

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE

ANÁLISE PRELIMINAR SOBRE ERROS DE AMOSTRAGEM E ALHEIOS À  
AMOSTRAGEM, OCORRIDOS NO CENSO DEMOGRÁFICO DE 1970

RIO DE JANEIRO

1974

ANÁLISE PRELIMINAR SOBRE ERROS DE AMOSTRAGEM E ALHEIOS À  
AMOSTRAGEM, OCORRIDOS NO CENSO DEMOGRÁFICO DE 1970

1 - INTRODUÇÃO

Em todo processo de medição, sempre ocorrem erros. Alguns devido a falhas do instrumento de medida, outros motivados pelo processo de operação humano, e ainda outros eventuais ou aleatórios, nem sempre detetados.

Na realização de um Censo esses erros também estão presentes porque estamos frente a um processo de medição e, em se adotando o uso da amostragem, acresce outro tipo de erro que lhe é próprio. Cabe portanto avaliar a intensidade e participação de cada tipo de erro no produto censitário. Eis o escopo do presente documento.

Nas pesquisas estatísticas, os erros podem ser agrupados em erros de amostragem e erros alheios à amostragem.

Os erros de amostragem decorrem das especificações técnicas, dos recursos financeiros e do esquema de amostragem adotado.

As especificações técnicas implicam em definir o grau de precisão e o nível de confiança dos resultados em termos de probabilidade e são estabelecidas no planejamento, propiciando a dimensão dos erros, antes do início da operação. Tais erros variam de acordo com o esquema de amostragem adotado e segundo os recursos disponíveis e utilizados.

Os erros alheios à amostragem independem das técnicas de amostragem utilizadas, mas são oriundos de falhas de planejamento, coleta defeituosa, preparação e processamento de dados deficientes, gerando as conhecidas tendenciosidades.

Não é demais esclarecer que os erros de amostragem são rigorosamente previstos e controláveis, o que não ocorre em relação às tendenciosidades que podem surgir, às vezes, de forma deliberada. Entretanto, na maioria dos casos, as tendenciosidades são introduzidas involuntariamente no decorrer dos trabalhos.

Com o aumento do tamanho da amostra diminui-se o erro de amostragem. Todavia, quanto maior for o tamanho da amostra, maior a possibilidade da introdução de erros alheios à amostragem.

No levantamento do Censo Demográfico de 1970 a amostragem foi aplicada de forma mais ampla.

Planejou-se a obtenção dos resultados finais de dados preliminares e ainda a avaliação da precisão respectiva, empregando-se plenamente a técnica de amostragem. Durante a elaboração dos resultados preliminares e finais, os erros de amostragem e as tendenciosidades estão presentes. A "Pesquisa de Avaliação da Precisão da Coleta do Censo" possibilita detectar e quantificar tanto as falhas de cobertura do Censo quanto as imprecisões na declaração aos quesitos investigados.

Dessa forma, as possíveis falhas do Censo de 1970, podem ser assim detetadas:

a) confrontando-se os resultados definitivos com os preliminares ou ainda com resultados de Censos anteriores.

b) através da Pesquisa de Avaliação no que diz respeito à cobertura e à qualidade do preenchimento dos quesitos.

É fora de dúvida que qualquer medição adquire maior crédito quando ao seu valor é associado o grau de confiança, ou seja, a sua precisão, e ainda são conhecidas e determinadas as falhas eventualmente cometidas. O conhecimento desses elementos, além de possibilitar a melhor utilização dos dados, constitui valioso e imprescindível subsídio para o planejamento de levantamentos futuros.

Com este propósito procura-se apontar e dimensionar algumas das principais falhas já identificadas.

## 2 - CONFRONTO ENTRE OS RESULTADOS DEFINITIVOS DO CENSO E OS RESULTADOS PRELIMINARES (TABULAÇÕES AVANÇADAS)

Para a perfeita compreensão da matéria, apresenta-se sucinto esboço dos procedimentos metodológicos adotados em ambos os trabalhos.

## 2.1 - SÍNTESE METODOLÓGICA DOS PLANOS DE AMOSTRAGEM

Os resultados definitivos do Censo, em sua grande maioria, foram obtidos através de amostra constituída por 25% dos domicílios particulares e pessoas neles recenseadas, e 25% das famílias ou grupos conviventes recenseados em domicílios coletivos.

As tabulações dos itens investigados por amostragem foram obtidas pelo processo de estimativa de razão, no qual os pesos ou fatores de expansão, a nível de município, resultaram da divisão do total de pessoas no universo pelo total de pessoas na amostra. Estas razões obtidas a nível de município, foram determinadas para os 46 grupos de controle, que se encontram definidas no Anexo 1.

Na elaboração dos grupos de controle, procurou-se estabelecer para 5 classes de idade, grupos homogêneos ou estratos quanto:

- à condição de presença;
- ao sexo;
- à situação do domicílio e
- à classificação da pessoa em chefe, cônjuge e outra categoria.

Nas estimativas das características das famílias considerou-se o peso atribuído ao chefe da família e, das características de domicílio, o peso atribuído ao chefe do domicílio.

Para os resultados preliminares do Censo foi utilizada uma subamostra da amostra selecionada na coleta do Censo, correspondendo a um esquema de amostragem desenvolvido em dois estágios, em que foram selecionados cerca de 1/20 das unidades da amostra do censo, reunindo 1,3% de pessoas.

O 1º estágio foi formado pela amostra selecionada na coleta do Censo. No 2º estágio as unidades de amostragem foram constituídas pelos Setores Censitários de Coleta, de modo que nos setores selecionados para a subamostra foram considerados apenas os questionários considerados para a amostra do 1º estágio.

As tabulações resultaram de um processo de estimativa de

razão em que os fatores de expansão ou pesos foram obtidos independentemente para 6 especificações, segundo a situação urbana e rural das pessoas moradoras, por sexo, e das pessoas não moradoras presentes, em cada Microrregião ou grupo de Microrregiões.

Os fatores de expansão foram calculados pela razão entre o total de pessoas recenseadas e o de pessoas selecionadas para a subamostra em cada um dos seis grupos.

Nas estimativas das características das famílias e dos domicílios adotou-se critério idêntico ao das estimativas finais do Censo.

## 2.2 - ERROS DE AMOSTRAGEM CALCULADOS

Foram elaboradas duas tabelas de erros de amostragem esperados. A primeira tabela (Anexo 2) diz respeito aos erros dos resultados definitivos e foi calculada para amostra aleatória simples com fração de amostragem de 1/4 e nível de confiança de 95%.

Na segunda tabela (Anexo 3) figuram os erros de amostragem dos resultados preliminares, calculados com a utilização de "Random Groups" ou subamostras interpenetrantes, aproximadamente equivalentes aos de uma amostra aleatória simples com fração de amostragem de 1/80.

Observa-se que os erros de amostragem esperados nos resultados definitivos do Censo são aproximadamente 5 vezes menores do que a dos resultados preliminares "Tabulações Avançadas", para o mesmo nível de confiança, devido a amostra ter sido 20 vezes maior.

## 2.3 - REALIZAÇÃO DO CONFRONTO

As estimativas dos resultados preliminares podem ser comparadas com os resultados finais mediante o confronto de ambas as estimativas. Em se tratando de resultados de amostra é esperado que as estimativas sejam próximas; entretanto como as tabulações preliminares contêm erros de amostragem mais elevados é natural admitir que a ordem de grandeza dos afastamentos resultantes do confronto não superem aos erros dos resultados preliminares.

Em algumas situações onde seja possível a presença de influências da correlação intraclasse, é provável que as diferenças apresentadas pelo confronto sejam maiores. Entretanto essas diferenças dificilmente ultrapassariam 1,5 o erro de amostragem dos resultados preliminares.

Quando o confronto dos resultados indica diferenças acentuadamente maiores do que as esperadas, podemos atribuir essas diferenças a tendenciosidades introduzidas nos trabalhos.

Teoricamente, os resultados das Tabulações Avançadas e os Resultados Definitivos deveriam ser aproximadamente iguais, apresentando variações da ordem de grandeza igual ao erro de amostragem esperado para cada tamanho da estimativa.

As tendenciosidades podem ser introduzidas em uma ou mais das quatro principais fases de operação:

- a) No Planejamento
- b) Durante a Coleta
- c) Durante a Crítica
- d) Durante a determinação dos Fatores de Expansão e processamento

Assim sendo, procura-se em seguida, caracterizar as principais tendenciosidades.

### 2.3.1 - TENDENCIOSIDADE IDENTIFICADA EM CARACTERÍSTICAS DA FAMÍLIA E DOS DOMICÍLIOS

O confronto dos Resultados Definitivos com os resultados das Tabulações Avançadas para o número de famílias e de domicílios, está demonstrado na tabela 1, a seguir.

Observa-se na referida tabela, que os resultados definitivos apresentam, em todas as Regiões, valores menores do que as Tabulações Avançadas. Essas diferenças são inaceitáveis, mesmo levando-se em conta o erro de amostragem esperado para as Tabulações Avançadas. Tal fato indica haver uma tendenciosidade.

CONFRONTO DOS RESULTADOS DEFINITIVOS COM OS RESULTADOS DAS TABULAÇÕES  
AVANÇADAS PARA O NÚMERO DE FAMÍLIAS E DE DOMICÍLIOS

BRASIL E REGIÕES	NÚMERO DE FAMÍLIAS				NÚMERO DE DOMICÍLIOS			
	Resultados Definitivos (A)	Tabulações Avançadas (B)	Diferença		Resultados Definitivos (C)	Tabulações Avançadas (D)	Diferença	
			Absoluta (B-A)	Rela- tiva (%)			Absoluta (D-C)	Rela- tiva (%)
BRASIL .....	18 554 426	19 104 086	+549 660	2,96	17 628 699	18 086 336	+457 637	2,60
Região I .....	631 739	641 437	+ 9 968	1,54	584 379	592 276	+ 7 897	1,35
Região II .....	901 664	947 190	+ 45 526	5,05	858 738	900 452	+ 41 714	4,86
Região III .....	2 857 861	2 911 641	+ 53 780	1,88	2 727 434	2 772 767	+ 45 333	1,66
REGIÃO IV .....	1 621 234	1 674 707	+ 53 473	3,30	1 554 696	1 602 967	+ 48 271	3,10
Região V .....	2 469 830	2 520 133	+ 50 303	2,04	2 381 841	2 420 555	+ 38 714	1,63
Região VI .....	2 004 712	2 078 511	+ 73 799	3,68	1 883 164	1 948 035	+ 64 871	3,44
Região VII .....	3 854 327	4 000 076	+145 749	3,78	3 636 138	3 755 186	+119 048	3,27
Região VIII .....	1 326 589	1 372 865	+ 46 276	3,49	1 272 355	1 307 798	+ 35 443	2,79
Região IX .....	1 930 203	1 954 149	+ 23 946	1,24	1 813 447	1 828 208	+ 14 761	0,81
Região X .....	956 267	1 003 377	+ 47 110	4,93	916 507	958 092	+ 41 585	4,54

Tal tendenciosidade pode ser proveniente de uma ou mais fases de trabalho:

- dos esquemas de amostragem adotados;
- dos processos de coleta;
- dos processos de crítica dos boletins;
- dos processos de estimação.

Qual seria o fator de tendenciosidade?

A análise de ambos os esquemas de amostragem mostra que os procedimentos adotados no seu planejamento não contêm razões que motivassem as discrepâncias ora encontradas.

A subamostra "Tabulações Avançadas" é parte da amostra do Censo, portanto o material coletado é da mesma qualidade, logo não estaria na coleta o fator de tendenciosidade.

Os procedimentos adotados na crítica dos Boletins de Amostra - CD 1.01 - foram os mesmos em ambas as fases de trabalho tanto na subamostra quanto no preparo dos resultados finais.

É oportuno esclarecer que, durante a crítica desses boletins, constatou-se que parte deles reunia mais de uma família. Essa ocorrência foi sanada mediante o desdobramento do boletim inicial, de modo que as famílias tivessem suas informações em boletins separados.

É fundamental ressaltar que este procedimento relativo ao desdobramento de famílias nos Boletins de Amostra, não foi adotado na crítica dos boletins CD 1.02, dada a carência de informações.

Após a crítica, é lógico que, se na amostra foram desdobrados questionários dando origem a um número X de novas famílias, a não-amostra deveria conter um número 3 vezes X a mais de novas famílias, entretanto isto não ocorreu.

Depreende-se desse fato que, depois de criticados, os Boletins de Amostra CD 1.01 isoladamente traduzem com maior fidelidade as características de família e de domicílios do que o universo dos Boletins CD 1.01 e CD 1.02. Convém esclarecer que esta crítica não alterou a população.

Ora, como a expansão da amostra da Tabulação Avançada es-

tá baseada apenas nas informações dos boletins de amostra CD 1.01, enquanto que os resultados finais são oriundos da reunião das famílias de ambos os boletins, o desdobramento de famílias nos boletins CD 1.01 introduziu a tendenciosidade do maior número de famílias em relação ao universo.

De fato, segundo a crítica, existe esse maior número de famílias, que não aparece nos resultados finais, devido estarem reunidas com outras famílias nos boletins CD 1.02, os quais não receberam crítica.

Convém lembrar que os fatores de expansão da subamostra levam em conta a população recenseada e a população na subamostra para cada uma das 6 categorias de condição de presença e sexo. O aumento do número de famílias não influi no peso do chefe da família, pois o número de pessoas recenseadas na amostra e na subamostra é o mesmo e próximo de 80, que é o inverso da fração geral de subamostragem. É natural que em cada nova família desdobrada o novo chefe criado passe a receber também peso em torno de 80. Portanto todas as características da proporção Y de famílias e de domicílios desdobradas na subamostra são multiplicadas por 80.

Por outro lado, na expansão da amostra do Censo, os pesos são determinados pela razão entre o número de pessoas no universo e na amostra, levando em conta também se essas pessoas são chefes de família, cônjuges e outros. É claro que os pesos dos chefes sofreram redução devido ao fato de haver relativamente maior número de chefes de família em boletins de Amostra que no Universo. Isto é consequência dos desdobramentos de boletins de amostra e não do processo de ponderação.

Exemplificação numérica:

Suponha-se a seguinte situação antes da crítica dos boletins:

a) no Censo

	Chefes de família	Pessoas
Universo.....	8 000	40 000
Boletins da Não-Amostra (CD 1.02)	6 000	30 000
Boletins da Amostra (CD 1.01) ...	2 000	10 000

91

Nota-se que a fração de amostragem é  $f = \frac{1}{4}$  e o peso de uma pessoa ou do chefe seria  $p = \frac{8\ 000}{2\ 000} = 4$

b) na Subamostra

	Chefes de família	Pessoas
Boletins da Amostra .....	100	500

Observa-se que, no primeiro estágio, a fração de amostragem é 1/20, e que em geral é de 1/80. Portanto o peso de uma pessoa ou do chefe seria:

$$p = \frac{\text{Pessoas no Censo}}{\text{Pessoas na Subamostra}} = \frac{40\ 000}{500} = 80$$

Suponha-se agora que após a crítica houve a inclusão de 5% de novas famílias, resultante do desdobramento nos boletins de amostra.

Teríamos:

a) no Censo

	Chefes de família	Pessoas
Universo .....	8 100	40 000
Boletins da Não-Amostra .....	6 000	30 000
Boletins da Amostra .....	2 100	10 000

Verifica-se que apenas alterou o número de chefes em 5% nos Boletins da Amostra e pouco mais de 1% no Universo.

O peso do chefe passou a ser:

$$p = \frac{2\ 100 + 6\ 000}{2\ 100} = \frac{8\ 100}{2\ 100} = 3,86, \text{ que é menor do que } 4.$$

A expansão do número de famílias, a partir da amostra, seria:  $2\ 100 \times 3,86 = 8\ 100$  famílias.

Como na amostra houve um acréscimo de 100 famílias, na não

-amostra deveria também haver um acréscimo correspondente, ou seja 3 vezes 100 = 300 famílias. Neste caso o peso não se alteraria pois  $\frac{2\ 100 + 6\ 300}{2\ 100} = \frac{8\ 400}{2\ 100} = 4$ .

b) na Subamostra

	Chefes de família	Pessoas
Boletins da Amostra .....	105	500

Nota-se que apenas foi alterado o número de chefes e quanto ao de pessoas permaneceu o mesmo.

O peso das pessoas continua sendo 80. Como o peso dos chefes é igual ao de pessoas, conseqüentemente as estimativas de características de famílias e de domicílios serão aumentadas na medida dos desdobramentos verificados.

A estimativa do número de chefes de família seria, então, 105 multiplicado por 80 = 8 400.

O confronto das estimativas do número de famílias provenientes da amostra com os da subamostra indica:

Expansão da amostra	8 100
Expansão da subamostra	<u>8 400</u>
Diferença	300

Essa diferença de 300 famílias para menos nos resultados finais do Censo é conseqüência do não desdobramento dos Boletins CD 1.02. Na realidade essas 300 famílias existem porque, sendo a subamostra um subconjunto representativo da amostra e esta um subconjunto maior, representativo do universo, o que ocorreu na subamostra, ocorreu também na amostra e deveria ocorrer no universo com a mesma intensidade.

#### Comprovação do raciocínio

Durante a operação de processamento dos resultados do Censo, várias tabelas de controle dos trabalhos de ponderação da amostra foram solicitadas ao Centro de Informática.

Uma dessas tabelas indica para cada "Grupo de Controle de Estimativa" o número de pessoas da amostra segundo o peso que teria recebido.

A referida tabela (Anexo 4), não só permite várias análises, como também possibilita comparar a distribuição percentual do número de pessoas na amostra que teriam recebido pesos 1, 2, 3, ....., 16 e mais, para os grupamentos de chefes de família, não chefes e total.

As distribuições percentuais desses três conjuntos de grupamentos: chefes, não chefes e total (chefes + não chefes) devem ser aproximadamente iguais se não houver tendenciosidade.

Havendo diferenças significativas nessas distribuições fica evidenciada a presença de tendenciosidade e em qual grupamento ela ocorreu. Quanto maior a diferença entre o total e as parcelas, maior será a tendenciosidade ocorrida.

Na tabela 2, a seguir, apresentam-se resumos provenientes da tabela IV do anexo, para algumas Regiões das Tabulações Avançadas.

Depreende-se desta tabela o seguinte:

a) o peso 4 é o mais freqüente e a proporção das pessoas que receberam este peso, chefes, não chefes e total, dentro de cada região, apresentam menores variações relativas;

b) em relação ao peso 3, verifica-se que as pessoas chefes representam proporção bem maior do que as não chefes;

c) conseqüentemente ao peso 5 as pessoas chefes representam uma proporção bem menor que as não chefes;

d) dentre as quatro regiões consideradas, a região IX é a que apresenta menores variações em cada pessoa.

De acordo com a tabela 1 a região IX está menos influenciada pela aludida tendenciosidade. Pela tabela 2 é exatamente essa região que apresenta menores variações relativas aos grupamentos citados em cada peso. Este fato comprova que a tendenciosidade provém dos desdobramentos dos boletins.

AS DISTRIBUIÇÕES CITADAS REFERENTES ÀS REGIÕES II, VII, VIII E IX DAS  
"TABULAÇÕES AVANÇADAS" SÃO APRESENTADAS NA TABELA ABAIXO

UNIDADES DA FEDERAÇÃO, CHEFES E NÃO CHEFES	PESSOAS NA AMOSTRA					
	Total	Pesos recebidos				
		1 e 2	3	4	5	6 e mais
VALORES ABSOLUTOS						
<b>II - MARANHÃO E PIAUÍ</b>						
Total .....	1 172 093	3 931	178 115	758 641	207 142	24 264
Chefes .....	231 063	916	49 297	155 327	22 740	2 783
Não Chefes .....	941 030	3 015	128 818	603 314	184 402	21 481
<b>VII - SÃO PAULO</b>						
Total .....	4 646 525	36 171	947 950	3 358 485	257 137	46 782
Chefes .....	1 032 927	10 053	289 830	699 468	23 941	9 635
Não Chefes .....	3 613 598	26 118	658 120	2 659 017	233 196	37 147
<b>VIII - PARANÁ</b>						
Total .....	1 849 451	24 012	553 577	1 121 730	127 075	23 057
Chefes .....	368 002	8 611	151 223	192 703	11 577	3 888
Não Chefes .....	1 481 449	15 401	402 354	929 027	115 498	19 169
<b>IX - RIO GRANDE DO SUL E SANTA CATARINA</b>						
Total .....	2 550 264	17 190	649 771	1 737 275	132 177	13 851
Chefes .....	520 394	3 763	154 704	344 956	14 924	2 047
Não Chefes .....	2 029 870	13 427	495 067	1 392 319	117 253	11 804

## VALORES RELATIVOS (%)

<b>II - MARANHÃO E PIAUÍ</b>						
Total .....	100,0	0,3	15,2	64,7	17,7	2,1
Chefes .....	100,0	0,4	21,3	67,3	9,8	1,2
Não Chefes .....	100,0	0,3	13,7	64,1	19,6	2,3
<b>VII - SÃO PAULO</b>						
Total .....	100,0	0,8	20,4	72,3	5,5	1,0
Chefes .....	100,0	1,0	28,1	67,7	2,3	0,9
Não Chefes .....	100,0	0,7	18,2	73,6	6,5	1,0
<b>VIII - PARANÁ</b>						
Total .....	100,0	1,3	29,9	60,7	6,9	1,2
Chefes .....	100,0	2,3	41,1	52,4	3,1	1,1
Não Chefes .....	100,0	1,0	27,2	62,7	7,8	1,3
<b>IX - RIO GRANDE DO SUL E SANTA CATARINA</b>						
Total .....	100,0	0,7	25,5	68,1	5,2	0,5
Chefes .....	100,0	0,7	29,7	66,3	2,9	0,4
Não Chefes .....	100,0	0,7	24,4	68,5	5,8	0,6

Reflexos dessa tendenciosidade nos resultados definitivos do Censo:

a) todas as características de família são afetadas no total, na mesma medida avaliada acima; em certos detalhes esta influência é variável;

b) todas as características de domicílios são afetadas em intensidade, ligeiramente menor do que a avaliada para as famílias;

c) os resultados de idades, baseados na amostra, foram afetados em menor intensidade;

d) todas as características nas quais o chefe de família tenha participação acentuada, foram atingidas, por exemplo, estado conjugal, setor de atividade, etc. A medida desta influência nestas características é menor do que a ocorrida em relação aos chefes.

3.2.2 - TENDENCIOSIDADE EXISTENTE NA INDAGAÇÃO DA IDADE

A investigação do quesito idade foi feita de duas maneiras: pela declaração da data de nascimento e pela idade presumida. A comparação das idades nestas duas formas de declaração mostra que em certas idades, especialmente nas idades mais avançadas, a idade presumida afasta-se mais da idade declarada e ainda que nestas idades há uma forte atração para as idades múltiplas de 5.

Durante a tabulação dos resultados definitivos do Censo, referentes ao Estado do Piauí, foi obtido um conjunto de três tabelas sobre a idade, a forma de declaração da idade e a origem da informação - se Boletim da Amostra ou Boletim da Não-Amostra.

O resumo destas tabelas, onde são apresentadas as participações de cada idade no total, segundo a forma de declaração e origem, figura na tabela a seguir.

TABELA - 3  
 CONFRONTO ENTRE AS DISTRIBUIÇÕES DE IDADES, SEGUNDO  
 A FORMA DE DECLARAÇÃO DE IDADE E A ORIGEM DOS  
 DADOS REFERENTES AO ESTADO DO PIAUÍ

IDADES QUINQUENAIS (Anos)	TOTAL			DATA DE NASCIMENTO			IDADE PRESUMIDA		
	Total	Amos- tra	Não Amos- tra	Total	Amos- tra	Não Amos- tra	Total	Amos- tra	Não Amos- tra
5 .....	3,5	3,2	3,5	3,7	3,4	3,9	3,0	2,8	3,1
10 .....	3,2	3,1	3,2	3,2	3,1	3,3	3,1	3,1	3,1
15 .....	2,5	2,7	2,4	2,5	2,7	2,4	2,4	2,6	2,3
20 .....	2,2	2,2	2,2	2,0	2,1	1,9	2,6	2,6	2,6
25 .....	1,6	1,5	1,7	1,3	1,4	1,3	2,1	1,9	2,2
30 .....	1,7	1,6	1,7	1,3	1,3	1,2	2,3	2,2	2,4
35 .....	1,1	1,0	1,1	0,9	0,9	0,8	1,6	1,4	1,6
40 .....	1,3	1,2	1,3	0,8	0,9	0,8	2,1	1,9	2,1
45 .....	0,9	0,8	0,9	0,7	0,7	0,6	1,2	1,2	1,3
50 .....	0,9	0,8	0,9	0,5	0,6	0,5	1,5	1,4	1,6
55 .....	0,7	0,6	0,7	0,5	0,5	0,5	0,9	0,9	1,0
60 .....	0,7	0,6	0,8	0,3	0,4	0,3	1,4	1,3	1,4
65 .....	0,3	0,3	0,4	0,2	0,2	0,2	0,6	0,6	0,6
70 .....	0,4	0,3	0,4	0,1	0,1	0,1	0,8	0,8	0,8
75 .....	0,1	0,1	0,1	0,0	0,0	0,0	0,3	0,3	0,2
80 .....	0,1	0,1	0,1	0,0	0,0	0,0	0,2	0,3	0,3

Da análise desta tabela, conclui-se:

a) para o total, conjunto das duas formas de declaração de idade, Data de Nascimento e Idade Presumida, praticamente não há diferença significativa quer seja na informação prestada no Boletim da Amostra quer no Boletim da Não-Amostra, isto em todas as idades;

b) para as formas de declaração de idade - Data de Nascimento - e - Idade Presumida - em todas as idades, o total indica a mesma ordem de grandeza encontrada para a origem da informação se proveniente do Boletim da Amostra ou do Boletim da Não-Amostra;

c) considerando indiferente a origem da informação, se proveniente do Boletim da Amostra ou do Boletim da Não-Amostra ou do seu conjunto, observa-se que os confrontos das participações segundo as formas de declaração de idade apresentam variações mais significativas, aumentando as diferenças à medida que as idades são mais elevadas;

d) isoladamente, para a declaração de idade - Data de Nascimento - a participação percentual apresenta uma distribuição contínua, mais elevada nas idades menores e decrescendo à medida que as idades aumentam, independentemente da origem da informação se proveniente dos Boletins da Amostra, ou dos Boletins da Não-Amostra ou seu conjunto;

e) o mesmo não ocorre em relação à forma de declaração de idade - Idade Presumida - a distribuição percentual apresenta saltos mais fortes, especialmente nas idades de 50, 60 e 70 anos.

Portanto, infere-se que a investigação segundo a forma de declaração com a Data de Nascimento apresenta maior exatidão do que a forma de declaração Idade Presumida.

Procurando-se melhor ilustrar este fato, apresenta-se em seguida uma tabela proveniente do conjunto das aludidas tabelas, onde se procura medir a participação da forma de declaração de idade com Data de Nascimento nos Boletins da Amostra e nos Boletins da Não-Amostra, em relação ao total de todas as formas.

TABELA - 4  
PARTICIPAÇÃO DA FORMA DE DECLARAÇÃO DE IDADE, SEGUNDO  
A DATA DE NASCIMENTO NOS BOLETINS DA AMOSTRA E DA  
NÃO AMOSTRA EM RELAÇÃO AO TOTAL  
PIAUI

IDADES QUINQUENAIS (Anos)	NÚMERO DE PESSOAS						
	Total			Com declaração da data de nascimento			
	Univer so (b)	Amos tra (c)	Nã o Amos tra (d)	Amostra		Nã o Amostra	
				Total (e)	% f=(e)/(c)	Total (g)	% h=(g)/(d)
(a)	(b)	(c)	(d)	(e)	f=(e)/(c)	(g)	h=(g)/(d)
TOTAL ...	1 680 573	452,460	1 228 113	327 501	72,4	733 599	59,7
5 .....	58 081	14 490	43 591	11 020	76,1	28 512	65,4
10 .....	53 617	14 099	39 518	10 251	72,7	24 147	61,1
15 .....	41 338	12 089	29 249	8 905	73,7	17 905	61,2
20 .....	36 689	10 171	26 518	6 901	67,8	13 852	52,2
25 .....	27 455	7 009	20 446	4 611	65,6	9 647	47,2
30 .....	27 776	7 029	20 747	4 327	61,6	9 109	43,9
35 .....	18 818	4 744	14 074	3 015	63,6	6 098	43,3
40 .....	21 504	5 300	16 204	2 891	54,5	5 884	36,3
45 .....	14 871	3 825	11 046	2 290	59,9	4 613	41,8
50 .....	15 175	3 566	11 609	1 808	50,7	3 880	33,4
55 .....	11 034	2 726	8 308	1 666	61,1	3 609	43,4
60 .....	12 133	2 778	9 355	1 176	42,3	2 507	26,8
65 .....	5 819	1 338	4 481	652	48,7	1 364	30,4
70 .....	6 108	1 423	4 685	477	33,5	872	18,6
75 .....	2 289	541	1 748	212	39,2	374	21,4
80 .....	2 055	483	1 572	137	28,4	263	16,7

De fato, observa-se nesta tabela que 72% das pessoas incluídas nos boletins de amostra declararam a data de nascimento, enquanto que nos Boletins da Não-Amostra esta percentagem caiu para 59,7%.

Este comportamento se verifica para todas as idades, e nas mais elevadas a declaração da data de nascimento representa uma proporção menor. Entretanto, à medida que as idades aumentam, maiores são as proporções relativas das pessoas que declararam a data de nascimento nos Boletins da Amostra, em relação às da Não-Amostra.

Houve portanto, uma tendenciosidade no quesito de idade durante a coleta, quando tudo leva a crer tenham sido os Boletins da Amostra preenchidos com maior cuidado do que os da Não-Amostra.

Admitindo-se que este fato observado para o Piauí também tenha ocorrido nas demais Unidades da Federação, pode-se afirmar que a obtenção das idades através dos Boletins da Amostra propicia resultados mais precisos do que levando-se em conta o Universo.

O confronto dos resultados definitivos do Censo com os da Tabulação Avançada, quanto à idade, apresenta afastamentos que fogem ao esperado, especialmente para o Brasil.

Considerando o fato de que os resultados das Tabulações Avançadas são oriundos apenas de Boletins de Amostra, e que os resultados definitivos do Censo levam em conta as informações do Universo, face ao exposto anteriormente, os resultados da subamostra seriam mais precisos que os resultados do Censo.

Durante o processamento dos trabalhos definitivos do Piauí, haviam sido notadas diferenças entre os resultados definitivos do Universo comparados com os da expansão da amostra. Na ocasião, analisando-se o problema, chegamos à hipótese de que a causa seria proveniente do fato de, ao serem selecionadas as pessoas para receberem os pesos menores e maiores que 4, elas o foram de forma sistemática e não aleatória.

Como o intervalo de seleção era próximo de 4, é provável que algum efeito de conglomeração tenha introduzido um efeito de causa sistemática no estabelecimento dos pesos, devido ao processo de seleção.

Finalmente, é admissível que as diferenças resultantes do confronto dos Resultados Definitivos do Censo com os das Tabulações Avançadas sejam superiores aos erros de amostragem esperados, principalmente em relação aos dados do conjunto do País. De fato considerando-se que:

- a) muitos chefes de família receberam pesos menores do que deveriam ter recebido;
- b) a melhor qualidade da declaração de idade nos Boletins da Amostra é neutralizada pelas informações provenientes dos Boletins da Não-Amostra;
- c) possivelmente, o processo de aplicação de peso referido acima, tenha exercido influência nos resultados, tal conclusão é perfeitamente válida.

### 3.2.2.1 - INCOMPATIBILIDADES EXISTENTES NAS IDADES BASEADAS NA COMPARAÇÃO DAS COMPOSIÇÕES ETÁRIAS DOS CENSOS DE 1940 A 1970

Viu-se anteriormente, que há algumas falhas nos resultados da investigação das idades no Censo de 1970, indicando-se as possíveis causas das tendenciosidades. Entretanto, isto já ocorreu em investigações censitárias anteriores.

A comprovação, de modo aproximado desta assertiva, é feita comparando-se os resultados da composição etária dos Censos de 1940 a 1970. Nessa comparação examina-se a população de certa faixa de idade numa determinada época com a população existente 10 anos depois.

As composições etárias dos referidos censos são apresentadas na tabela que segue.

Nesta tabela são observadas algumas incompatibilidades, vejamos:

- a) a população existente em 1960 - na faixa etária de 10 a 14 anos foi 8 574 mil habitantes, entretanto ela é superior à população existente em 1950 na faixa etária de 0 a 4 anos, ou seja 8 371 mil habitantes;
- b) semelhantemente, a população existente em 1970 na faixa etária de 10 a 14 anos foi 11 766 mil habitantes, por outro lado, esta população é oriunda dos 11 196 mil habitantes com 0 a 4 anos em 1960.

TABELA - 5

COMPOSIÇÃO ETÁRIA DA POPULAÇÃO PRESENTE, SEGUNDO  
OS RECENSEAMENTOS GERAIS

GRUPOS DE IDADE	POPULAÇÃO (1 000 hab)			
	1º-IX-1940	1º-VII-1950	1º-IX-1960	1º-IX-1970
TOTAL .....	41 236	51 944	70 119	92 342
0 a 4 anos .....	6 440	8 371	11 196	13 827
5 a 9 anos .....	5 759	7 016	10 161	13 416
10 a 14 anos .....	5 328	6 309	8 574	11 766
15 a 19 anos .....	4 444	5 502	7 142	10 122
20 a 24 anos .....	3 813	4 991	6 161	8 147
25 a 29 anos .....	3 356	4 132	5 209	6 424
30 a 39 anos .....	4 902	6 286	8 505	10 620
40 a 49 anos .....	3 442	4 365	5 960	7 973
50 a 59 anos .....	2 045	2 650	3 787	5 172
60 a 69 anos .....	1 076	1 451	2 185	2 995
70 anos e mais .....	599	754	1 128	1 714
Idade ignorada .....	32	117	111	165

A fim de melhor comparar os resultados acima, elaborou-se uma tabela que mostra as razões do COORTE num período decenal.

Estas razões de Coorte apresentam as alterações provenientes de todas as causas, representando de certa maneira os sobreviventes do período considerado. Elas devem ser inferiores a 1,00, sendo geralmente altas nas idades menores, e decrescendo nas idades mais avançadas. Valores superiores a 1,00 ou afastando-se de forma acentuada da tendência da série, indicam de modo aproximado as incompatibilidades.

COMPARAÇÃO DAS RAZÕES DE COORTE NOS CENSOS 1940 a 1970

IDADES (L <sub>x</sub> )	RAZÕES DECENAIS DE COORTE (1)		
	$\frac{L_x + 10}{L_x}$		
	$\frac{\text{Censo de 1950}}{\text{Censo de 1940}}$	$\frac{\text{Censo de 1960}}{\text{Censo de 1950}}$	$\frac{\text{Censo de 1970}}{\text{Censo de 1960}}$
De 0 a 4 anos .....	0,98	1,02	1,05
De 5 a 9 anos .....	0,96	1,02	1,00
De 10 a 14 anos .....	0,94	0,98	0,95
De 15 a 19 anos .....	0,93	0,95	0,90
De 20 a 29 anos .....	0,88	0,93	0,93
De 30 a 39 anos .....	0,89	0,95	0,94
De 40 a 49 anos .....	0,77	0,87	0,87
De 50 a 59 anos .....	0,71	0,82	0,79
De 60 a 69 anos .....	0,70	0,78	0,78

(1) Para dois Censos decenais realizados nos anos t e t + 10 anos, a razão decenal de Coorte ( $L_x + 10/L_x$ ), é calculada dividindo-se a população sobrevivente  $L_x + 10$  com a idade (x + 10) anos, na época (t + 10) anos pela população original ( $L_x$ ) que tinha a idade x anos na época t.

As razões de Coorte do período 1940/1950, confirmam o exposto no parágrafo anterior.

No período 1950/1960 seria esperado que os valores das razões fossem de um modo geral superiores aos do período anterior, devido às reduções nas taxas de mortalidade. Entretanto observam-se dois valores superiores a 1,0 indicando a existência de incompatibilidade nas populações de 1960 nos grupos de 10 a 14 anos e de 15 a 19 anos, que possivelmente estariam superestimadas.

No período 1960/1970 seriam esperadas razões iguais ou ligeiramente superiores às do período anterior. Todavia observa-se que as quatro primeiras razões apresentam incompatibilidades, pois são superiores a 1,0 ou apresentam valores destoantes da série. Aparentemente isto indica que as populações de 0 a 4 anos e de 5 a 9 anos em 1960 estariam subestimadas e por outro lado, as populações de 10 a 14 anos e de 15 a 19 anos estariam superestimadas, aliás como já foi visto.

Portanto é provável que a composição etária da população de 1960 esteja discrepante, especialmente nas quatro primeiras classes de idade. Nota-se que justamente em 1960 o critério de investigação de idade limitou-se à indagação da idade presumida.

### 3.2.3 - TENDENCIOSIDADE OCORRIDA NA INDAGAÇÃO SOBRE A ALFABETIZAÇÃO

O confronto dos Resultados Definitivos do Censo com os da Subamostra (Resultados Preliminares) mostra que o número de pessoas alfabetizadas na Subamostra foi sempre superior aos Resultados Definitivos.

Geralmente quando uma situação semelhante a esta ocorre, isto significa que é provável a existência de tendenciosidade.

A tabela que segue reúne ambas as estimativas e possibilita melhor visualização dessa ocorrência, além disso, realça a ordem de grandeza absoluta e relativa dessa tendenciosidade e em que Regiões as diferenças entre os resultados são maiores.

TABELA - 6

PESSOAS ALFABETIZADAS SEGUNDO OS RESULTADOS DEFINITIVOS  
E PRELIMINARES DO CENSO, POR REGIÕES

BRASIL E REGIÕES	PESSOAS ALFABETIZADAS			
	Resultados definitivos (b)	Resultados preliminares (c)	Diferença	
			Absoluta (d)=(b)-(c)	Relativa %
BRASIL . . . . .	47 864 531	48 821 471	- 956 940	2,0
Região I . . . . .	1 610 711	1 657 752	- 47 041	2,9
Região II . . . . .	1 302 337	1 331 480	- 29 143	2,2
Região III . . . . .	4 947 764	5 063 930	- 116 166	2,4
Região IV . . . . .	2 924 551	3 027 896	- 103 345	3,5
Região V . . . . .	6 578 539	6 759 936	- 181 397	2,8
Região VI . . . . .	6 148 825	6 303 144	- 154 319	2,5
Região VII . . . . .	12 093 640	12 166 806	- 73 166	0,6
Região VIII . . . . .	3 602 663	3 689 688	- 87 025	2,4
Região IX . . . . .	6 241 770	6 350 670	- 108 900	1,7
Região X . . . . .	2 413 731	2 470 169	- 56 438	2,3

Observa-se que, em média, houve uma subestimação da ordem de 2% nos resultados de cada Região. Essa diferença de 2% é bem mais elevada do que os erros de amostragem esperados para as estimativas citadas. Deve ter ocorrido uma tendenciosidade nessa investigação, pois há sempre subestimação nos resultados de todas as Regiões.

O Boletim da Amostra contém o quesito de alfabetização no meio do boletim e ainda dispõe de quesitos sobre o nível de instrução, de modo que permite uma crítica mais precisa da indagação sobre a alfabetização.

No Boletim da Não-Amostra o quesito sobre alfabetização figura no fim do questionário, sem possuir outros quesitos que possibilitem crítica.

O Boletim da Não-Amostra não recebeu crítica de espécie alguma, enquanto que o Boletim da Amostra, após a crítica, teve a sua consistência melhorada. O confronto entre os resultados definitivos do Censo e os da subamostra, para a soma do número de pessoas sem declaração ao quesito alfabetização e das que declararam serem analfabetas, apresenta uma diferença da mesma ordem de grandeza da medida da subestimação ao quesito alfabetização que as anula. A tabela seguinte comprova este fato para o Brasil.

TABELA - 7  
ALFABETIZAÇÃO SEGUNDO OS RESULTADOS DEFINITIVOS  
E PRELIMINARES  
BRASIL

ESPECIFICAÇÃO	PESSOAS		
	Resultados definitivos (a)	Resultados preliminares (b)	Diferença absoluta (c)=(a)-(b)
Alfabetizados .....	47 864 531	48 821 471	- 956 940
Analfabetos .....	30 718 597	30 424 191	+ 294 406
Sem declaração .....	744 103	60 095	+ 684 008

Depreende-se que houve de fato uma tendenciosidade de omissão no preenchimento dos Boletins da Não Amostra, no quesito alfabetização. A diferença observada de analfabetos, provenientes dos resultados aludidos é devido à correção que foi feita nos Boletins de Amostra.

É oportuno ressaltar que a crítica procedida nos Boletins da Amostra beneficiou a qualidade das informações com a conseqüente melhoria dos resultados provenientes da utilização somente dos Boletins da Amostra. Realmente, as estimativas do Censo utilizando apenas as informações da amostra, ao invés do universo, contém a mesma consistência e são da mesma ordem de grandeza das estimativas das Tabulações Avançadas.

#### 3.2.4 - TENDENCIOSIDADE NA ELABORAÇÃO DOS PROGRAMAS E NO PROCESSAMENTO DA TABULAÇÃO AVANÇADA

No decorrer das tarefas atinentes ao processamento das Tabulações Avançadas, foram por várias vezes refeitos alguns programas e tabulações, devido a incompatibilidades que os resultados apresentavam.

Além disso, houve interpretações errôneas quanto à utilização de alguns códigos, tanto na elaboração de programas quanto na complementação da crítica eletrônica das fitas da subamostra, especialmente no que diz respeito aos itens estudantes, anos de estudo e fecundidade.

Estes fatos ocasionaram tendenciosidades diversas e, em conseqüência, algumas estimativas das Tabulações Avançadas foram influenciadas, acarretando subestimações e superestimações em relação aos resultados definitivos do Censo.

#### 3.2.5 - TENDENCIOSIDADES OCORRIDAS NA INVESTIGAÇÃO DO NÚMERO DE FILHOS NASCIDOS VIVOS NO ANO ANTERIOR À DATA DO CENSO, NA POPULAÇÃO COM MENOS DE UM ANO DE IDADE E COM UM ANO.

A população brasileira com menos de 5 anos de idade, segundo os resultados definitivos do Censo de 1970, é a seguinte:

TABELA - 8  
POPULAÇÃO COM MENOS DE 4 ANOS DE IDADE  
BRASIL

IDADES	POPULAÇÃO
Menos de 1 ano .....	2 798 798
1 ano .....	2 499 387
2 anos .....	2 833 056
3 anos .....	2 841 621
4 anos .....	2 859 944

O número de filhos tidos nascidos vivos no ano anterior ao Censo, segundo a amostra e a subamostra, foram 2 897 900 e 2 943 603 respectivamente. Tais valores representam estimativas compatíveis, considerando o erro de amostragem esperado.

Durante o decênio de 1960/1970 a taxa bruta de natalidade para o Brasil (1) foi estimada em 0,0377.

As populações residentes estimadas para 1-7-1968 e 1-7-1969 no Brasil (2) seriam respectivamente 88 222 000 e 90 768 000.

Obtêm-se estimativas aproximadas do número de filhos nascidos vivos durante os períodos 2-9-1968 a 1-9-1969, e de 2-9-1969 a 1-9-1970, aplicando-se a taxa de natalidade 0,0377 às populações estimadas, indicadas acima, quais sejam 3 326 000 e 3 422 000, respectivamente.

Convém esclarecer que estes 3 326 mil nascidos em 1968/69 teriam, por ocasião do Censo de 1970, de 1 a 2 anos de idade. Da mesma forma os 3 422 mil nascidos em 1969/70 teriam menos de 1 ano por ocasião do Censo.

(1) Boletim Demográfico - CBED (julho/setembro - 1971).

(2) " " " (abril/junho - 1972).

O confronto da estimativa de 3 422 000 filhos nascidos vivos no período 1/9/69 a 1/9/70, com o respectivo resultado definitivo do Censo - 2 898 mil - indica ter havido omissão na declaração censitária de 524 mil pessoas (3).

Considerando o nº de filhos nascidos vivos no período 1/9/69 a 1/9/70 e a população com menos de 1 ano, segundo os resultados definitivos do Censo, ou sejam 2 898 mil e 2 799 mil respectivamente, infere-se que teriam morrido 99 mil pessoas, representando uma taxa de mortalidade de apenas 0,034. Esta taxa é muito baixa para o Brasil, pois em 1950 e 1960 as taxas de mortalidade infantil foram estimadas em 0,170 e 0,120 respectivamente (4). Em 1969/70 a taxa de mortalidade infantil deve ter sido da ordem de 0,100 ou, na melhor das hipóteses de 0,080 (5).

Aplicando-se a taxa de mortalidade infantil de 0,080 à estimativa de filhos nascidos vivos durante o período 1/9/69 a 1/9/70, teríamos a mortalidade infantil de 274 mil pessoas. Conseqüentemente haveriam 3 148 mil crianças com menos de 1 ano de idade por ocasião do Censo.

Comparando-se este resultado de 3 148 mil crianças com os 2 798 mil com menos de 1 ano de idade, segundo o Censo, teríamos uma omissão da ordem de 350 mil pessoas. Por outro lado, considerando a estimativa de 3 326 000 filhos nascidos vivos no período 2/9/68 a 1/9/69 e aplicando-se a essa estimativa à taxa de mortalidade infantil de 0,080, teríamos 266 mil óbitos infantis. A população com 1 ano de idade em 1/9/69, resultante dos nascimentos do ano imediatamente anterior, seria de 3 060 000 pessoas.

Aplicando-se novamente a taxa de mortalidade (0,015) (6) às 3 060 000 pessoas com 1 ano de idade em 1/9/69 teríamos uma estimativa de 46 000 óbitos para essa idade, resultando uma estimativa de 3 014 000

---

(3) Estudo Preliminar do CBED, publicado na R. bras. Estat. de jul/set de 1971, estima uma omissão da mesma ordem de grandeza.

(4) Diagnóstico Preliminar de Demografia - EPEA.

(5) Anuário Estatístico do Brasil - 1972.

(6) Valor adotado com base na taxa argentina, cuja mortalidade infantil era de 0,080. (Demographic Yearbook - 1966).

pessoas de 1 a menos de 2 anos de idade em 1/9/70.

Comparando-se essa estimativa de 3 014 000 pessoas com o resultado do Censo, 2 499 000 pessoas, poderíamos dizer que houve omissão de 515 000 pessoas de 1 a menos de 2 anos, na investigação do Censo de 1970.

Do exposto conclui-se que teria havido uma tendenciosidade no Censo nas declarações do número de filhos nascidos vivos no período de 1/9/69 a 1/9/70, na população com menos de 1 ano e na população de 1 a menos de 2 anos, respectivamente da ordem de grandeza de 15%, 11% e 20%.

#### 4 - AVALIAÇÃO DA PRECISÃO DA COLETA

Durante a Coleta do Censo vários erros são cometidos. Alguns ligados ao processo de levantamento dos dados, como: omissão de domicílios e pessoas, invasão de setores censitários, interpretação indevida dos conceitos de domicílio vago e fechado, duplicidade, etc. Essas ocorrências constituem os Erros de Cobertura do Censo.

Outros são oriundos do preenchimento incorreto dos Boletins de Coleta, devido à má aplicação de conceitos, interpretações errôneas, má vontade do informante, coleta das informações através de terceiros ou moradores incapacitados a prestá-las, etc. Tais ocorrências constituem os erros de preenchimento.

A Pesquisa de Avaliação foi planejada para detetar a ocorrência desses dois grandes grupos de erro e avaliar a medida da intensidade de suas ocorrências.

As informações da Pesquisa de Avaliação foram obtidas através de um esquema de amostragem desenvolvido em dois estágios. No primeiro estágio foram selecionados 1/100 dos setores censitários, estratificados por situação urbana e rural. No segundo estágio, dentro de cada setor selecionado no 1º estágio, foram investigados 1/10 dos domicílios particulares ocupados.

Durante o primeiro estágio, foi preenchida novamente uma Caderneta de Coleta fazendo-se o cadastramento completo de domicílios e pes

soas em cada setor selecionado, ignorando-se os trabalhos anteriormente executados durante a realização do Censo.

No segundo estágio, preencheu-se um Boletim de Coleta em cada domicílio selecionado para a amostra.

As tendenciosidades ocorridas na coleta foram reunidas em dois grupos: cobertura e preenchimento dos questionários. Os processos adotados na caracterização dessas tendenciosidades são desenvolvidos a seguir.

#### 4.1 - TENDENCIOSIDADE DE COBERTURA OCORRIDA NA COLETA

##### 4.1.1 - Caracterização e dimensão das ocorrências

As omissões e duplicações de registros, bem como a invasão de setores e a não cobertura de áreas foram caracterizadas mediante o confronto das Cadernetas do Recenseador e Folhas coletadas no Censo com a Caderneta do Pesquisador e Folhas de Coleta preenchidas na coleta da Pesquisa de Avaliação, em cada setor selecionado para a amostra da Pesquisa de Avaliação.

Sempre que as informações coletadas na Pesquisa de Avaliação apresentavam divergências superiores a 5% em relação ao Censo, especialmente no caso de omissões, o confronto foi feito também com as cadernetas do Censo dos setores limítrofes dos selecionados para a Amostra.

Para a caracterização das ocorrências resultantes do confronto dos cadastros do Censo e da Pesquisa, foram estabelecidos dez códigos correspondentes às situações:

Código 1 - mesma família recenseada no Censo e na Pesquisa.

Código 7 - família pesquisada e não recenseada; e família recenseada e não pesquisada.

O significado dos códigos empregados figura no Anexo 5.

Ao realizar-se o confronto das informações registradas em cada uma das linhas das Folhas de Coleta do Censo com as registradas nas Folhas de Coleta da Pesquisa, caracterizaram-se as ocorrências, registrando-se em cada linha o código correspondente.

Na situação urbana o confronto foi feito procurando-se casar os registros de cada linha da Folha de Coleta de Censo com os registros das linhas da Folha de Coleta da Pesquisa. Assim, buscou-se principalmente caracterizar se um domicílio registrado no Censo foi registrado na Pesquisa, e ainda se estava ocupado em ambas ocasiões pelos mesmos responsáveis ou por outros, havendo razões para tal.

Na situação rural, o procedimento foi análogo ao indicado acima sempre que possível, pois a falta de numeração tornam mais difícil a identificação. Além disso, quando o confronto revelava alta taxa de omissão, também foram utilizadas as Cadernetas do Censo Agrícola para auxiliar a explicação das divergências de arrolamento entre o Censo e a Pesquisa.

A apuração das ocorrências segundo os códigos possibilitou a obtenção dos resultados das tendenciosidades.

As estimativas das tendenciosidades de cobertura foram obtidas através do processo de estimativa de razão, onde a razão encontrada na amostra, por exemplo, da omissão de pessoas, foi aplicada aos resultados do Censo.

#### 4.1.2 - RESULTADOS PRELIMINARES DA AVALIAÇÃO DA PRECISÃO DA COBERTURA DA COLETA

Até o presente momento, dezembro de 1973, foram apurados o material referente a 24 Unidades da Federação correspondendo a 75% dos setores que compõem a amostra.

As apurações preliminares das Regiões: Norte, Sul e Centro-Oeste e as respectivas estimativas de suspeita de omissão já estão concluídas.

Para as Regiões: Nordeste e Sudeste os resultados disponíveis referem-se, apenas às Unidades da Federação já apuradas.

Procedeu-se a estimativas para o conjunto do Brasil, admitindo-se que as taxas de omissão avaliadas para o total das 24 Unidades da Federação, representem o País.

Na tabela a seguir, são apresentadas, em caráter preliminar, as estimativas de suspeita de omissão da população, segundo as Regiões.

TABELA - 9

ESTIMATIVA DA SUSPEITA DA POPULAÇÃO OMITIDA, SEGUNDO  
A SITUAÇÃO URBANA E RURAL, POR REGIÕES

REGIÕES FISIOGRÁFICAS	POPULAÇÃO OMITIDA					
	Total		Urbana		Rural	
	Absoluto	Rela- tivo (%)	Absoluto	Rela- tivo (%)	Absoluto	Rela- tivo (%)
BRASIL (1) ...	6 491 790	6,9	2 539 428	4,8	3 952 362	9,5
TOTAL APURADO	4 855 457	6,9	1 934 730	4,8	2 920 727	9,5
Norte .....	273 340	7,4	59 294	5,7	214 046	10,6
Nordeste (2) .....	1 807 160	8,6	416 131	4,7	1 391 029	11,4
Sudeste (3) .....	1 445 707	6,1	1 096 094	5,7	349 613	8,0
Sul .....	926 200	5,6	267 160	3,6	659 040	7,1
Centro-Oeste .....	403 050	7,8	96 051	3,9	306 999	11,5

(1) Estimativa baseada nas taxas para o total já apurado.

(2) Exclusive Bahia, ainda não apurado.

(3) Exclusive Minas Gerais e Rio de Janeiro, não apurados.

De acordo com estas estimativas preliminares, avalia-se que, no final das apurações, cerca de 6,4 milhões de pessoas teriam sido omitidas na coleta do Censo, ou sejam, 6,9%.

A omissão esperada nas áreas urbanas é de 2,5 milhões de pessoas enquanto que, nas áreas rurais, é da ordem de 3,9 milhões, constituindo taxas de omissão de 4,8% e 9,5% respectivamente.

A análise a nível regional mostra que, em relação à situação urbana, as taxas de omissão das regiões sul e centro-oeste são as mais baixas, 3,6% e 3,9%, enquanto que, nas regiões norte e sudeste, apresentam taxa mais elevada: 5,7%.

Quanto à situação rural, observa-se que, de modo quase geral, as taxas de omissão representam o dobro das constatadas nas áreas urbanas. Para a Região Sul encontrou-se a taxa mais baixa - 7,1% - enquanto que na Centro-Oeste, a mais elevada - 11,5%.

É oportuno salientar que, de certa maneira, as taxas de omissão encontradas, segundo as situações, são razoavelmente próximas.

#### 4.1.3 - TENDENCIAS OCORRIDAS NO PREENCHIMENTO DO QUESTIONÁRIO

O questionário da Pesquisa de Avaliação foi planejado de modo a medir a dificuldade de preenchimento de vários quesitos, especialmente quanto a condição de presença, parentesco, frequência à escola, rendimentos, etc.

Conjugando o Boletim de Família da Pesquisa de Avaliação com os Boletins de Família da Amostra e da Não-Amostra do Censo, para os mesmos domicílios, é possível avaliar índices de consistência das informações coletadas e, portanto, medir as tendencias ocorridas.

Algumas das tendencias que já analisamos neste trabalho podem ser testadas através da Pesquisa de Avaliação, como por exemplo a forma de declaração da idade, quer através dos Boletins da Amostra quer dos da Não-Amostra; frequência à escola; filhos nascidos vivos durante o úl-

timo ano anterior à data do Censo; número de pessoas com menos de 1 ano e com 1 ano de idade; além disso, torna-se possível a obtenção da taxa de mortalidade infantil.

Ainda não se dispõe de nenhuma informação relativa a tais aspectos, nesta fase da pesquisa, pois não foi iniciado o confronto das informações coletadas pelo Censo e pela Pesquisa.

##### 5 - ALGUMAS CONSIDERAÇÕES SOBRE OS ERROS ESTUDADOS

Neste estudo preliminar foram apontadas algumas tendenciosidades ocorridas em Censos Demográficos brasileiros.

É fundamental ressaltar que em todo levantamento censitário, especialmente em países da dimensão populacional do Brasil, tendenciosidades sempre deverão ocorrer, mesmo em países altamente desenvolvidos como os Estados Unidos, devido à complexidade e às múltiplas tarefas e elevado contingente de mão-de-obra que envolvem operações desta natureza.

A existência de tendenciosidades isoladas prejudica os resultados obtidos, porém não invalida um Censo, salvo quando os erros sejam generalizados e de grande intensidade, tal como ocorreu na França em 1946 (1), o que levou o Instituto Nacional de Estatística a não divulgar parte do Recenseamento, tal como estavam os resultados obtidos.

As tendenciosidades apontadas no Censo Demográfico brasileiro no presente estágio dos estudos, são de dimensões tais que indicam níveis de qualidade condizentes ao estágio de desenvolvimento cultural do Brasil, sendo portanto plenamente aceitáveis.

A propósito, em termos de comparação, recomenda-se a leitura da publicação ACCURACY OF DATA ON POPULATION CHARACTERISTICS AS MESURED BY CPS - CENSUS MATCH SERIES ER 60 N° 5 - BUREAU OF THE CENSUS.

---

(1) Veja-se trabalho publicado no "Journal de la Societé de Statistique de Paris (93° année, Ns. 1-2-3, Janv. - Fév. - Mars. 1952), cuja tradução foi publicada na Revista Brasileira de Estatística.



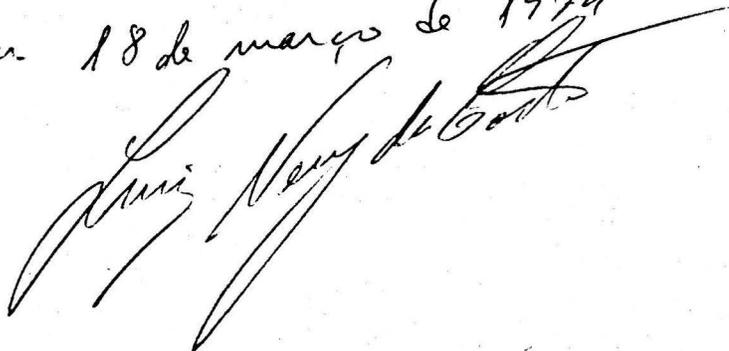
É indiscutível que os resultados de qualquer levantamento inspiram maior confiança e são melhor utilizados quando conhecida a ordem de grandeza da precisão deles.

A grande preocupação da estatística é conhecer, controlar, e evitar as tendenciosidades nos levantamentos estatísticos. Geralmente as tendenciosidades são de magnitude maior do que os erros de amostragem, daí a razão da preocupação atual.

A propósito, no presente estudo, as tabelas anexas 2 e 3 contêm os erros de amostragem dos resultados definitivos do Censo e dos Resultados das Tabulações Avançadas.

Os erros esperados de amostragem das Tabulações Avançadas são superiores aos do Censo, é lógico. Entretanto os níveis desses erros para estimativas da dimensão das tendenciosidades encontradas, são bem inferiores aos erros que as tendenciosidades encerram, corroborando tudo que já foi afirmado.

em 18 de março de 1974

A handwritten signature in dark ink, appearing to read "Rui Nery de Costa". The signature is written in a cursive, flowing style with long, sweeping strokes.

GRUPOS DE CONTROLE PARA ESTIMATIVAS DE AMOSTRAGEM

GRUPOS DE IDADE	MORADORES PRESENTES									MORADORES AUSENTES				NÃO MORADORES PRESENTES			
	Homens				Mulheres					Chefes		Cônjuges e outros		Chefes		Cônjuges e outros	
	População Urbana		População Rural		População Urbana			População Rural									
	Chefes	Cônjuges e outros	Chefes	Cônjuges e outros	Chefes	Cônjuges	Outros	Chefes	Cônjuges	Outros	Homens	Mulheres	Homens	Mulheres	Homens	Mulheres	
0 a 9 anos ...	G <sub>4</sub>		G <sub>12</sub>			G <sub>23</sub>			G <sub>34</sub>								
10 a 19 anos ..	} G <sub>1</sub>	G <sub>5</sub>	} G <sub>9</sub>	G <sub>13</sub>	} G <sub>17</sub>	G <sub>20</sub>	} G <sub>28</sub>	} G <sub>31</sub>	G <sub>35</sub>	} G <sub>39</sub>	} G <sub>40</sub>	} G <sub>41</sub>	} G <sub>42</sub>	} G <sub>43</sub>	} G <sub>44</sub>	} G <sub>45</sub>	} G <sub>46</sub>
20 a 29 anos ..		G <sub>6</sub>		G <sub>14</sub>		G <sub>25</sub>			G <sub>36</sub>								
30 a 39 anos ..	G <sub>2</sub>	G <sub>7</sub>	G <sub>10</sub>	G <sub>15</sub>	G <sub>18</sub>	G <sub>21</sub>	G <sub>26</sub>	G <sub>29</sub>	G <sub>32</sub>	G <sub>37</sub>							
40 anos e mais e idade ignorada.	G <sub>3</sub>	G <sub>8</sub>	G <sub>11</sub>	G <sub>16</sub>	G <sub>19</sub>	G <sub>22</sub>	G <sub>27</sub>	G <sub>30</sub>	G <sub>33</sub>	G <sub>38</sub>							

XXVIII

CENSO DEMOGRÁFICO DE 1970

Erros de Amostragem (valores absolutos aproximados)

TAMANHO DO UNIVERSO	TAMANHO DA ESTIMATIVA																		
	50	100	250	500	1 000	2 500	5 000	10 000	25 000	50 000	100 000	250 000	500 000	1 000 000	2 500 000	5 000 000	10 000 000	25 000 000	
500	23	30	38	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1 000	23	32	45	53	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2 000	24	33	50	66	76	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3 000	24	33	51	68	85	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4 000	24	34	51	70	93	110	-	-	-	-	-	-	-	-	-	-	-	-	-
5 000	24	34	52	72	96	120	-	-	-	-	-	-	-	-	-	-	-	-	-
7 500	24	34	52	72	96	140	-	-	-	-	-	-	-	-	-	-	-	-	-
10 000	24	34	52	74	100	150	170	-	-	-	-	-	-	-	-	-	-	-	-
15 000	24	34	52	74	100	150	210	-	-	-	-	-	-	-	-	-	-	-	-
20 000	24	34	52	75	110	160	210	240	-	-	-	-	-	-	-	-	-	-	-
30 000	24	34	53	75	110	160	220	280	-	-	-	-	-	-	-	-	-	-	-
40 000	24	34	53	75	110	160	220	290	320	-	-	-	-	-	-	-	-	-	-
50 000	24	34	54	76	110	170	230	300	380	-	-	-	-	-	-	-	-	-	-
75 000	24	34	54	76	110	170	230	310	440	-	-	-	-	-	-	-	-	-	-
100 000	24	34	54	76	110	170	230	320	470	540	-	-	-	-	-	-	-	-	-
150 000	24	34	54	76	110	170	240	330	490	620	-	-	-	-	-	-	-	-	-
200 000	24	34	54	76	110	170	240	330	500	660	760	-	-	-	-	-	-	-	-
250 000	24	34	54	76	110	170	240	330	510	680	830	-	-	-	-	-	-	-	-
500 000	24	34	54	76	110	170	240	330	520	720	960	1 200	-	-	-	-	-	-	-
750 000	24	34	54	76	110	170	240	330	530	730	990	1 380	-	-	-	-	-	-	-
1 000 000	24	34	54	76	110	170	240	340	530	740	1 020	1 470	1 700	-	-	-	-	-	-
2 000 000	24	34	54	76	110	170	240	340	530	750	1 050	1 580	2 080	2 400	-	-	-	-	-
3 000 000	24	34	54	76	110	170	240	340	540	750	1 050	1 620	2 190	2 780	-	-	-	-	-
4 000 000	24	34	54	76	110	170	240	340	540	750	1 050	1 640	2 250	2 940	3 170	-	-	-	-
5 000 000	24	34	54	76	110	170	240	340	540	760	1 060	1 650	2 280	3 040	3 800	-	-	-	-
7 500 000	24	34	54	76	110	170	240	340	540	760	1 070	1 670	2 320	3 140	4 380	-	-	-	-
10 000 000	24	34	54	76	110	170	240	340	540	760	1 070	1 670	2 340	3 220	4 650	5 370	-	-	-
25 000 000	24	34	54	76	110	170	240	340	540	760	1 070	1 690	2 370	3 330	5 090	6 790	8 320	-	-
50 000 000	24	34	54	76	110	170	240	340	540	760	1 070	1 690	2 390	3 360	5 230	7 200	9 600	12 000	-

NOTA: Erros calculados para o coeficiente de confiança de 95% em um esquema de amostragem aleatória simples.

UTILIZAÇÃO DA TABELA

Para conhecimento do valor do erro absoluto de uma estimativa observe o procedimento compatível com uma das seguintes situações específicas:

1a. - Dados existentes na tabela. Exemplo: estimativa 50 000 e universo 750 000. O erro absoluto de amostragem (730) será encontrado no cruzamento da linha 750 000 do universo, com a coluna correspondente a estimativa 50 000.

2a. - Inexistência de um dos dados na tabela. Exemplo: estimativa 9 000 (não existente) e universo 750 000. Como as estimativas mais próximas de 9 000 são o par 5 000 e 10 000, terá de ser feito um cálculo de interpolação linear, com base nesses valores. No cruzamento das colunas (5 000 e 10 000) com a linha (750 000) do universo constam os erros absolutos 240 e 330. O valor procurado, resultante da interpolação, será 312.

Observe-se que o erro para a estimativa 10 000 (próxima a 9 000) sendo 330 poderia, sempre que fosse prescindível um maior rigor, ser adotado como uma aproximação, dispensando o cálculo de interpolação.

3a. - Dados inexistentes na tabela. Exemplo: estimativa 6 000 (compreendida entre as colunas 5 000 e 10 000) e universo 25 000 (situado entre as linhas 20 000 e 30 000). Proceder-se à primeira interpolação, com relação a linha 20 000 do universo, encontrando-se o erro absoluto 216; a segunda, relativa a linha 30 000 fornece o erro 232; e a terceira, em relação ao universo de 25 000, usando os resultados 216 e 232: o erro absoluto será 224.

4a. - Estimativa muito próxima do universo. Exemplo: estimativa 40 000 para universo de 50 000. Reformula-se o problema, calculando o erro relativo a essa estimativa, para um universo igual à diferença entre o universo e a estimativa. No exemplo, o universo de referência passa a ser 10 000 e o erro absoluto procurado será 100. Observe-se que, em alguns casos, nos quais as estimativas representam valores próximos de 50% do universo, o erro aproximado é obtido pelo maior valor indicado na tabela para aquele universo.

TABULAÇÕES AVANÇADAS DO CENSO DEMOGRÁFICO DE 1970  
 ESTIMATIVAS DOS ERROS DE AMOSTRAGEM RELATIVOS E ABSOLUTOS  
 PARA OS COEFICIENTES DE CONFIANÇA DE 68% e 95%

Nº DE ORDEM (1)	TAMANHO DA ESTIMATIVA (2)	ERROS DE AMOSTRAGEM			
		Para o Coeficiente de Confiança de 68%		Para o Coeficiente de Confiança de 95%	
		Relativo (%) (3)	Absoluto (4)	Relativo (%) (5)	Absoluto (6)
1	1 000 .....	24,54	245	48,10	481
2	2 500 .....	14,02	351	27,48	687
3	5 000 .....	11,24	562	22,03	1 102
4	7 500 .....	9,24	693	18,11	1 358
5	10 000 .....	8,03	803	15,74	1 574
6	20 000 .....	5,74	1 148	11,25	2 250
7	30 000 .....	4,71	1 413	9,23	2 769
8	40 000 .....	4,10	1 640	8,04	3 216
9	50 000 .....	3,68	1 840	7,21	3 605
10	60 000 .....	3,37	2 022	6,61	3 966
11	70 000 .....	3,13	2 191	6,13	4 291
12	80 000 .....	2,93	2 344	5,74	4 592
13	90 000 .....	2,77	2 493	5,43	4 887
14	100 000 .....	2,63	2 630	5,15	5 150
15	150 000 .....	2,16	3 240	4,23	6 345
16	200 000 .....	1,88	3 760	3,68	7 360
17	300 000 .....	1,54	4 620	3,02	9 060
18	400 000 .....	1,34	5 360	2,63	10 520
19	500 000 .....	1,21	6 050	2,37	11 850
20	750 000 .....	0,99	7 425	1,94	14 550
21	1 000 000 .....	0,86	8 600	1,69	16 900
22	2 000 000 .....	0,62	12 400	1,22	24 400
23	3 000 000 .....	0,51	15 300	1,00	30 000
24	4 000 000 .....	0,44	17 600	0,86	34 400
25	5 000 000 .....	0,39	19 500	0,76	38 000
26	7 500 000 .....	0,32	24 000	0,63	47 250
27	10 000 000 .....	0,28	28 000	0,55	55 000

NOTA: 1. Utilizando-se os Erros de Amostragem para os Coeficientes de Confiança de 68% e 95%, isto significa que 68 e 95 ocorrências, em 100 casos, respectivamente, os erros de amostragem não deverão ser superiores aos valores da tabela. - 2. As variâncias foram estimadas utilizando-se 10 "Random Groups" ou Subamostras Interpenetrantes.

UTILIZAÇÃO DA TABELA - Desejando-se conhecer o Erro de Amostragem de uma estimativa, por exemplo, da estimativa 310 460 estudantes do Brasil do curso superior com idade de 20 a 24 anos, que figura na última coluna da tabela 4 da página 4 das Tabulações Avançadas, com o Coeficiente de Confiança de 95%, procede-se da seguinte maneira:

- procura-se na coluna 2, o número mais próximo do Tamanho da Estimativa 310 460, que no caso é 300 000, na linha 17;
- no cruzamento da coluna 5 com a referida linha, lê-se 3,02% que é o Erro Relativo de Amostragem para a Estimativa de 300 000. Esta percentagem pode ser aplicada à estimativa de 310 460 que está sendo observada, obtendo-se o Erro Absoluto. O Erro Absoluto 1 aproximado, sem efetuar cálculo, é dado pela coluna 6; no exemplo, ele é 9 060;
- O Intervalo de Confiança para a Estimativa de 310 460, é: 310 460 ± 9 060, com a probabilidade de 95 em 100 casos, conter o valor teórico desconhecido.

NÚMERO DE PESSOAS NA AMOSTRA, SEGUNDO OS PESOS RECEBIDOS E GRUPOS DE CONTROLE DE ESTIMATIVA

ESTADO DO PARANÁ

ESPECIFICAÇÃO	GRUPAMENTOS	FREQUÊNCIA DOS PESOS																
		Total	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16 e mais
TOTAL .....	-	1 849 451	139	23 873	553 577	1 121 730	127 075	15 578	3 789	1 091	715	424	284	272	135	312	116	341
CHEFE .....	-	368 002	108	8 503	151 223	192 703	11 577	1 687	877	356	218	165	100	78	53	129	30	195
NÃO-CHEFE .....	-	1 481 449	31	15 370	402 354	929 027	115 498	13 891	2 912	735	497	259	184	194	82	183	86	146
MORADOR PRESENTE																		
Chefes homens urbano .....	1+2+3	120 718	-	1 372	39 280	76 701	3 031	271	20	19	24	-	-	-	-	-	-	-
Outros urbano .....	4 a 8	195 206	-	1 103	44 951	136 583	10 749	1 310	406	36	-	-	-	19	20	29	-	-
Chefes homens rural .....	9+10+11	214 106	89	6 071	101 072	101 654	5 056	162	2	-	-	-	-	-	-	-	-	-
Outros rural .....	12 a 16	394 540	-	4 153	99 989	247 945	36 370	5 211	405	205	93	20	92	6	4	47	-	-
Chefes mulheres urbano ...	17 a 19	15 268	10	427	5 568	8 479	655	85	25	7	6	2	-	1	-	1	-	2
Outros urbano .....	20 a 27	315 677	-	1 854	78 480	218 117	15 523	1 248	322	30	58	13	-	-	17	15	-	-
Chefes mulheres rural .....	28 a 30	9 764	7	543	4 506	4 091	531	70	8	8	-	-	-	-	-	-	-	-
Outros rural .....	31 a 38	544 704	-	7 555	173 776	310 984	46 823	4 503	757	87	54	-	21	89	16	34	5	-
MORADOR AUSENTE																		
Chefes homens .....	39	802	-	-	-	265	537	-	-	-	-	-	-	-	-	-	-	-
Chefes mulheres .....	40	7 344	2	90	797	1 513	1 767	1 099	822	322	188	163	100	77	53	128	30	193
Outros .....	41+42	15 363	11	224	2 150	5 244	4 311	1 252	932	321	261	209	67	80	25	55	78	143
NÃO MORADOR PRESENTE																		
Chefes homens .....	43	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Chefes mulheres .....	44	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Outros .....	45+46	15 959	20	481	3 008	10 154	1 722	367	90	56	31	17	4	-	-	3	3	3

CARACTERIZAÇÃO DE ALGUMAS OCORRÊNCIAS OBSERVADAS NO  
CONFRONTO DAS CADERNETAS DO CENSO E DA PESQUISA

A coleta do Censo Demográfico de 1970 e o da Pesquisa de Avaliação tiveram início respectivamente em 1º de setembro de 1970 e janeiro de 1971.

No interregno destas coletas, vários fatos ocorreram em relação a cada setor censitário, por exemplo: famílias deixaram ou passaram a morar, outras trocaram de residência, etc.

Estes fatos, bem como a necessidade de caracterizar com razoável segurança as omissões, duplicidades e evasões de domicílios e famílias do Censo, deram origem a criação de dez ocorrências básicas (código de 1 a 10).

Nos setores selecionados para a Pesquisa de Avaliação, todos os registros lançados nas folhas de coleta CD-1.60 são confrontados com os respectivos, registrados nas folhas de coleta do Censo CD 1.07 e CD 1.08. Cada linha com registro, nas citadas folhas, recebe um código conforme a natureza da ocorrência que lhe é atribuída. Em relação aos códigos 1 a 5 e 8, sempre que eles são atribuídos às linhas das folhas de coleta do Censo, correspondem a idênticas ocorrências nas folhas de coleta da Pesquisa.

Em anexo expõe-se a definição, o roteiro para caracterização e o código das referidas ocorrências, bem como as indicações técnicas a serem obtidas a partir delas.

CARACTERIZAÇÃO DE ALGUMAS OCORRÊNCIAS OBSERVADAS NO  
CONFRONTO DAS CADERNETAS DO CENSO E DA PESQUISA

A coleta do Censo Demográfico de 1970 e o da Pesquisa de Avaliação tiveram início respectivamente em 1º de setembro de 1970 e janeiro de 1971.

No interregno destas coletas, vários fatos ocorreram em relação a cada setor censitário, por exemplo: famílias deixaram ou passaram a morar, outras trocaram de residência, etc.

Estes fatos, bem como a necessidade de caracterizar com razoável segurança as omissões, duplicidades e evasões de domicílios e famílias do Censo, deram origem a criação de dez ocorrências básicas (código de 1 a 10).

Nos setores selecionados para a Pesquisa de Avaliação, todos os registros lançados nas fôlhas de coleta CD-1.80 são confrontados com os respectivos, registrados nas fôlhas de coleta do Censo CD 1.07 e CD 1.08. Cada linha com registro, nas citadas fôlhas, recebe um código conforme a natureza da ocorrência que lhe é atribuída. Em relação aos códigos 1 a 5 e 8, sempre que eles são atribuídos às linhas das fôlhas de coleta do Censo, correspondem a idênticas ocorrências nas fôlhas de coleta da Pesquisa.

Em anexo expõe-se a definição, o roteiro para caracterização e o código das referidas ocorrências, bem como as indicações técnicas a serem obtidas a partir delas.

OCORRÊNCIAS CARACTERIZADAS E CODIFICADAS NO CONTEÚDO DOS CAPÍTULOS DO CENSO E DA PESQUISA

OCORRÊNCIAS			INDICADORES TÉCNICOS
Código	Definição	Roteiro para caracterização	
1	Mesma família recenseada e pesquisada, residindo no mesmo domicílio	Considerados como tal sempre que: a) Identificam-se os mesmos chefes de domicílios residindo nos mesmos endereços, especialmente na situação Urbana; b) Identificam-se os mesmos chefes de domicílios na situação rural nas mesmas localidades e logradouros; c) Identificam-se os mesmos nomes dos cônjuges, informantes, sobrenomes e pronomes de chefes, com nº de pessoas iguais e mantendo uma sequência na recida de arrolamento nas localidades e logradouros.	A Totalização das pessoas caracterizadas nas ocorrências 1, 2 e 5, possibilita obter: a) uma taxa de acerto do Censo; b) testes de hipóteses sobre níveis de pessoas por domicílios; c) taxa de permanência das famílias nas referidas áreas.
2	Mesma família recenseada e pesquisada residindo em domicílios diferentes, porém ambos arrolados no Censo e Pesquisa	As famílias que entraram para o domicílio, após a coleta do Censo, são identificadas por uma pergunta específica. (coluna 18 da Folha de Coleta CI-1.80) A identificação das famílias é feita análogamente como na ocorrência 1, entretanto, na Pesquisa, a resposta dada na coluna 18 foi "NÃO"	
3	Mesma família recenseada e pesquisada, porém, o domicílio ocupado na pesquisa não foi recenseado ou estava em construção	A identificação das famílias é feita análogamente como acima, entretanto, por ocasião do Censo o domicílio estava em construção (sem ser habitado) ou não fora registrado.	
4	Família pesquisada residindo em domicílio desocupado no Censo, isto é família que se mudou para o setor, após a coleta do Censo	A família pesquisada, cuja resposta a esta pergunta foi Não, isto é, não residia no setor por ocasião do Censo e o domicílio, ora ocupado, estava vago durante a coleta do Censo, foi considerada como mudança (entrada para o setor)  Considerando como tal, famílias nas seguintes situações:	Taxas sobre as mudanças que caracterizam entradas
5	Família pesquisada residindo em domicílio que na ocasião do Censo estava ocupado por outra, isto é, o mesmo domicílio ocupado por famílias diferentes por ocasião de ambas as coletas	a) O domicílio arrolado na pesquisa é o mesmo arrolado no Censo, entretanto, as famílias recenseada e pesquisada são outras; para a família pesquisada a resposta dada à pergunta da coluna 18 foi Não isto é, entrou para o domicílio, após a coleta do Censo b) O domicílio pesquisado é o mesmo recenseado, entretanto a família pesquisada não é a mesma recenseada e a resposta registrada na coluna 18 é Sim. Atribui-se erro de resposta neste caso, somente tolerado, quando há perfeita correspondência do endereço do domicílio.	Taxas sobre mudanças de domicílios
6	Família pesquisada residindo em domicílio não arrolado no Censo	A família pesquisada respondeu Não à pergunta da coluna 18 da folha de coleta e o domicílio em que reside não foi arrolado no Censo, isto é emitido ou construído após o Censo.  Todas as famílias registradas no Censo ou na Pesquisa sem se enquadrarem nas ocorrências 1 a 5 e 8 foram consideradas nesta ocorrência.	Taxas sobre as entradas porém em domicílio novos e/ou emitidos no Censo.
7	Família pesquisada e não recenseada ou vice-versa (omissões).	Na Pesquisa, para alguns domicílios fechados não se registrou o nome da família e o nº de pessoas; entretanto, no Censo, para os mesmos domicílios constam os registros de pessoas residentes e vice-versa. Estas situações foram consideradas também como ocorrência de omissão na Pesquisa ou no Censo, conforme o caso.	Taxas de omissão a) de pessoas b) de famílias
8	Família recenseada morando em domicílio desocupado na Pesquisa, isto é, família que se mudou após o Censo e antes da Pesquisa	Considerou-se como tal, toda família recenseada, cujo respectivo domicílio foi registrado na pesquisa como <u>NÃO</u> (somente nos casos em que o endereço era bem definido ou a sequência de arrolamento no Censo e Pesquisa foram bem semelhantes).	Taxas sobre as mudanças que caracterizam saídas
9	Domicílios desocupados (vagos e fechados)	Todos os domicílios registrados como vagos e fechados que apareceram, respectivamente no Censo e na Pesquisa, foram considerados nesta ocorrência.	Taxas dos domicílios desocupados em relação aos ocupados
10	Unidades não domiciliárias	Foram considerados todos os registros de unidades não domiciliárias, seja que houvesse pessoas residentes.	Taxa de relação de unidades não domiciliárias em relação aos domicílios e população

Obs. - Foram anotadas em observações os domicílios coletivos, as invasões de setores por parte da Pesquisa ou do Censo e outras eventuais ocorrências, tais como desdobramento, aglutinação de domicílios, etc.

TABELA 1

CONFRONTO DOS RESULTADOS DEFINITIVOS COM OS RESULTADOS DAS TABULAÇÕES  
AVANÇADAS PARA O NÚMERO DE FAMÍLIAS E DE DOMICÍLIOS

BRASIL E REGIÕES	NÚMERO DE FAMÍLIAS				NÚMERO DE DOMICÍLIOS			
	Resultados Definitivos (A)	Tabulações Avançadas (B)	Diferença		Resultados Definitivos (C)	Tabulações Avançadas (D)	Diferença	
			Absoluta (B-A)	Rela tiva (%)			Absoluta (D-C)	Rela tiva (%)
BRASIL .....	18 554 426	19 104 086	+549 660	2,96	17 628 699	18 086 336	+457 637	2,60
REGIÃO I .....	631 739	641 437	+ 9 968	1,54	584 379	592 276	+ 7 897	1,35
REGIÃO II .....	901 664	947 190	+ 45 526	5,05	858 738	900 452	+ 41 714	4,86
REGIÃO III .....	2 857 861	2 911 641	+ 53 780	1,88	2 727 434	2 772 767	+ 45 333	1,66
REGIÃO IV .....	1 621 234	1 674 707	+ 53 473	3,30	1 554 696	1 602 967	+ 48 271	3,10
REGIÃO V .....	2 469 830	2 520 133	+ 50 303	2,04	2 381 841	2 420 555	+ 38 714	1,63
REGIÃO VI .....	2 004 712	2 078 511	+ 73 799	3,68	1 883 164	1 948 035	+ 64 871	3,44
REGIÃO VII .....	3 854 327	4 000 076	+145 749	3,78	3 636 138	3 755 186	+119 048	3,27
REGIÃO VIII .....	1 326 589	1 372 865	+ 46 276	3,49	1 272 355	1 307 798	+ 35 443	2,79
REGIÃO IX .....	1 930 203	1 954 149	+ 23 946	1,24	1 813 447	1 828 208	+ 14 761	0,81
REGIÃO X .....	956 267	1 003 377	+ 47 110	4,93	916 507	958 092	+ 41 585	4,54