de modo a produzirem resultados com menores erros de amostragem ou variabilidades.

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 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.

2.2.10.3 - Controle das Tendenciosidades e das Flutuações de Amostragem

A estimativa de razão é obtida a partir da divisão do total de pessoas (unidade da amostragem) no universo pelo total de pessoas 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 a dimensão (quanto ao número de unidade de amostragem) da área de coleta ou região para qual serão elaboradas as estimativas e por causas aleatórias. Esses fatores influem diretamente nos resultados acarretando superestimções ou subestimções conforme o caso.

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 controlada na coleta, especialmente na situação rural, e em áreas faveladas, e os seus efeitos afetam as estimativas,

previniu-se no planejamento do censo 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 deficiência. Porém os resultados definitivos do censo foram controlados como o previsto antecipadamente.

No anexo 1 apresenta-se uma tabela mostrando 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. Note-se que mais de 95% dos domicílios do país correspondem a uma única família.

Qualquer amostra apresenta flutuação em relação ao universo. Como no censo são elaboradas estimativas para áreas de tamanhos variáveis, e para municípios com pouca população (menos de 5000 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.

2.2.10.4 - 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 torna-se difícil planejar e executar o censo com uma estratificação a priori, geralmente faz-se uso da pós-estratificação, onde os dados dos 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 ten -

denciosidades de coleta e as flutuações de amostragem de forma tal que os benefícios da estratificação sejam diretos, para essas variáveis e os efeitos da estratificação, para outras variáveis, sejam apreciáveis e obtidos indiretamente.

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 vinculados à idade;

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

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

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 devido:

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

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 frequência 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;

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, a nível 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 pela tabela originada, pelos pós-estratos formados pelas variáveis de estratificação, contendo $48 \times 15 = 720$ células.

2.2.10.5 - Consistência

Os resultados do censo são originários pela apuração do universo, para algumas variáveis, e pelas estimativas para as demais variáveis que figuram principalmente no boletim da amostra.

A fim de evitar resultados diferentes entre as apurações do universo e as respectivas estimativas provenientes da amostra, em diversas tabulações, 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 margi -

nais. É oportuno salientar que o processo ideal é aquele que obtém melhor consistência tanto a nível de célula quanto a nível de marginal.

2.2.10.6 - 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 tendência de coleta ocorre independentemente em cada setor censitário, pois depende das falhas cometidas pelo recenseador. Assim sendo, o setor censitário poderia constituir no menor nível geográfico de estimação, com a vantagem de restringir à própria área as suas deficiências de coleta.

Se por esse lado a tendenciosidade fica limitada, por outro, as estimativas tornam-se menos consistentes, face ao número de células da matriz de ponderação que necessariamente exigiria reagrupamentos devido às baixas frequências de unidades de amostragem. As estimativas globais obtidas pelas somas das estimativas elaboradas a nível de 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 possua um número suficiente de unidades de amostragem na amostra. Como os municípios variam substancialmente de tamanho, torna-se difícil de se estabelecer qual seria esse número suficiente, pois os municípios pequenos são áreas de ponderação.

Em 1980 havia 665 municípios com menos de 2000 habitantes. As estimativas elaboradas para esses municípios, naturalmente são menos eficientes do que 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 resultados para níveis geográficos menores do que o município, especialmente para os municípios

maiores e mais importantes, seria interessante, por ocasião do planejamento do censo definir esses níveis geográficos, sem constituírem áreas pequenas, para representarem a área de ponderação, desde que as estimativas não sejam prejudicadas, o que já foi feito em parte em 1980.

2.2.10.7 - 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 esses pontos a estratégia da ponderação prevê critérios restritivos e regras de reagrupamentos 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 nos censos dos EUA, Canadá, Brasil, etc.

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

O tamanho da área de ponderação, o detalhe 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 de 1980 foram estabelecidos a partir dos estudos contidos em (20), onde foram fixados os valores $n_{ij} \geq 5$ e $1 \leq w_{ij} \leq 8$. No censo de 1980 a restrição foi levada para as marginais, conforme o processo RREP, assim sendo as restrições foram: $n_{i.} \geq 5$ e $n_{.j} \geq 5$ e $1 \leq w_{i.} \leq 8$ e $1 \leq w_{.j} \leq 8$.

Quando os valores de n ou de w para uma célula 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 frequências das variáveis que definem os estratos, evitando-se reagrupamento de estratos diferentes por exemplo: pessoas moradoras com não moradoras porque as características dessas pessoas são diferentes, pois são estratos distintos, salvo em última circunstância.

2.2.10.8 - Uso de Duas ou Mais Matrizes de Ponderação

Nos censos demográficos geralmente são elaboradas tabulações referentes às características das pessoas, domicílios e das famílias. Para cada um desses conjuntos, são calculados os 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 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 caracterizem 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 do censo demográfico referentes aos domicílios e às famílias apresentam algumas inconsistências devido ao uso de duas matrizes de ponderação.

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

2.2.11 - Cálculo e apresentação dos erros

Nos censos por amostragem ocorrem duas espécies de erros, o erro alheio à amostragem e os de amostragem. O primeiro é mais difícil de ser avaliado e controlado enquanto o segundo depende dos procedimentos de amostragem adotados.

2.2.11.1. - Erros alheios e amostragem

Alguns erros alheios à amostragem podem ser avaliados através de análises demográficas ou procedendo-se comparações com outras fontes de dados externas que inspirem confiança. Entretanto, os erros de cobertura, geralmente são avaliados por meio de Pesquisas de Avaliação, realizadas por amostragem, de modo que as estimativas das taxas de erro são avaliadas 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 do conhecer e medir o grau e o sentido 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 indefinidos não por falta de recursos técnicos, mas sim por falta de decisão superior. Os resultados disponíveis provêm das apurações preliminares que possibilitam avaliar as taxas de omissão da coleta de domicílios e seus moradores. Os procedimentos e metodologia de cálculo constam de vários documentos.

Outros países também realizam pesquisa de avaliação, pois não há censo sem erro. Não basta promover pesquisas de avaliação, é necessário e indispensável que as tarefas a elas pertinentes sejam acompanhadas e cobra-

das 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.

2.2.11.2 - Erros de amostragem

Os erros de amostragem são estimados com mais precisão utilizando-se os estimadores mais adequados aos processos de estimação de totais, médias e proporções, seguidos 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 de cálculo dos erros podem ser utilizados, dentro deles citam-se (19): 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 (19), os erros de amostragem, expressos através do coeficiente de variação, calculados pelo processo indireto dos grupos aleatórios, geralmente superestimam a variabilidade, devida a utilização da estimativa total, nesses grupos de controle. As estimativas do coeficiente de variação obtidas através da pós-estratificação e da estratificação, aproximadamente são da mesma ordem de grandeza e substancialmente menores do que as estimativas obtidas através do método dos grupos aleatórios, ocasionadas pelo benefício do uso dos pós-estratos. Para as características que não figuram nos pós-estratos, as diferenças 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 grandezas dos coeficientes de variação adotados em 1970 e 1980, para os diversos tamanhos das es

timativas 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. Para estimar os erros da amostragem, foi utilizado inicialmente o método de Keyfitz (10) para cada área de ponderação. Esse método em linhas gerais consiste em calcular a variabilidade entre pares de segmento 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. Posteriormente foram calculados os efeitos do desenho (Deff) para oito áreas distintas em que o país foi dividido, isto devido o esquema de amostragem adotado neste censo ser o de conglomerados.

Todos os censos brasileiros por amostragem já realizados, 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 na pesquisa 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 e útil para a utilização geral dos resultados dos censos, porém isto não é suficiente.

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 tamanhos diferentes e para as variáveis principais, reunindo os resultados em tabelas.

Um outro aspecto a considerar e explicitar é o nível de confiança da estimativa do erro obtido, isto é, se refere a 68% ou a 95%. Os resultados dos erros absolutos do censo de 1970, foram expressos para um nível de confiança de 95%, enquanto as estimativas referentes ao censo de 1980 devem corresponder a nível de 68%. A adoção de um ou outro nível de confiança é válido, 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 níveis mais elevados.

ANEXO I

EM RESENHEAMENTO GERAL - 1990

I- FAMILIAS

1.5- FAMILIAS E PESSOAS RESIDENTES EM DOMICILIOS PARTICULARES, POR NUMERO DE COMPONENTES.

SEGUNDO ALGUMAS DAS PRINCIPAIS CARACTERISTICAS DO CHEFE DA FAMILIA

NUMERO DE ORDEM	CARACTERISTICAS DO CHEFE DA FAMILIA	PESSOAS	FAMILIAS		
			TOTAL	NUMERO DE COMPONENTES	
				1 PESSOA	2 PESSOAS
1	TOTAL.....	117 960 301	26 806 753	1 354 972	4 440 200
SEXO					
2	HOVENS.....	104 742 834	22 616 470	826 248	3 217 387
3	MULHERES.....	13 147 467	4 190 278	728 724	1 222 813
GRUPOS DE IDADE					
4	15 A 19 ANOS.....	738 267	290 362	34 695	150 068
5	20 A 24 ANOS.....	6 377 381	2 112 039	114 355	577 863
6	25 A 29 ANOS.....	12 612 356	3 215 543	125 705	632 194
7	30 A 34 ANOS.....	15 857 763	3 560 510	107 271	359 660
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	106 564	221 036
11	50 A 54 ANOS.....	11 861 477	2 346 504	125 908	289 350
12	55 A 59 ANOS.....	8 330 136	1 864 421	130 443	325 006
13	60 A 64 ANOS.....	5 284 107	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.....	4 454 235	1 526 740	249 708	574 967
16	IDADE IGNORADA.....	216 940	49 764	4 833	8 959
ANOS DE ESTUDO					
17	SEM INSTRUCCAO E MENOS DE 1 ANO.....	39 082 955	8 481 495	697 351	1 424 524
18	1 ANO.....	6 223 456	1 279 447	21 485	102 274
19	2 ANOS.....	10 154 967	2 115 406	58 109	295 039
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	21 474	81 340
23	6 ANOS.....	1 994 487	515 303	22 505	79 431
24	7 ANOS.....	1 918 197	506 189	23 709	107 210
25	8 ANOS.....	5 334 182	1 384 359	66 985	266 066
26	9 ANOS.....	6 08 045	1 69 443	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	256 334
29	12 ANOS.....	343 274	97 885	6 146	22 554
30	13 ANOS.....	426 053	120 401	7 515	27 574
31	14 ANOS.....	583 898	161 953	11 142	35 332
32	15 ANOS.....	1 904 512	497 465	32 359	88 816
33	16 ANOS.....	1 240 882	326 533	18 479	51 264
34	17 ANOS OU MAIS.....	584 383	146 755	10 956	25 328
35	ANOS DE ESTUDO NAO DETERMINADOS.....	96 949	24 278	1 801	4 418
36	SEM DECLARACAO.....	-	-	-	-
SETOR DE ATIVIDADE (1)					
37	ATIVIDADES AGROPECUARIAS, DE EXTRACAO VEGETAL E PESCA.....	36 535 894	7 162 467	266 551	872 307
38	INDUSTRIA DE TRANSFORMACAO.....	15 362 200	3 538 967	106 004	507 559
39	INDUSTRIA DA CONSTRUCCAO.....	9 728 202	2 032 935	64 493	240 359
40	OUTRAS ATIVIDADES INDUSTRIAIS.....	2 149 065	455 410	15 899	52 431
41	COMERCIO DE MERCADORIAS.....	9 054 372	2 036 114	78 742	282 385
42	TRANSPORTES E COMUNICACOES.....	5 650 617	1 259 141	28 840	153 774
43	PRESTACAO DE SERVICIOS.....	10 655 717	2 631 640	181 572	484 403
44	ATIVIDADES SOCIAIS.....	3 544 568	887 102	68 292	168 854
45	ADMINISTRACAO PUBLICA.....	5 088 168	1 075 435	37 583	134 015
46	OUTRAS ATIVIDADES.....	2 257 058	569 890	36 893	101 513
POSICAO NA OCUPACAO (1)					
47	EMPREGADO.....	57 408 560	13 023 830	580 007	1 947 282
48	AUTONOMO.....	37 669 148	7 570 482	267 559	914 654
49	EMPREGADOR.....	4 347 851	911 763	70 037	49 568
50	NAO REQUERIDO.....	121 381	36 080	2 545	17 425
51	SEM DECLARACAO.....	513 001	126 946	14 921	24 733
RENDIMENTO MEDIO MENSAL (SALARIO MINIMO)					
52	ATE 1/2.....	10 306 740	2 655 274	359 221	636 789
53	MAIS DE 1/2 A 1.....	22 059 497	4 927 147	371 694	877 190
54	MAIS DE 1 A 3.....	31 784 758	6 874 321	336 256	1 045 654
55	MAIS DE 3 A 5.....	16 330 859	3 555 174	134 907	501 854
56	MAIS DE 5 A 10.....	15 014 604	3 348 463	115 857	475 797
57	MAIS DE 10 A 20.....	10 243 511	2 400 373	88 297	347 429
58	MAIS DE 20 A 30.....	4 932 612	1 155 916	44 463	161 440
59	MAIS DE 30 A 40.....	2 715 624	603 559	20 985	71 736
60	SEM RENDIMENTO.....	4 056 921	1 042 852	58 982	291 038
61	SEM DECLARACAO.....	465 755	133 579	74 106	31 261

(1) FAMILIAS E PESSOAS CUJOS CHEFES TINHAM 15 ANOS OU MAIS E QUE TRABALHARAM NO ANO DE REFERENCIA.

CENSO DEMOGRAFICO - BRASIL

1- FAMILIAS

1.3- FAMILIAS E PESSOAS, RESIDENTES EM DOMICILIOS PARTICULARES, POR NUMERO DE COMPONENTES, SEGUNDO ALGUMAS DAS PRINCIPAIS CARACTERISTICAS DO CHEFE DA FAMILIA

FAMILIAS							NUMERO DE CASAS	NUMERO DE PESSOAS
NUMERO DE COMPONENTES								
3 PESSOAS	4 PESSOAS	5 PESSOAS	6 PESSOAS	7 A 10 PESSOAS	11 A 14 PESSOAS	15 PESSOAS		
5 028 241	4 839 945	3 772 492	2 543 165	4 124 242	483 495	19 486	1	
4 185 568	4 285 064	3 415 677	2 331 837	3 867 579	488 145	18 945	2	
812 653	554 881	357 295	211 358	256 663	15 350	541	3	
69 591	20 437	8 199	3 552	3 670	186	4	4	
730 200	357 249	134 518	45 843	30 737	1 157	47	5	
1 010 644	874 522	431 426	219 047	174 715	2 908	90	6	
645 210	875 514	670 053	419 702	445 532	14 479	219	7	
371 485	615 048	611 318	455 751	737 342	58 408	1 101	8	
311 468	475 174	512 027	421 120	871 677	175 462	4 101	9	
372 258	388 320	346 943	319 004	704 679	118 873	5 650	10	
351 030	365 225	328 954	257 475	541 075	85 321	4 101	11	
335 714	292 964	237 339	173 939	222 745	43 120	2 221	12	
297 713	270 202	159 352	103 601	164 845	18 255	1 008	13	
265 290	169 831	110 025	65 255	92 157	8 479	409	14	
328 270	175 060	97 072	53 418	62 736	5 154	301	15	
8 345	7 369	6 142	4 642	8 290	1 139	45	16	
1 350 933	1 205 965	1 051 997	813 465	1 686 346	231 914	9 500	17	
208 024	203 349	178 312	140 017	279 098	37 510	1 508	18	
343 721	351 124	304 429	234 541	432 251	43 421	1 871	19	
481 923	497 748	416 419	300 017	574 579	57 647	2 232	20	
1 254 645	1 267 363	935 193	580 566	812 534	74 645	3 274	21	
115 357	103 379	70 228	39 414	39 852	1 728	145	22	
124 444	114 017	72 232	39 414	39 775	2 751	133	23	
127 509	109 356	68 940	37 214	55 875	2 530	96	24	
370 287	306 311	198 682	101 279	103 386	7 037	311	25	
45 078	35 330	21 324	10 195	9 052	534	25	26	
85 757	54 418	34 918	16 910	14 861	401	53	27	
298 552	269 604	191 772	95 761	77 350	3 295	116	28	
24 557	21 792	12 881	5 914	3 875	135	8	29	
29 039	26 281	16 485	7 735	5 156	193	-	30	
36 501	35 357	24 109	11 621	7 811	261	-	31	
98 856	116 123	88 394	43 726	30 385	810	26	32	
54 856	66 415	57 979	33 163	25 562	701	37	33	
25 567	30 186	25 767	15 830	12 746	373	22	34	
4 551	4 842	3 893	2 221	2 379	142	11	35	
-	-	-	-	-	-	-	36	
1 127 792	1 132 040	1 022 655	797 614	1 685 660	246 640	10 183	37	
759 147	788 029	559 232	334 716	440 711	42 098	1 471	38	
368 659	386 435	316 715	224 339	384 691	41 316	1 288	39	
82 117	91 602	74 345	50 187	80 012	5 379	427	40	
395 271	425 667	376 961	258 971	287 202	28 824	1 240	41	
255 750	285 695	214 617	133 857	175 149	15 787	632	42	
329 579	512 219	375 183	231 324	241 724	24 666	973	43	
174 176	188 342	125 975	77 490	54 162	4 236	405	44	
191 944	227 124	185 814	124 297	174 901	18 942	811	45	
117 707	121 132	89 385	48 494	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	1 142 195	849 033	1 619 973	272 462	9 475	48	
142 740	193 737	173 953	113 204	146 735	15 829	922	49	
8 837	4 860	2 751	1 541	2 629	299	13	50	
27 556	20 066	14 909	10 123	17 227	2 113	86	51	
487 738	362 495	272 666	193 401	343 278	40 370	1 316	52	
458 287	737 751	610 302	458 095	896 957	113 024	3 953	53	
1 284 303	1 167 572	936 721	682 955	1 260 053	176 683	6 194	54	
710 141	705 565	531 807	355 605	572 243	70 019	3 011	55	
678 145	731 319	540 811	334 393	459 838	50 087	2 423	56	
470 540	658 873	409 004	234 545	245 717	24 666	1 245	57	
707 957	774 573	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	

TAMANHO DE AMOSTRA ALEATORIA SIMPLES NECESSÁRIO À ESTIMAÇÃO DA MÉDIA
 (Para $\delta = 1,96 \approx 2$)

COEFICIENTE DE VARIÇÃO $\gamma = \frac{\sigma}{\mu} \left\{ \frac{x}{x} \right\}$

TAMANHO DA POPULAÇÃO N	$\gamma = 0,5$				$\gamma = 1,0$				$\gamma = 1,5$			
	Nível de significância %				Nível de significância %				Nível de significância %			
	$\alpha = 0,01$	$\alpha = 0,03$	$\alpha = 0,05$	$\alpha = 0,10$	$\alpha = 0,01$	$\alpha = 0,03$	$\alpha = 0,05$	$\alpha = 0,10$	$\alpha = 0,01$	$\alpha = 0,03$	$\alpha = 0,05$	$\alpha = 0,10$
1 000	909	526	286	91	976	816	615	286	989	909	783	474
2 500	2 000	769	345	96	2 353	1 905	976	345	2 432	2 000	1 475	652
5 000	3 333	909	370	98	4 444	2 353	1 212	370	4 738	3 333	2 094	752
10 000	5 000	1 000	385	99	8 000	3 077	1 379	385	9 000	5 000	2 647	825
25 000	7 143	1 064	394	100	15 385	3 774	1 504	394	19 565	7 143	3 147	869
50 000	8 333	1 087	397	100	22 222	4 032	1 550	397	32 143	8 333	3 358	884
75 000	8 824	1 095	398	100	26 087	4 225	1 567	398	40 909	8 824	3 435	889
100 000	9 090	1 098	398	100	28 571	4 255	1 575	398	47 368	9 090	3 474	892
250 000	9 615	1 106	399	100	34 483	4 369	1 590	399	66 176	9 615	3 549	897
500 000	9 804	1 108	400	100	37 037	4 405	1 595	400	76 271	9 804	3 574	898
1 000 000	9 901	1 110	400	100	38 462	4 425	1 597	400	82 569	9 901	3 587	899
∞	10 000	1 111	400	100	40 000	4 444	1 600	400	90 000	10 000	3 600	900

NOTA: O erro relativo correspondente ao nível de confiança (α) é de $\epsilon_r = \alpha \%$

Utilizou-se a seguinte fórmula: $n = \frac{N \delta^2 \gamma^2}{\delta^2 \gamma^2 + N \epsilon_r^2}$; onde

- n = tamanho da amostra
- N = " da população
- δ = múltiplo do desvio padrão
- δ^2 = variância relativa
- ϵ_r = erro relativo correspondente ao nível de confiança α

R E F E R Ê N C I A S

- (1) Proposições iniciais para a investigação de itens por amostragem no Censo Demográfico de 1980 - versão preliminar - Documento elaborado por *Luiz Nery da Costa* - DECME/SUESP/DT/IBGE - 1979/80.
- (2) Los Metodos de Muestreo y Los Censos - ONU/FAO - 1967.
- (3) Metodologia do Censo Demográfico de 1980 - Fundação IBGE - Rio de Janeiro - 1983.
- (4) U. S. Bureau of the Census - Censuses of Population and Housing - Procedural History - 1960.
- (5) U. S. Bureau of the Census - Censuses of Population and Housing - 1979.
- (6) Estudos sobre a Avaliação da Precisão da Coleta e de Algumas Tendenciosidades do Censo Demográfico de 1970 - Resultados não divulgados - IBGE/SUESP/DECME - 1975.
- (7) Análise preliminar sobre erros da amostragem e alheios à amostragem, ocorridas no Censo Demográfico de 1970 - IBGE/DECEN - Estudo elaborado por *Luiz Nery da Costa* - 1975.
- (8) Census of Canada, Selected Observations - 1981 - Dr. Edward Pryor - Census Manager, Social Statistics Field Statistics Canada.
- (9) Arora, H.R. and Brackstone, G.H. - An Investigation of the Properties of Raking Ratio Estimators with Simple Random Sampling (Census Survey Methods Division), Canada Statistics - 1977.
- (10) Diseño de la Muestra del Censo Nacional de Poblacion y Vivienda de 1980 - Argentina: Procedimientos de Selección, Estimación, Cálculo de Efectos de Diseño y Presentación de Errores de Muestreo-Estadística vol. XXXVII - Avance nº 1 - Santiago - Chile - julho 1985.
- (11) National Report prepared by Canada for the XV Session of Coins
- (12) Estudos sobre a avaliação de precisão da coleta e de algumas tendenciosidades no Censo Demográfico de 1970 - II Resultados definitivos - Brasil
Estudo não divulgado.
- (13) Censo Demográfico Experimental de 1980 - Pesquisa de Avaliação da Precisão da Coleta - IBGE - DT - DECME - agosto/80.

- (14) Censo Escolar de 1977 do Estado do Espírito Santo - Governo do Estado do Espírito Santo - 1979.
- (15) Brackstone, G.J. & Rao, J.N.K. - Raking Ratio Estimators - Canada, Statistics - 1975.
- (16) Arora H.R. & Brackstone, G.J. - An Investigation of the Properties of Raking Ratio Estimators with Simple random sampling - Census Surveys Methods Division Canada - Statistics - 1978.
- (17) Arora, H.R. & Brackstone, G.J. - Weithting Procedures of Raking Ratio Estimation, Census Survey Methods Division - Canada - Statistics - 1978.
- (18) Brackstone, G.J. - The Reliability and Consistency of census data- Census Survey Methods Division - Canada - Statistics - 1976.
- (19) Expansão da Amostra do Censo Experimental de 1980 - Estudos Experimentais- IBGE - DT - DECME - Abril - 1982.
- (20) Costa, Luiz Nery - Uso do Processo de Estimativa de Razão no Censo Demográfico. FIBGE - 1978 - Documento não divulgado.