

ANEXO A

DESENHO DA PESQUISA

Descrevem-se, neste anexo, as características principais do desenho e implementação da amostra da PNDS 1996. A caracterização do desenho da amostra inclui: tamanho da amostra; escolha dos domínios geográficos; estágios do processo amostral; estratificação da amostra; grau de aglomeração; relação entre desenho da amostra e natureza do marco amostral. Uma descrição mais completa dos desenhos de amostra das pesquisas DHS encontra-se no Manual de Amostragem, Série Documentação Básica nº 8, pp. 59-66.

A implementação da amostra refere-se a qualquer trabalho cartográfico e de listagem necessário para atualizar, aprimorar ou gerar a seleção definitiva de domicílios e entrevistas individuais.

A.1 Introdução

A amostra utilizada para a Pesquisa PNDS 1996 é uma amostra probabilística selecionada em duas etapas: a primeira etapa consiste na seleção dos setores censitários e a segunda, na seleção de domicílios. A amostra se constitui de 842 setores censitários, distribuídos proporcionalmente nas zonas urbana e rural de cada estado da federação, com exceção dos estados de Rondônia, Acre, Amazonas, Roraima, Pará, Amapá e Tocantins (inclui um setor na área rural), todos eles localizados na região Norte do país. A amostra permite sólidas estimativas de um número de variáveis importantes para cada uma das sete regiões da PNAD e para quatro estados da federação que foram sobreamostrados (Rio Grande do Norte, Pernambuco, Bahia e Ceará). O desenho também permite estimativas independentes para os estados do Rio de Janeiro, São Paulo, Minas Gerais e Rio Grande do Sul. A amostra não é autoponderada em nível de regiões, total urbano, total rural e todo o país. Acrescenta-se ainda, que para os estados de Pernambuco, Bahia e Ceará a amostra é representativa a nível urbano e rural.

Uma amostra total de 16.838 domicílios foi selecionada com o objetivo de obter 13.613 entrevistas completas de mulheres de 15-49 anos de idade. O número de domicílios ocupados foi 14.252, sendo neles realizadas 13.283 entrevistas completas de domicílios. Nesses domicílios encontrou-se um total de 14.579 mulheres elegíveis e obteve-se um total de 12.612 entrevistas completas.

A.2 Desenho e seleção

Universo de estudo

O universo da amostra estudada representa, aproximadamente, 97% da população total do país. Excluiu-se a área rural da região Norte, com exceção do estado de Tocantins, área extensa e pouco habitada, uma vez que o custo total implicado no estudo de fração populacional muito pequena do país (3%) seria extremamente alto.

Unidades amostrais

As Unidades Primárias da Amostra (UPA) são áreas censitárias definidas pelo censo populacional de 1991. Todo domicílio localizado dentro de cada área censitária é considerado como uma Unidade Secundária da Amostra (USA).

Estratificação

Com o objetivo de assegurar maior precisão das estimativas (com menor erro amostral) as UPA's estão agrupadas por áreas de residência rural e urbana, dentro de cada estado da federação.

Marco Amostral

O Instituto Brasileiro de Geografia e Estatística (IBGE) mantém desde 1967, uma amostra principal que tem sido ampliada e aperfeiçoada para atender a diferentes objetivos de pesquisa de base domiciliar nas últimas décadas. A última seleção dessa amostra, denominada PNAD (Pesquisa Nacional por Amostra de Domicílios), foi elaborada usando como marco amostral o censo populacional de 1991. O IBGE mantém um banco de dados com informações sobre os setores censitários da amostra da PNAD, tais como os parâmetros de seleção e a última atualização de domicílios em cada setor dessa amostra. A atualização de domicílios é realizada anualmente e a última corresponde a 1995.

Em consequência, decidiu-se utilizar a amostra principal da PNAD como marco amostral para os objetivos da pesquisa PND 1996, devido, primeiro, à viabilidade de uma subseleção de setores para a pesquisa PND 1996 e, segundo, à atualização dos endereços dos domicílios em cada setor selecionado, realizada durante o ano de 1995 pelo IBGE.

Composição da Amostra

Com o objetivo de obter um número aproximado de 13.600 entrevistas completas de mulheres elegíveis, utilizaram-se informações da Pesquisa sobre Saúde Familiar no Nordeste Brasil, 1991, como instrumento de auxílio para estimar alguns parâmetros do desenho. Estimou-se que seria necessário um total de 16.800 domicílios a serem selecionados para chegar a aproximadamente 13.600 entrevistas completas. Esse total foi distribuído por regiões da federação, não de acordo com sua população, mas para cumprir com os objetivos da pesquisa, como se pode observar na Tabela A.1.

Tabela A.1 Entrevistas de mulheres e número de setores selecionados

Tamanho esperado de entrevista de mulheres e número de setores selecionados por região e estado. Brasil, PNDS 1996.

Região e estado	Número de domicílios		Número esperado de mulheres	Número de setores		
	Desenho proporcional	Desenho não proporcional		Área urbana	Área rural	Total
Rio de Janeiro	1701	1500	1213	72	3	75
São Paulo	3974	2200	1779	103	7	110
Sul						
Paraná	1,027	700	566	26	9	35
Santa Catarina	556	500	404	18	7	25
Rio Grande do Sul	1229	1000	809	39	11	50
Centro-Leste						
Minas Gerais	1835	1200	970	46	14	60
Espírito Santo	303	500	404	19	6	25
Nordeste						
Maranhão	488	435	352	9	13	22
Piauí	253	225	182	6	5	11
Ceará	657	1049	848	35	17	52
Rio Grande do Norte	253	632	511	22	10	32
Paraíba	337	315	255	11	5	16
Pernambuco	791	1265	1023	46	17	63
Alagoas	253	240	194	7	5	12
Sergipe	168	150	121	5	2	7
Bahia	1246	1522	1231	47	29	76
Norte						
Rondônia	67	172	139	9	NA	9
Acre	34	69	56	3	NA	3
Amazonas	135	309	250	15	NA	15
Roraima	17	35	28	2	NA	2
Pará	253	516	417	26	NA	26
Amapá	17	35	28	2	NA	2
Tocantins	101	103	83	5	1	6
Centro-Oeste						
Distrito Federal	185	400	323	19	1	20
Mato Grosso do Sul	219	400	323	18	2	20
Mato Grosso	219	866	700	34	9	43
Goiás	488	500	404	22	3	25
Total	16838	16838	13613	666	176	842

NA = Não se aplica.

Nota: Para o estado de Mato Grosso só foi possível completar 30 setores, sendo 25 na área urbana e 5 na área rural.

Seleção da Amostra

A amostra para cada domínio ou região é obtida em duas etapas: a primeira etapa consiste na seleção sistemática dos setores censitários com probabilidade proporcional ao seu tamanho. Na realidade, essa primeira etapa é uma combinação de procedimentos: a seleção dos setores do censo para a pesquisa PNAD e uma subseleção desses setores para a pesquisa PNDS 1996. Depois de selecionados os setores, utilizou-se a lista atualizada de domicílios fornecida pelo IBGE para realizar a seleção última dos domicílios. Finalmente definiu-se o número de domicílios a serem selecionados em cada setor censitário, de tal maneira que se mantivesse uma fração amostral uniforme para cada domicílio em cada região. Essa fração amostral não é a mesma para cada região, uma vez que o dimensionamento da amostra não foi proporcional em todas as regiões estudadas.

O processo de seleção dos setores censitários deu-se da seguinte forma:

O total da amostra foi distribuído entre cada uma de suas regiões, de acordo com o tamanho requerido e, dentro de cada região, entre seus correspondentes estados, de acordo com sua população;

O número de setores censitários em cada estado foi calculado de acordo com a divisão da amostra por estado e tamanho médio da amostra em cada setor do censo.

Para cada região se obteve uma lista de setores censitários por estado, de acordo com a pesquisa PNAD.

Em cada estado realizou-se uma seleção sistemática do número de setores censitários requeridos para esta pesquisa, digamos a , e um total de setores censitários da pesquisa PNAD, designado por b . Temos então uma taxa de subamostragem para a pesquisa PNDS 1996 como: $s = a/b$.

Originalmente, a probabilidade de seleção de i -ésimo setor censitário para a pesquisa PNAD foi:

$$P_{ii} = (b * m_i) / (\sum m_i)$$

onde:

b : número de setores censitários da pesquisa PNAD,
 m_i : número de domicílios no i -ésimo setor censitário do censo de 1991,
 $\sum m_i$: número de domicílios no estrato de seleção segundo o censo de 1991.

A combinação da seleção dos setores censitários com probabilidades de seleção proporcional ao tamanho da pesquisa PNAD com a subseleção sistemática destes setores resulta numa amostra final de setores com probabilidade proporcional ao tamanho, similar à amostra original.

A probabilidade final do i -ésimo setor censitário para a pesquisa PNDS 1996 é dada por:

$$P_{ii}^* = s * (b * m_i) / (\sum m_i)$$

$$P_{ii}^* = (a * m_i) / (\sum m_i)$$

onde:

a : número de setores censitários para a pesquisa PNDS 1996;

m_i : número de domicílios no i-ésimo setor censitário segundo o censo de 1991;

Σm_i : número de domicílios no estrato de seleção segundo o censo de 1991.

Depois de obter as listas atualizadas para cada setor censitário selecionado, determina-se o número final de domicílios de acordo as seguintes fórmulas:

$$P_{2ij}^* = (n_i / L_i)$$

$$P_{li}^* * P_{2ij}^* = f$$

daí obtém-se

$$n_i = (b * f * L_i) / (a * P_{li})$$

onde

n_i : número de domicílios selecionados no setor para pesquisa **PNDS 1996**,

L_i : número total de domicílios no setor censitário na PNAD de 1995,

f : fração amostral na região'

a : número de setores censitários para pesquisa **PNDS 1996**,

b : número de setores censitários da PNAD,

P_{li} : probabilidade de seleção do setor censitário na PNAD e cujo valor é calculado com base nos parâmetros do desenho da amostra principal como :

$$P_{li} = (n\text{-mun-se})(\text{prob-mun})(n\text{-set-se})(\text{prob-set})$$

onde :

$N\text{-mun-se}$: número de municípios selecionados na PNAD,

Prob-mun : probabilidade de seleção do município na PNAD,

$N\text{-set-se}$: número de setores selecionados no município na PNAD e

Prob-set : probabilidade de seleção do setor na PNAD.

Uma vez estabelecido o número de domicílios a serem selecionados, o procedimento de seleção dos mesmos depende de o setor ser urbano ou rural. Para cada setor urbano, a seleção dos domicílios foi realizada de forma sistemática, com intervalo de seleção de $I_{ij} = L_i / n_i$. Para cada setor rural, a seleção dos domicílios foi realizada com grupos compactos de domicílios contínuos sobre a lista de domicílios e de acordo com o seguinte procedimento:

Numeramos os domicílios de forma consecutiva, de acordo com a lista do setor censitário rural.

Em cada múltiplo de oito (ou seja, 8, 16, 24 etc.) traça-se uma linha com um marcador colorido.

Cada grupo é formado por domicílios entre linhas horizontais consecutivas, sendo o primeiro grupo constituído pelo primeiro domicílio na lista até o anterior imediato à primeira linha horizontal. No caso de existir mais de quatro domicílios depois da última linha horizontal, o último grupo é constituído por esses domicílios depois da linha horizontal. Todos esses grupos assim formados se denominam grupos compactos, sendo G_i o total desses grupos.

Se houver domicílios (menos de quatro) depois da última linha horizontal, os mesmos foram distribuídos uniformemente entre os grupos compactos anteriores, em forma sistemática e com início do sorteio.

Calcula-se o número de grupos compactos a selecionar como :

g_i = número arredondado mais próximo de $(n_i/8)$.

Selecionamos então g_i grupos compactos do total G_i numa forma sistemática e com início de sorteio.

Todos os domicílios dentro do grupo compacto no setor censitário foram considerados como amostra de domicílios do setor censitário.