

2. DESCRIÇÃO DO PLANO AMOSTRAL PESB¹

A amostra da PESB, representativa para a população adulta brasileira, foi probabilística com três estágios de seleção. No primeiro estágio, 102 unidades primárias de amostragem (UPA's), ou municípios, foram selecionados probabilisticamente e proporcionalmente ao tamanho. No segundo estágio, 280 unidades secundárias de amostragem (USA's) foram selecionadas probabilisticamente e proporcionalmente em cada município. No terceiro estágio, os domicílios foram selecionados proporcionalmente ao tamanho de forma sistemática. No final, um adulto era selecionado aleatoriamente dentro de cada domicílio para responder à pesquisa.

Cada unidade primária de amostragem constitui um município. Foram utilizados os dados da Contagem de 1996 do IBGE e a divisão político-administrativa das 5 regiões, com 27 estados (26 estados mais o Distrito Federal) e 5507 municípios. Foram sorteados 102 municípios. Destes, 27 foram considerados como estrato certo (auto-representativos) e 75 como não sendo auto-representativos. Os municípios auto-representativos são as capitais dos estados. A amostra é representativa das 5 regiões (Centro-Oeste, Norte, Nordeste, Sul e Sudeste).

Para reduzir custos, todos os municípios com até 20.000 habitantes das regiões Norte e Centro-Oeste foram excluídos. Assim, o equivalente a 3,1% da população foi excluído da população amostrada.

2.1. Estratificação e distribuição da amostra

A PESB empregou dois estágios de estratificação nos quais os municípios foram explicitamente estratificados, ou agrupados, por variáveis sócio demográficas de interesse. Já os municípios dentro dos estratos explícitos foram classificados por outras variáveis de maneira implícita. A estratificação tem duas finalidades: (1) controlar a distribuição das variáveis chaves (p.ex.: região) e (2) diminuir o erro total da amostra.

Estratificação explícita

Os dois estratos utilizados na PESB foram a região e o tipo de município (capital de estado ou não). A decisão de estratificar explicitamente a amostra baseada nessas duas variáveis se deveu à importância delas para algumas explicações e análises sociais. Em primeiro lugar, estudos mostram que existe uma grande diferença econômica e cultural entre as regiões do Brasil. Em segundo lugar, embora não seja provado empiricamente, uma opinião comum para vários pesquisadores é de que existam diferenças substanciais em atitudes políticas e sociais entre os habitantes que vivem em capitais de estado em oposição àqueles que vivem fora delas.

A PESB estratifica a população de domicílios (total de domicílios = 38.487.336) primeiro pelas regiões e depois dentro destas por município considerando-se se são capitais de estado ou não. Todos municípios auto-representativos (27) foram selecionados com certeza, enquanto os demais municípios foram selecionados por alguns critérios probabilísticos.

Tabela 2.1: Distribuição de domicílios por Região e Tipo de município

Região	Não-capital	Capital
Centro-oeste	1.678.714	997.912
Norte	1.484.299	738.378

¹ O plano amostral da PESB foi realizado por Clifford Young.

Nordeste	7.352.564	2.226.247
Sudeste	12.831.515	4.998.216
Sul	5.277.372	902.119
Total	28.624.464	9.862.872

Estratificação implícita

Dentro de cada região, os municípios que não eram auto-representativos foram estratificados de forma implícita com base em três variáveis: (1) número de domicílios; (2) taxa de urbanização; e (3) taxa de alfabetização. Essas variáveis foram selecionadas por duas razões. Primeiro, são de fácil acesso. Segundo, análises estatísticas demonstraram que essas variáveis sócio-demográficas discriminam bem os municípios. Outras pesquisas mostram que variáveis sócio-econômicas no Brasil são importantes para explicar as diferenças de atitude e comportamento entre os indivíduos.

Na ordenação dos municípios (estratificação implícita), foi utilizada uma lógica que alternava em ordem ascendente e descendente (ordem de serpentina) sucessivos grupos de municípios de acordo com as variáveis de porte, urbanização e alfabetização.

Por exemplo, na classificação dos municípios de um determinado porte, ordenou-se da mais baixa urbanização até a mais alta; no segundo porte, ordenou-se os municípios da mais alta taxa de urbanização até a mais baixa, e assim sucessivamente (ver tabela 2.2 abaixo).

Tabela 2.2: Ordem Serpentina das Variáveis

Porte 1	Taxa de urbanização 1	Alfabetização 3
Porte 1	Taxa de urbanização 2	Alfabetização 2
Porte 1	Taxa de urbanização 3	Alfabetização 1
Porte 2	Taxa de urbanização 3	Alfabetização 1
Porte 2	Taxa de urbanização 2	Alfabetização 2
Porte 2	Taxa de urbanização 1	Alfabetização 3
Porte 3	Taxa de urbanização 1	Alfabetização 3
Porte 3	Taxa de urbanização 2	Alfabetização 2
Porte 3	Taxa de urbanização 3	Alfabetização 1

A ordem de serpentina dos municípios assegura: 1) que a amostra será representativa por meio de todos os tipos de variáveis, e, 2) diminuir variações amostrais ao maximizar as homogeneidades dentro dos municípios e a heterogeneidades entre os municípios.

2.2. Lógica da Estratificação

A estratificação explícita da amostra considera as cinco grandes regiões como definidas pelo IBGE: Sul, Sudeste, Nordeste, Centro-Oeste e Norte.

Tabela 2.3: Distribuição do Número de Residências, proporções e erro amostral para as 5 Regiões

Região	Nº de domicílios	Proporção	Amostra Proporcional	Erro Amostral
Norte	2.222.677	0,058	116	13

Centro-Oeste	2.676.626	0,070	140	12
Nordeste	9.578.811	0,249	500	6
Sudeste	17.829.731	0,463	931	5
Sul	6.179.491	0,161	323	8
Total	38.487.336	1,000	2.010	3

A Tabela 3 mostra a distribuição do número de domicílios pelas regiões do Brasil: Norte e Centro-Oeste representam apenas 6 e 7% da população, respectivamente.

Para aumentar o potencial analítico dos usuários para as regiões menores, a PESB empregou um desenho amostral desproporcional, aumentando a amostra das quatro menores regiões (Centro-Oeste, Norte, Nordeste e Sul) e diminuindo o número de entrevistas da maior região (Sudeste).

Na tabela 2.4 encontra-se a distribuição proporcional, desproporcional, o número de entrevistas e o erro amostral estimado por estrato.

Tabela 2.4: Distribuição de domicílios, nº de entrevistas e erro amostral por estrato

Região	Nº de domicílios	Proporção	Amostra Proporcional	Erro	Amostra Desproporcional	Erro
Centro-Oeste	1.678.714	0,044	88	15	120	13
Norte	1.484.299	0,039	78	16	100	14
Nordeste	7.352.564	0,191	384	7	420	7
Sudeste	12.831.515	0,333	670	5	540	6
Sul	5.277.372	0,137	276	8	320	8
Capitais	9.862.872	0,256	515	8	510	6
Total	38.487.336	1,000	2.010	3	2.010	3

Finalmente, os municípios (UPA's) foram estratificados como sendo capital ou não. A Tabela acima mostra que aproximadamente 26% da população reside nas capitais de estados, correspondendo ao tamanho da amostra de 510 e à margem de erro de 6 pontos percentuais.

2.3. Qualificação do plano amostral

A PESB foi desenhada para representar as cinco grandes regiões brasileiras e as capitais de estado. Entretanto, um número pequeno de entrevistas em algumas regiões corresponde a uma margem de erro maior por estrato (ver tabela 4). Recomenda-se aos usuários bastante atenção quando analisarem os dados com um enfoque regional.

Como outras pesquisas sociais existentes, a PESB tem o objetivo de ser realizada regularmente, o que facilitará obter uma série cumulativa de dados. Este procedimento aumenta efetivamente o tamanho da amostra, permitindo a análise de subgrupos pequenos. Por exemplo, utilizando este procedimento, o General Social Survey (GSS) totaliza mais de 35.000 respondentes, permitindo que os analistas examinem grupos extremamente pequenos, como os judeus americanos que representam aproximadamente 2% da população.

2.4. Seleção dos municípios, setores censitários e domicílios

Na seleção dos 75 municípios não auto-representativos, os municípios foram classificados pela quantidade de residências, pela taxa de urbanização e pela taxa de analfabetismo. Em seguida, calculou-se o intervalo da amostra e um começo aleatório para cada região e, de maneira sistemática, os municípios foram selecionados probabilisticamente e proporcionalmente. Na tabela abaixo podemos observar a distribuição das UPA's.

Região	Nº de UPA's auto-representativas	Nº de UPA's não auto-representativas	Total
Centro-Oeste	4	6	10
Norte	7	5	12
Nordeste	9	21	30
Sudeste	4	27	31
Sul	3	16	19
Total	27	75	102

Foram também selecionadas 280 unidades secundárias (setor censitário). Listou-se dentro de cada setor censitário todos os domicílios. Foram selecionados entre 4 e 10 domicílios para serem pesquisados. A variação no número de domicílios sorteados por setor censitário se deveu ao tamanho do município. Após a seleção do domicílio, um morador adulto era aleatoriamente selecionado para responder à pesquisa.

2.5. Taxa de não-resposta, amostra dividida e ponderação dos dados²

2.5.1. Taxa de não-resposta

A taxa de não-resposta é definida para suprir os casos de recusa ou de não-contato. Como a amostra é probabilística, não é desejável a substituição. Assim, foram sorteados domicílios a mais para se obter o número esperado de entrevistas. A taxa de não-resposta considerada para a PESB foi de 34%. Ou seja, considerou-se que, no mínimo, a taxa de resposta seria de 66%. Na tabela abaixo pode-se observar as taxas efetivas de respostas por estado.

REGIÃO	UF	TAXA DE RESPOSTA
Centro-Oeste	DF	78%
	GO	80%
	MS	71%
	MT	93%
Centro-Oeste Total		80%
Nordeste	AL	75%
	BA	89%
	CE	94%
	MA	91%
	PB	81%
	PE	74%
	PI	83%
	RN	86%

² A ponderação dos dados foi feita conjuntamente por Andréia Schroeder e Clifford Young.

	SE	86%
Nordeste Total		85%
Norte	AC	95%
	AM	61%
	AP	75%
	PA	85%
	RO	83%
	RR	83%
	TO	83%
Norte Total		84%
Sudeste	ES	65%
	MG	83%
	RJ	69%
	SP	61%
Sudeste Total		68%
Sul	PR	72%
	RS	80%
	SC	90%
Sul Total		79%
TOTAL GERAL		77%

Desse modo, como a taxa de não-resposta da pesquisa foi menor do que a taxa inicialmente prevista, foram obtidas 2362 entrevistas.

2.5.2. Amostra Dividida

Na PESB foi utilizado um recurso chamado “split ballot” ou amostra dividida: metade dos entrevistados respondeu à versão 1 do questionário e a outra metade respondeu à versão 2. As diferenças entre as duas versões estão especificadas mais adiante.

Em que consiste a amostra dividida?

O *split* é um tipo de desenho experimental puro que permite aos pesquisadores testar determinadas diferenças entre o uso de palavras e fraseados de perguntas aplicadas em um mesmo *survey*.

Este tipo de desenho experimental resulta em duas sub-amostras que são representativas da população alvo. Por esta razão, permite o teste de escalas diferentes, o teste de formulação de perguntas, a influência de determinadas informações antes da aplicação de determinada pergunta e, conseqüentemente, o teste de hipóteses sociológicas.

O procedimento adotado foi de alternar as versões durante a aplicação dos questionários. Ou seja, aplicava-se um questionário da versão 1 e outro da versão 2, um da versão 1 e assim por diante. No total, obteve-se 1221 casos da versão 1 e 1143 questionários da versão 2.

2.5.3. Ponderação

Como já explicado acima, a amostra da PESB foi probabilística. Neste tipo de desenho todas as etapas de seleção são realizadas por critérios que envolvem a aleatoriedade. Tudo é feito para garantir que o perfil pesquisado seja representativo da população brasileira. No entanto, algumas vezes ocorrem pequenas diferenças entre o perfil pesquisado e o perfil da população. Os dados são ponderados para corrigir estas diferenças.

No caso da PESB, além da diferença entre os perfis existem outros fatores que levaram à realização da ponderação: a desproporcionalidade entre as regiões e as diferenças entre as

versões. Foi considerado, para efeito de ponderação a distribuição das variáveis básicas (sexo, idade, escolaridade, pea³ e não-pea) por região e por versão de questionário.

O peso corrige as desproporcionalidades da amostra.

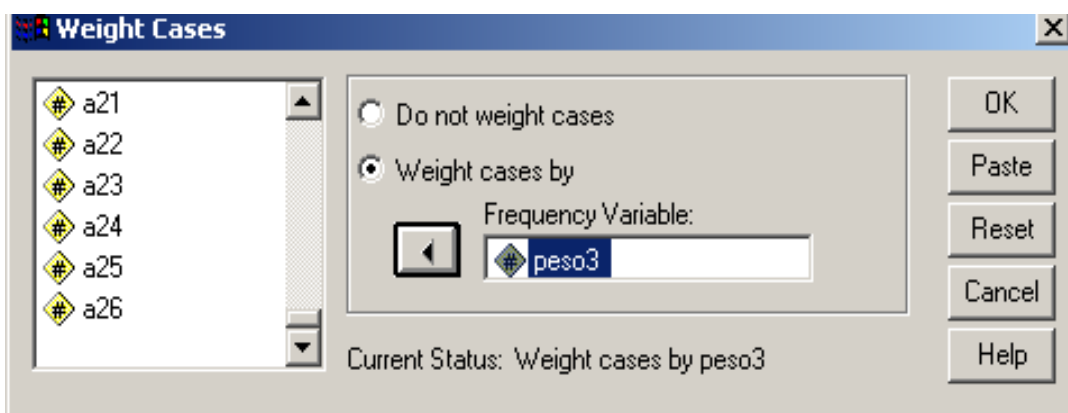
2.5.4. Uso do peso

Como os dados da PESB são ponderados, ao se fazer qualquer tipo de uso dos dados o peso deve estar ativado. Assim, ao abrir a base de dados a seguinte informação no canto inferior à direita deve estar aparecendo: **Weight On.**

Caso a variável “peso” não esteja ativada você deve seguir o seguinte procedimento:

- Menu Data / Weight Cases / Weight cases by;
- Inserir variável “peso3” no campo que fica disponível e clicar em OK (veja ilustração abaixo);
- Salvar a base de dados;

Após estes procedimentos a base volta a ficar com o peso aplicado.



³ População Economicamente Ativa - PEA