

FUNDAÇÃO IBGE - INSTITUTO BRASILEIRO DE ESTATÍSTICA
DEPARTAMENTO DE CENSOS - DECEN

VIII RECENSEAMENTO GERAL DO BRASIL

TABULAÇÃO AVANÇADA
DQ
CENSO DEMOGRÁFICO

PLANO DE SUBAMOSTRAGEM

TRABALHO ELABORADO PELO SETOR ESPECIAL DE AMOSTRAGEM
DA DIVISÃO DE PLANEJAMENTO

TABULAÇÃO AVANÇADA DO CENSO DEMOGRÁFICO DE 1970

Projeto do Plano de Subamostragem

Introdução - A apuração de um recenseamento geral, como é o caso do Censo Demográfico de 1970, exige vários anos para a sua concretização, motivado pelas dificuldades inerentes aos levantamentos de tal amplitude, principalmente a existência da grande massa de informações a serem tratadas.

O desenvolvimento do País alcançou estágio tal a não mais permitir atraso nos resultados do Censo Demográfico, exigindo a necessidade de divulgação dos resultados do Censo de 1970 com maior brevidade.

Dêsse modo, a fim de atender a demanda das informações, por parte dos consumidores governamentais e particulares de dados básicos demográficos o DECEM veio a planejar a solução viável para que, dentro de um período relativamente curto de até um ano após a coleta, já estivessem disponíveis os principais resultados globais, pela apuração de uma amostra dos boletins coletados.

A seguir expõe-se tal plano de amostragem.

CARACTERÍSTICAS DO ANTEPROJETO DO PLANO DE AMOSTRAGEM

1. Objetivos - Obtenção de tabulações prévias das características de maior demanda necessárias aos estudos demográficos do País, investigados no Censo Demográfico de 1970.
2. Representatividade das informações - Serão obtidos resultados representativos para os totais de cada Unidade da Federação. Os totais para o Brasil serão obtidos pela soma dos totais das Unidades da Federação.
3. Âmbito da divulgação dos dados - A divulgação será feita por Unidade da Federação e para o conjunto do País.
4. Profundidade da investigação - Serão obtidas as tabulações de acordo com o plano de tabulação elaborado pela Divisão do Censo Demográfico. Não serão divulgadas as informações mais diluídas que apresentem erros de amostragem considerados elevados.
5. Esquema de amostragem adotado - Consoante a metodologia a ser adotada na coleta do Censo Demográfico de 1970, quando todos os domicílios serão recenseados e apenas 25% preencherão um boletim mais detalhado, denominado "Boletim da Amostra", enquanto os restantes 75% preencherão um boletim reduzido denominado "Boletim da não

Amostra" cujo elenco de informações também figura no "Boletim da Amostra", optou-se por uma amostra aleatória simples dos domicílios selecionados para a amostra do Censo.

Dessa forma o esquema adotado constituirá uma subamostra do Censo Demográfico de 1970.

5.1 - Razões que influíram na determinação do tamanho da amostra - Várias razões pesaram nas decisões para a fixação do plano ora apresentado. Dentre elas destacam-se:

- a) as informações serão divulgadas em forma de tabelas cruzadas. A frequência, isto é, a proporção do valor que figura em cada célula é na maioria dos casos pequena em relação ao total geral; assim, fixou-se que tais proporções seriam estimadas com erros de amostragem variáveis quanto menor a proporção maior erro de amostragem, erro este com limite máximo previamente fixado adiante;
- b) a maioria das tabelas dizem respeito às informações da população. Considerou-se básico que o tamanho das amostras deveriam ser determinados levando em conta as características da população;
- c) as frações de amostragem cujos inversos constituem números inteiros facilita os trabalhos de seleção e estimação. Assim, deveriam os tamanhos das amostras, após calculados teoricamente, ser aproximados para números superiores cujos inversos das frações de amostragem facilitassem os trabalhos de seleção.

5.2 - Determinação do tamanho das amostras - O cálculo do tamanho das amostras de cada Unidade da Federação foi obtido com base na projeção de população referente a 1º de setembro de 1970 (Volume Brasil - Sinopse Preliminar do Censo Demográfico de 1960).

Os tamanhos das amostras devem permitir estimar as distribuições: número de moradores por grupos de idade, número de pessoas por domicílios, rendimento mensal e ainda estimar proporções de características da população até um milésimo ($p = 0,001$).

Considerando, ser a proporção $p = 0,001$ muito pequena e que o tamanho global das amostras deveria ser tal que permitisse reduzir o prazo da divulgação dos resultados, admitiu-se que o erro de amostragem de 20% na estimação da referida proporção seria aceitável. Assim, o valor verdadeiro P estaria dentro do intervalo $0,0008 \leq P \leq 0,0012$. Observe-se que com o crescimento de P diminui o erro de amostragem esperado, na sua estimação.

Levou-se em conta que as estimativas estariam dentro de um intervalo de confiança de 95%, isto é, tem-se a probabilidade de $\frac{95}{100}$ de que as estimativas a serem obtidas atendam às condições previstas.

Calculou-se, o tamanho da amostra de modo que

$$Pr \left\{ \hat{\theta} - s \hat{v} \left\{ \hat{\theta} \right\} \leq \theta \leq \hat{\theta} + s \hat{v} \left\{ \hat{\theta} \right\} \right\} \approx (1 - \alpha)$$

onde:

- $\alpha = 0,05$ nível de confiança
- θ = parâmetro em estudo
- $\hat{\theta}$ = estimativa do parâmetro
- δ = múltiplo do desvio padrão
- $\sqrt{V\{\hat{\theta}\}}$ = raiz quadrada da estimativa da variância de $\hat{\theta}$

Utilizou-se a seguinte fórmula: (1)

$$n = \frac{N \hat{\delta}^2 (1 - p)}{\delta^2 (1 - p) + (N - 1) p \cdot \xi_r^2}$$

onde n = tamanho da amostra

N = tamanho da população (número de pessoas no Censo)

p = proporção desejada; p = 0,001

ξ_r^2 = quadrado do erro relativo de amostragem, ou variância relativa.

Na tabela 2, em anexo, apresentam-se alguns tamanhos de amostras para estimar a proporção P satisfazendo aos níveis de significância mais usuais.

No presente caso, atendendo as condições estabelecidas, utiliza-se a quarta coluna da referida tabela para encontrar-se o tamanho da amostra necessária a estimação de P = 0,001 com um nível de significância $\alpha = 0,20$, inerentes ao tamanho N da população que figura na 1.ª coluna.

A população de cada Unidade da Federação, projetada para 1.º de setembro de 1970, na maioria dos casos, é superior a um milhão de habitantes. Em tais situações tomou-se o tamanho da amostra como sendo n = 100 mil que é válido para o caso de N (tamanho da população) ser muito grande. Os tamanhos das amostras figuram na coluna (4) da tabela 1.

Para cada Unidade da Federação calculou-se a respectiva fração geral de amostragem $f = \frac{n}{N}$ (razão entre o número esperado de pessoas a participar da subamostra e a população projetada para 1970), arredondando-se os valores encontrados. Tais valores figuram na coluna (5) da tabela 1.

Sendo $f = f_1 \cdot f_2$, isto é, a fração geral de amostragem é igual ao produto da fração de amostragem f_1 , da primeira fase, pela fração de amostragem f_2 , da 2.ª fase, então, conhecendo-se f e f_1 obtém-se f_2 pela relação $f_2 = \frac{f}{f_1}$.

Observe-se que o valor esperado de f_1 é a constante: $f_1 = \frac{1}{4} = 0,25$.

(1) Sendo o valor de N muito grande no Censo, - é da ordem de muitos milhares -, então a distribuição hipergeométrica tem por limite a distribuição binominal, e o tratamento dado fica justificado.

Exemplo: Sendo a população projetada da Guanabara para 1º de setembro de 1970 igual a $N = 4498$ mil habitantes, devendo participar da subamostra $n = 100$ mil habitantes, tem-se a fração teórica geral de amostragem

$$f = \frac{n}{N} = \frac{100}{4498} = \frac{1}{45} .$$

A fração de subamostragem f_2 será

$$f_2 = \frac{f}{f_1} = \frac{\frac{1}{45}}{\frac{1}{4}} = \frac{1}{11} .$$

O valor de f_2 , isto é, da fração de subamostragem a ser adotado para cada Unidade da Federação figura na coluna (8) da tabela 1.

Em se adotando um processo de seleção sistemática, o intervalo de seleção é igual ao inverso da fração de subamostragem, ou seja, é igual a $\frac{1}{f_2}$.

O número de pessoas que deverão participar da subamostra em cada Unidade da Federação é obtido aplicando-se a fração de amostra geral final à estimativa da população em 1970. (Veja-se coluna (6) da tabela 1).

Uma estimativa aproximada do número de domicílios que deverão participar da subamostra, em cada Unidade da Federação é obtida dividindo-se o número de pessoas que participarão da subamostra pela média de pessoas por domicílio em 1960 (veja-se coluna 7 do anexo 1).

Os tamanhos dessas subamostras permitirão a obtenção de estimativas das características de domicílios e de famílias que correspondam a proporções $p = 0,05$, ou seja 5% com erros de amostragem provavelmente inferiores a 10%.

Na amostra do censo as famílias conviventes moradoras em domicílios selecionados para a amostra preenchem boletins de amostra. Nessa subamostra o critério será idêntico, isto é, todos os boletins das famílias conviventes de um domicílio selecionado para a subamostra serão incluídos nela.

6. Seleção da subamostra - As unidades de seleção da subamostra são as mesmas da amostra, ou seja, os domicílios particulares ocupados e os grupos conviventes e famílias recenseadas em domicílios coletivos.

A seleção das unidades da subamostra será sistemática e independente para cada Município, com o início da série estabelecida por um número aleatório.

A seleção será feita através das folhas de coleta (CD-1.07 e CD-1.08) e das listas de domicílios coletivos (CD-1.03), após a conferência dos instrumentos de coleta e durante a recepção.

Através da coluna 20 da folha de coleta serão identificados os domicílios particulares selecionados para a amostra. Mediante a contagem destes far-se-á a seleção da subamostra. Nesta ocasião será assinalado o número do questio-

nário (coluna 18 da fôlha de coleta) de cada domicílio selecionado.

A seleção dos grupos conviventes em domicílios coletivos será feita através da contagem dessas unidades registradas nas linhas de amostra das 1 listas de domicílios coletivos. Nessa ocasião o número do questionário selecionado para a subamostra será circundado.

Posteriormente, através dos números dos questionários correspondentes às unidades selecionadas para a subamostra, serão separados os boletins e guardados em pastas especiais que terão prioridade nos trabalhos. Todos os questionários pertencentes às famílias conviventes dos domicílios particulares selecionados para a subamostra nela serão incluídos.

7. Estimativas. - Uma série de variáveis e atributos pesquisados no censo demográfico são fortemente influenciados pelo fato da população viver em regiões urbanas e rurais.

Do mesmo modo a maioria das tabelas dos planos de tabulação preliminar e avançada deverão conter resultados segundo as situações urbana e rural.

A utilização de processos de estimação que levem em conta êsse fato melhoram sensivelmente a qualidade dos resultados.

Por outro lado a adoção de estimadores que limitem as tendenciosidades da coleta e da operação ao próprio Município tornam mais eficientes as estimativas.

Assim sendo, será adotado um processo de estimativa de razão com estratificação. Serão utilizados dois fatores de ponderação, um correspondente ao estrato urbano e outro para o rural.

Êsse processo de estimação fica assim definido:

$$\hat{T}_{h \neq 1} = \sum_{i=1}^2 \hat{T}_{i, h \neq 1}$$

sendo $i=1, 2$

$$\hat{T}_{i, h \neq 1} = K_i T_{i, h \neq 1}^*$$

$$K_i = \frac{T_{i, h=1}}{T_{i, h=1}^*}$$

onde

$\hat{T}_{h \neq 1}$ = estimativa do total da variável ou atributo $h \neq 1$;

$\hat{T}_{i, h \neq 1}$ = estimativa do total da variável ou atributo $h \neq 1$ correspondente ao estrato geográfico i ;

K_i = fator de ponderação correspondente ao estrato i ;

$T_{i, h=1}$ = total da população recenseada no estrato i , resultado da apuração preliminar do Censo;

$T_{i, h=1}^*$ = total da população recenseada, na subamostra do estrato i ;

$T_{i, h \neq 1}^*$ = total da variável ou atributo $h \neq 1$, da subamostra do estrato i ;

As estimativas referentes a população para o conjunto da Unidade da Federação, será a soma das estimativas municipais obtidas pela fórmula (1).

As estimativas referentes às características de domicílios particulares e de famílias serão obtidas de modo semelhante ao estimado (1). Para isso basta ponderar os resultados da tabulação da subamostra pelos fatores K_i referentes aos estratos urbano e rural.

Cabe alertar que as estimativas isoladas para cada Município apresentam erros elevados. Em geral quanto maior for o Município, quanto à população, menores serão esses erros.

Portanto as estimativas municipais isoladas não deverão ser utilizadas sem prévio exame e condensação de tabelas.

TABULAÇÃO AVANÇADA DO CENSO DEMOGRÁFICO DE 1970

Estimativa da população, do número de domicílios e do tamanho da subamostra

UNIDADES DA FEDERAÇÃO	ESTIMATIVA PARA 1-IX-1970		AMOSTRA (1)		AMOSTRA FINAL		FRAÇÃO DE SUBAMOS- TRAGEM f_2
	População (1000 hab) N	Domicílios ocupados (1000 dom) (3)	Tamanho (pessoas) (1000 hab) n	Fração geral de amostragem f	Número de pessoas (n) (1000 hab) (6)	Domicílios ocupados (1000 dom) (7)	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
BRASIL	97 164	18 403	2 376	$\frac{1}{40} = 0,025$	2 521	470,6	-
<u>Norte</u>							
Rondônia	132	33	50	1/3	33	8	1
Acre	219	44	71	1/3	55	11	1
Amazonas	986	164	100	1/10	123	21	1/2
Roraima	47	8	33	1	12	2	1
Pará	2 095	349	100	11/20	105	18	1/5
Amapá	123	21	50	1/3	31	5	1
<u>Nordeste</u>							
Maranhão	3 849	770	100	1/38	107	21	1/9
Piauí	1 494	249	100	1/15	120	20	1/3
Ceará	4 062	677	100	1/41	102	17	1/10
Rio G. Norte	1 358	272	100	1/14	109	21	1/3
Paraíba	2 349	470	100	1/23	117	23	1/5
Pernambuco ..	4 973	995	100	1/50	104	21	1/12
Alagoas	1 458	292	100	1/15	117	23	1/3
Fern.Noronha	3	1	5	1	3	0,6	1
Sergipe	887	177	83	1/9	111	22	1/2
Bahia	7 309	1 462	100	1/73	102	20	1/18
<u>Sudeste</u>							
Minas Gerais	12 404	2 067	100	1/124	103	17	1/30
Esp. Santo ..	2 331	389	100	1/23	117	20	1/5
Rio Janeiro	4 947	989	100	1/49	103	21	1/12
Guanabara ...	4 498	900	100	1/45	103	21	1/11
São Paulo ...	18 102	3 620	100	1/181	100	20	1/45
<u>Sul</u>							
Paraná	8 482	1 696	100	1/85	106	21	1/20
Santa Catarina	2 915	486	100	1/29	104	17	1/7
Rio G. do Sul	7 014	1 403	100	1/70	103	21	1/17
<u>Centro-Oeste</u>							
Mato Grosso .	1 552	259	100	1/16	96	16	1/4
Goiás	3 075	513	100	1/31	110	18	1/7
Dist. Federal	500	100	83	1/5	125	25	1

NOTA - Os dados para o Brasil foram obtidos por soma das Unidades da Federação.

(1) - Tamanho calculado para estimar a proporção $p = 0,001$ com erro relativo de amostragem de 20%.

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

TAMANHO DA POPULAÇÃO N	P = 0,001				P = 0,01				P = 0,05				P = 0,10			
	Nível de significância %				Nível de significância %				Nível de significância %				Nível de significância %			
	$\alpha=0,03$	$\alpha=0,05$	$\alpha=0,10$	$\alpha=0,20$	$\alpha=0,01$	$\alpha=0,03$	$\alpha=0,05$	$\alpha=0,10$	$\alpha=0,01$	$\alpha=0,03$	$\alpha=0,05$	$\alpha=0,10$	$\alpha=0,01$	$\alpha=0,03$	$\alpha=0,05$	$\alpha=0,10$
1 000	1 000	999	998	990	1 000	998	994	976	999	988	968	884	997	976	852	783
2 500	2 499	2 496	2 484	2 439	2 498	2 486	2 461	353	2 492	2 428	2 310	1 881	2 484	2 353	2 130	1 475
5 000	4 994	4 984	4 938	4 762	4 993	4 950	4 841	4 444	4 973	4 720	4 289	3 016	4 932	4 444	3 711	2 094
10 000	9 978	9 938	9 756	9 091	9 975	9 778	9 407	8 888	9 870	8 941	7 525	4 318	9 730	8 000	5 901	2 647
25 000	24 860	24 615	23 529	20 000	24 845	23 669	21 622	15 385	24 204	19 289	13 718	5 828	23 377	15 385	9 137	3 147
50 000	49 444	48 485	44 444	33 333	49 383	44 898	38 004	22 222	47 500	31 405	18 905	6 597	43 902	22 222	11 180	3 358
75 000	73 755	71 642	63 158	42 857	73 620	64 009	50 856	26 087	68 345	39 694	21 624	6 901	62 069	26 087	12 081	3 435
100 000	97 800	94 118	80 000	50 000	97 537	81 481	61 300	28 571	88 372	45 783	23 313	7 063	78 260	28 571	12 587	3 474
250 000	236 686	216 216	153 846	71 429	235 154	159 420	96 964	34 483	188 119	63 129	27 104	7 376	147 541	34 483	13 616	3 549
500 000	449 438	380 952	222 222	83 333	443 946	234 043	120 300	37 037	301 587	72 243	28 658	7 486	209 302	37 037	13 997	3 574
1 000 000	816 327	615 385	285 714	90 909	798 387	305 556	136 740	38 462	431 818	77 869	29 503	7 542	264 706	38 462	14 195	3 587
∞	4 444 444	1 600 000	400 000	100 000	3 968 000	440 000	158 400	40 000	760 000	84 444	30 400	7 600	360 000	40 000	14 400	3 600

NOTA: O erro relativo correspondente ao nível de confiança α é de $\xi_r = \alpha\%$
 Utilizou-se a seguinte fórmula:

$$n = \frac{N \delta^2 (1 - P)}{\delta^2 (1 - P) + \xi_r^2 P (N - 1)}; \text{ onde}$$

n = tamanho da amostra

N = tamanho da população

δ = múltiplo do desvio padrão

δ^2 = variância relativa

ξ_r = erro relativo correspondente ao nível de confiança α

P = proporção do atributo X

TAMANHO DE AMOSTRA ALEATÓRIA SIMPLES NECESSÁRIO À ESTIMAÇÃO DA MÉDIA
 (Para $\xi = 1,96 = 2$)

TAMANHO DA POPULAÇÃO N	COEFICIENTE DE VARIAÇÃO $\gamma = \frac{\sigma \{x\}}{\mu \{x\}}$											
	$\gamma = 0,5$				$\gamma = 1,0$				$\gamma = 1,5$			
	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	662
5 000	3 333	909	370	98	4 444	2 353	1 212	370	4 738	3 333	2 094	762
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	4 082	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 $\xi r = \alpha \%$

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

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