



Revista de Saúde Pública

ISSN: 0034-8910

revsp@usp.br

Universidade de São Paulo  
Brasil

de Oliveira Bussab, Wilton; Grupo de Estudos em População, Sexualidade e Aids  
Plano amostral da Pesquisa Nacional sobre Comportamento Sexual e Percepções sobre  
HIV/Aids, 2005

Revista de Saúde Pública, vol. 42, núm. 1, junio, 2008, pp. 12-20

Universidade de São Paulo  
São Paulo, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=67240172004>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica

Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal

Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

Wilton de Oliveira Bussab

Grupo de Estudos em  
População, Sexualidade e Aids\*

# Plano amostral da Pesquisa Nacional sobre Comportamento Sexual e Percepções sobre HIV/Aids, 2005

## Sampling plan for the National Survey Sexual Behavior and Perceptions of the Brazilian Population concerning HIV/AIDS, 2005

---

### RESUMO

**OBJETIVO:** Descrever o plano amostral e os métodos de estimação utilizados na coleta e análise dos dados da Pesquisa sobre o Comportamento Sexual e Percepções sobre HIV/Aids da População Brasileira em 2005.

**MÉTODOS:** São apresentadas as decisões adotadas quanto à definição do universo da pesquisa, estratos de interesse da pesquisa e do plano amostral, principais procedimentos para análise dos dados e desempenho da amostra no campo.

**RESULTADOS DA AMOSTRAGEM:** Foi elaborado plano probabilístico, com 5.040 unidades amostrais, obtidas sobre a população brasileira: indivíduos com idades entre 16 e 65 anos, residentes nos grandes centros urbanos brasileiros. Trata-se de plano amostral complexo, distribuído em oito domínios principais de estimação, desenhado em múltiplos estágios, com um homem ou mulher entrevistada no último desses estágios. Cada unidade entrevistada e cada domicílio têm probabilidade específica de pertencer à amostra.

**DESCRITORES:** Amostragem. Técnicas de Estimação. Interpretação Estatística de Dados. Coleta de Dados. Questionários. Estudos Populacionais em Saúde Pública. Brasil. Estudos transversais.

Departamento de Informática e Métodos Quantitativos. Escola de Administração de Empresas de São Paulo. Fundação Getúlio Vargas. São Paulo, SP, Brasil

\* Integrantes: Elza Berquó, Francisco Inácio Pinkusfeld Bastos, Ivan França Junior, Regina Barbosa, Sandra Garcia, Vera Paiva, Wilton Bussab.

### Correspondência | Correspondence:

Wilton de Oliveira Bussab  
DIMQ-EAESP-FGV  
Av. Nove de Julho 2029  
01313-902 São Paulo, SP, Brasil  
E-mail: wilton.bussab@fgv.br

Recebido: 31/7/2007  
Revisado: 25/2/2008  
Aprovado: 20/3/2008

---

## ABSTRACT

**OBJECTIVE:** To describe the sampling plan and estimation methods used to collect and analyze data in the survey Sexual Behavior and Perceptions of the Brazilian Population concerning HIV/AIDS in 2005.

**METHODS:** The study presents the decisions that were made concerning population definition, strata of interest to the survey and to the sampling plan, main procedures for data analysis and sample performance in the field.

**SAMPLING RESULTS:** A probabilistic plan was designed with 5,040 sampling units obtained from the Brazilian population, with individuals aged between 16 and 65 years living in large Brazilian urban centers. It is a complex sampling plan distributed over eight main estimation domains, designed in multiple stages. A man or a woman was interviewed in the last stage. Each interviewed unit and each household have specific probability of belonging to the sample.

**DESCRIPTORS:** Sampling studies. Estimation Techniques. Data Interpretation, Statistical. Data Collection. Questionnaires. Population Studies in Public Health. Brazil. Cross-sectional studies.

---

## INTRODUÇÃO

Pesquisas de abrangência nacional na área dos comportamentos sexuais, riscos e proteção frente ao HIV/Aids e demais infecções sexualmente transmissíveis (IST) são necessárias em quaisquer sociedades que desejem formular e avaliar políticas públicas neste campo, com base em dados empíricos consistentes.<sup>1</sup> Da mesma forma, é importante investigar algumas das interfaces das IST com o consumo de substâncias psicoativas ou o campo mais abrangente da saúde sexual e reprodutiva, como decisões acerca da utilização de métodos contraceptivos.

Com raras exceções, as pesquisas realizadas no Brasil e na maioria dos países de renda baixa e média têm abrangência local ou, no máximo, regional. Contudo, tais pesquisas nacionais são imprescindíveis, especialmente em um país de dimensões continentais, atravessado por heterogeneidades e contrastes sociais, econômicos e culturais, como o Brasil. Um retrato fidedigno desta sociedade e seus fenômenos dinâmicos – como as mudanças no âmbito dos comportamentos sexuais ou a epidemia de HIV/Aids – exigem a realização sistemática de inquéritos nacionais, lastreados em planos amostrais consistentes.

O objetivo do presente artigo foi descrever aspectos metodológicos sobre a definição do universo da pesquisa e o plano amostral elaborado para pesquisa de âmbito nacional.

## MÉTODOS

### Definição do universo da pesquisa

A “Pesquisa Sobre Comportamento Sexual e Percepções da População Brasileira Sobre HIV/Aids” de 2005 compreende a totalidade das unidades da Federação, enquanto que sua edição realizada em 1998, abrangeu 24 estados brasileiros e o Distrito Federal, excetuando-se os estados de Tocantins e Roraima.<sup>a</sup>

Os dados do levantamento de 2005 foram obtidos por meio de amostra probabilística em múltiplos estágios, totalizando 5.040 respondentes, com idades entre 16 e 65 anos, moradores das grandes regiões urbanas do Brasil. Portanto, a amostra pesquisada no levantamento de 2005 foi maior que a da pesquisa realizada em 1998, que foi de 3.600 indivíduos.

O Ministério da Saúde fixou quatro estratos geográficos de interesse, correspondendo a agrupamentos de estados brasileiros, a saber: estados das regiões Norte e Nordeste; estados da região Centro-Oeste e Sudeste exceto São Paulo; estados da Região Sul e, por fim, o Estado de São Paulo, como domínio adicional.

### População-alvo

O sistema de referência adotado para definir a população de interesse da pesquisa foi o Censo Demográfico

---

<sup>a</sup> Berquó E, coordenador. In: Comportamento sexual da população brasileira e percepções do HIV/AIDS. Brasília (DF): Ministério da Saúde, Secretaria de Políticas de Saúde, Coordenação Nacional DST e Aids; 2000. (Série avaliação, 4).

do ano 2000, realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE).

A pesquisa utilizou como unidade de análise uma das unidades de agregação de dados utilizadas pelo IBGE, o sistema de referência informações, agrupado em 558 microrregiões. Além das informações identificadoras das microrregiões (nome e unidade da federação), foram utilizadas as seguintes variáveis: população total, população urbana e população com idade entre 16 e 64 anos, esta como uma aproximação da população de interesse, cujo limite superior é de 65 anos.

As microrregiões recobrem grandes áreas territoriais, o que poderia dificultar o acesso a algumas unidades e, conseqüentemente, aumentar o custo do levantamento. Para minimizar esse problema, foram tomadas decisões operacionais. A primeira delas foi restringir a pesquisa aos moradores das áreas urbanas das microrregiões

A segunda decisão foi pesquisar apenas os grandes conglomerados urbanos, definidos como microrregiões que, em 2000, contavam com mais do que 100.000 habitantes moradores em suas áreas urbanas. Esta medida reduziu para 276 o número de microrregiões, correspondendo à diminuição de 12% no número de moradores de interesse da pesquisa, conforme exibido na Tabela 1. Finalmente, assim como na pesquisa de 1998, foram eliminadas 17 microrregiões da Região Norte que não incluíam as respectivas capitais dos estados, em virtude da dificuldade de acesso. Ao final, 259 microrregiões constituíram o universo de interesse da pesquisa de 2005.

Desse modo, a população-alvo da pesquisa foi definida pela inclusão de todos os moradores, com idade entre 16 e 65 anos, residentes em áreas urbanas das microrregiões, que em 2000, possuíam mais de 100.000 habitantes em sua zona urbana, excetuando-se microrregiões da Região Norte que não continham as capitais dos respectivos estados.

Esse contingente representava, em 2000, 88% da população brasileira da faixa etária residente em áreas urbanas do País, o que correspondia a aproximadamente 80 milhões de pessoas. Na Tabela 1, podem ser observados os efeitos das medidas operacionais sobre os totais populacionais, bem como sobre os estratos de interesse. Como esperado, a maior perda ocorreu no estrato Norte/Nordeste, com 76% de cobertura, enquanto que em São Paulo esta taxa foi de 98%.

### Plano Amostral

O plano amostral escolhido foi do tipo estratificado em múltiplos estágios, com probabilidades desiguais de inclusão dos eventos sob análise.<sup>2,3</sup>

Foram adotados os quatro estratos geográficos de interesse do Ministério da Saúde: estados das regiões Norte/

**Tabela 1.** Contingentes populacionais e percentagem da população abrangida pela pesquisa. Brasil, 2005.

Estrato amostral	População*	Cobertura (%)
Brasil		
População total**	168.370.893	
População urbana	137.015.685	
População urbana de 16 a 64 anos***	90.645.118	
População-alvo	79.546.966	87,7
Norte/Nordeste		
População total**	60.144.982	
População urbana	41.722.290	
População urbana de 16 a 64 anos	26.348.629	
População-alvo	20.027.545	76,6
Centro-Oeste/Sudeste		
População total**	46.609.259	
População urbana	40.769.874	
População urbana de 16 a 64 anos	27.339.181	
População-alvo	24.678.821	90,3
Estado de São Paulo		
População total**	36.713.804	
População urbana	34.332.375	
População urbana de 16 a 64 anos	23.415.976	
População-alvo	22.849.375	97,6
Sul		
População total**	24.902.848	
População urbana	20.191.146	
População urbana de 16 a 64 anos	13.541.332	
População-alvo	11.991.225	88,6

\* Fonte: Censo Demográfico 2000. Fundação IBGE.

\*\* População residente em domicílios particulares permanentes. Elaboração dos autores.

\*\*\* Devido às informações agregadas do Censo, usou-se como medida do tamanho da população objetivo o número de habitantes entre 16 e 64 anos.

Nordeste; das regiões Centro-Oeste/Sudeste exceto São Paulo; da Região Sul; e do Estado de São Paulo.

Além de produzir estimativas confiáveis para esses quatro estratos, procurou-se produzir boas estimativas acerca do comportamento da população das capitais brasileiras. Para esse fim, no âmbito de cada estrato, separaram-se as microrregiões que continham as capitais das demais microrregiões, distribuindo-se então as 259 microrregiões pelos oito domínios, conforme apresentado na Tabela 2.

Entretanto, no desenho do plano amostral decidiu-se que todas as microrregiões contendo capitais deveriam

**Tabela 2.** Regiões geográficas segundo distribuição das microrregiões por domínios e população de 16-64 anos. Brasil, 2005.

Região geográfica	Microrregião que inclui Capital					
	Não		Sim		Total	
	Microrregião	População*	Microrregião	População*	Microrregião	População*
Norte/Nordeste	62	8.052	16	11.975	78	20.027
Centro-Oeste/Sudeste	72	10.372	7	14.307	79	24.678
Estado de São Paulo	49	14.472	1	8.377	50	22.849
Sul	49	7.608	3	4.383	52	11.991
Total	232	40.505	27	39.041	259	79.546

\* em 1.000 hab

estar representadas na amostra. Isto equivale, do ponto de vista de amostragem, a considerar cada uma destas microrregião como um estrato. Assim, a população da pesquisa foi dividida em oito domínios de interesse para a análise e 31 estratos para o desenho do plano amostral: quatro contendo as microrregiões do interior por região geográfica e 27 para cada microrregião contendo a capital do respectivo estado.

O desenho do plano amostral atendeu a algumas restrições e premissas:

- em função dos recursos orçamentários disponíveis e dos objetivos, o tamanho viável da amostra foi estabelecido em 5.040 indivíduos;
- visando a obter a mesma precisão em cada estrato, optou-se por fixar o tamanho de amostra em 1.260 domicílios por estrato, alternativa mais adequada para subpopulações com a mesma variabilidade;
- a precisão das estimativas foi calculada na suposição de um desenho amostral aleatório simples, que acarretaria em estimativas de proporções com erro amostral da ordem de 3% em cada estrato geográfico, e poder de detectar diferenças significativas entre estratos da ordem de 4,0 pontos percentuais;<sup>1</sup>
- assumindo-se que a similaridade entre as respostas de moradores de um mesmo setor censitário (correlação intraclasse<sup>1-3</sup>) aumenta o erro amostral,<sup>6</sup> e que investigações sobre características sociais e econômicas da população têm sugerido que o número ótimo por setor censitário não ultrapasse 15 domicílios, fixou-se como nove domicílios o número de referência para sorteio dentro de cada setor. Isso correspondeu ao sorteio de 140 setores por estrato geográfico;
- cada microrregião contendo a capital do estado constituiu um substrato especial dentro do respectivo domínio de interesse. Desse modo, passou-se a contar com 27 microrregiões pré-selecionadas.

O plano amostral escolhido foi do tipo estratificado em quatro estágios para as os estratos que não incluem capitais, reduzindo-se a três naqueles com as mi-

corregiões das capitais, em virtude da eliminação do sorteio da microrregião. As unidades amostrais em cada estágio foram assim definidas:

- unidade primária de amostragem (UPA) – utilizou-se a microrregião;
- unidade secundária de amostragem (USA) – referente ao sorteio no segundo estágio, corresponde ao setor censitário urbano. Foram utilizados para tal propósito os setores censitários definidos pelo IBGE para o Censo Demográfico de 2000;
- unidade terciária de amostragem (UTA) – corresponde ao domicílio particular;
- unidade quaternária de amostragem (UQA) – indivíduo com idade entre 16 e 65 anos.

Em suma, a amostra inicial de 5.040 unidades foi igualmente dividida em 1.260 unidades por estrato geográfico, visando a obter a mesma precisão estatística para as estimativas em cada um dos quatro estratos regionais. O sorteio de nove domicílios em cada setor censitário, como controle dos efeitos de conglomeração, implicou distribuir 560 setores censitários pelos quatro grandes estratos geográficos, alocando-se 140 setores por estrato. Em consonância com as razões apresentadas e pretendendo-se obter estimadores com igual precisão nos oito domínios de interesse da pesquisa, recomendar-se-ia utilizar 70 setores censitários em cada um deles. Porém, a particularidade relativa à composição do número de capitais existentes nos diferentes domínios geográficos determinou algumas adaptações.

Como o estrato São Paulo-Capital tem apenas uma microrregião e o Sul-capital tem três, decidiu-se alocar a cada um desses estratos 49 setores censitários. Essa medida, combinada com o sorteio de nove domicílios em todos os setores, garantiria 441 domicílios no âmbito desses estratos, número adequado para produzir estimativas com grau de confiabilidade aceitável. A diferença de 91 setores deste quantitativo para 140 foi atribuída ao estrato que não contém a microrregião da capital. Para cada microrregião da amostra, no estágio subsequente, foram sorteados sete setores censitários.

Para a região Norte/Nordeste Capital, com 16 microrregiões e população pouco maior, decidiu-se por atribuir 77 setores censitários, destinando-se os 63 restantes para as microrregiões que não incluem as capitais. A combinação dessas decisões remete à Tabela 3, que descreve a alocação final do número de microrregiões, setores censitários e domicílios alocados pela amostra.

Para o sorteio da amostra em cada estrato das microrregiões que não incluíam a capital, tanto as microrregiões como os setores censitários, foram sorteados com probabilidades proporcionais aos respectivos tamanhos. O número de habitantes para microrregiões e o número de domicílios ocupados para setores censitários segundo os dados do Censo de 2000 foram adotados como medidas de seus respectivos tamanhos. Os domicílios e moradores foram sorteados com igual probabilidade.

Para o domicílio selecionado, uma pessoa com idade entre 16 e 65 anos foi sorteada por meio de uma tabela de

sorteio previamente atribuída ao domicílio.<sup>4</sup> Esta tabela permitiu um balanceamento no total da amostra entre homens e mulheres, controlado pela faixa etária.

O sorteio dos setores dentro da microrregião se deu do seguinte modo:

- para cada microrregião sorteada, foram alocados sete setores censitários;
- os municípios da microrregião sorteada foram ordenados pelo tamanho, criando assim, sete estratos implícitos (zonas) de setores;
- a cada zona correspondeu o sorteio de um setor censitário;
- o setor assim sorteado, com probabilidade proporcional ao seu tamanho com referência ao ano 2000, foi recontado, de modo a obter o número de domicílios atualizados, ou seja, o seu tamanho real. Essa contagem foi realizada antes do sorteio dos domicílios do setor.

Desse modo, garantiu-se uma amostra com representação proporcional do tamanho das cidades e, consequentemente, um maior espalhamento das unidades sorteadas, denominadas “unidades titulares”.

Para cada domicílio, foi sorteado outro, dentro do mesmo setor censitário e próximo, que serviu de substituto em caso de perda da unidade titular.

Para cada um dos quatro grandes estratos geográficos, o número de setor censitário em cada capital foi repartido proporcionalmente à sua população, garantindo, porém, que cada capital tivesse pelo menos dois setores. Após essa alocação, o sorteio para cada uma das 27 microrregiões das capitais seguiu o mesmo procedimento descrito para as microrregiões que não incluíam capitais. Assim, os municípios de cada microrregião contendo a capital foram ordenados pelo tamanho, os setores censitários divididos em zonas e sorteado um de cada zona. O processo para selecionar domicílio e morador foi o mesmo utilizado para os estratos do interior, inclusive o sorteio de unidades de substituição.

## RESULTADOS DA AMOSTRAGEM

Para o trabalho de coleta de dados no campo foi contratada uma empresa de pesquisa de mercado para realizar as 5.040 entrevistas. O desenvolvimento do questionário, treinamento dos entrevistadores e coordenadores de campo foi todo executado e supervisionado pela coordenação da pesquisa.

No trabalho de campo houve a necessidade de substituir 13 setores censitários (2,3% do total): seis na região Norte/Nordeste, cinco na região Sul e dois em São Paulo, basicamente por apresentarem características rurais ou por impossibilidade de acesso, esgotados

**Tabela 3.** Alocação da amostra. Brasil, 2005.

Estrato amostral	Número de			Domicílio por setor censitário
	Domicílio	Microrregião	Setor Censitário	
Total				
Norte/Nordeste	1.260	25	140	9
Centro-Oeste/Sudeste	1.260	17	140	9
Estado de São Paulo	1.260	14	140	9
Sul	1.260	16	140	9
Total	5.040	72	560	-
Interior				
Norte/Nordeste	567	9	63	9
Centro-Oeste/Sudeste	630	10	70	9
Estado de São Paulo	819	13	91	9
Sul	819	13	91	9
Total	2.835	45	315	-
Capital				
Norte/Nordeste	693	16	77	9
Centro-Oeste/Sudeste	630	7	70	9
Estado de São Paulo	441	1	49	9
Sul	441	3	49	9
Total	2.205	27	245	-

Fonte: Censo Demográfico 2000, Fundação IBGE

todos recursos possíveis (condomínios). Estes foram então substituídos por setores próximos e com características semelhantes, de renda média.

As Tabelas 4 e 5 apresentam os números absolutos e relativos que permitem avaliar o comportamento da amostra.

Cerca de 70% da amostra foi obtida com os domicílios titulares. Os principais motivos que determinaram a utilização das unidades reservas se referiram a problemas com os domicílios (e.g. fechados, de veraneio, sem função de moradia). A proporção de recusa situou-se próxima a 10%, proporção que pode ser considerada baixa para pesquisas desta natureza. Além disso, 5% dos domicílios foram descartados por não possuírem moradores com idades estabelecidas pelos critérios de inclusão da pesquisa.

a apuração de medidas simples descritivas (médias, proporções, razões, índices, combinações lineares de variáveis, tabelas cruzadas e outras). Para obtenção de estimativas com adequada precisão, os procedimentos recomendados na apuração dessas estatísticas foram:

- uso de sistema de ponderação para produzir estimativas;
- emprego intensivo de estimadores razões;
- estimativas de totais populacionais de modo indireto, e
- uso de pacotes específicos do SPSS para estimar os erros amostrais.

**Tabela 4.** Características do número e proporção de entrevistas realizadas. Brasil, 2005.

Região	Prevista	Realizada Titular	Domicílio inelegível	Realizada substituta	Fora da faixa etária	Total de recusa
Norte/Nordeste	1260	886	167	374	41	64
Centro-Oeste/Sudeste	1260	732	200	528	60	121
Estado de São Paulo	1260	675	151	585	41	107
Sul	1260	674	193	586	81	118
Total	5040	2967	711	2073	223	410

Avaliando-se o desempenho da amostra pelas regiões, observou-se que a amostra coletada de titulares esteve mais próxima do planejado para o estrato Norte/Nordeste, enquanto que para as outras três regiões os desempenhos foram muito parecidos entre si, com substituição mais freqüente de titulares. A análise desses números sugere a ausência de vícios relevantes decorrentes do trabalho de coleta dos dados, podendo-se considerar essas substituições como tendo ocorrido aleatoriamente.

A análise dos dados obtidos pelo plano amostral complexo exige o uso de técnicas desenvolvidas especificamente para esse fim, principalmente de estatísticas com dados ponderados para estimação não viesada, ou quase não viesada, dos parâmetros populacionais.<sup>5,6</sup> O uso de estimadores desenhados para amostras aleatórias simples pode introduzir vícios relevantes nessas estimativas<sup>6</sup> e as inadequações seriam ainda mais acentuadas se empregadas para o cálculo dos erros amostrais. A maioria dos programas computacionais disponíveis no mercado possui opções para o cálculo de estatísticas ponderadas. O mesmo não ocorre para o cálculo das variâncias amostrais de estimadores, o que exige a presença de módulos especiais desenvolvidos com essa finalidade.<sup>6</sup>

Os dados da pesquisa foram utilizados na análise das várias dimensões do comportamento sexual do brasileiro, por diferentes pesquisadores e empregando distintas técnicas de análise. Os estudos iniciam-se com

**Tabela 5.** Alguns indicadores do desempenho das entrevistas em relação ao total de domicílios elegíveis. Brasil, 2005.

Região	Proporção de		
	Realizada no titular	Recusa	Fora da faixa etária
Norte/Nordeste	81,1%	5,9%	3,8%
Centro-Oeste/Sudeste	69,1%	11,4%	5,7%
Estado de São Paulo	60,9%	9,6%	3,7%
Sul	63,2%	11,1%	7,6%
Total	68,5%	9,5%	5,2%

Cada entrevistado teve associado um peso  $w_i$  que é o inverso da sua probabilidade de inclusão. Embora tenha sido sorteada uma pessoa por domicílio, a probabilidade de inclusão do domicílio não é a mesma do entrevistado. Desse modo, construiu-se outro conjunto de pesos para estimar, quando necessário, características referentes ao domicílio. Detalhes do cálculo dos pesos encontram-se no Anexo, que explicita os procedimentos estatísticos. Habitualmente, a soma desses pesos reflete o total do universo descrito no sistema de referência. Essa soma correspondeu a 5.040, que é o tamanho efetivo da amostra coletada. Passou-se, portanto, de um sistema de ponderação para outro por meio da multiplicação de uma constante. Resumindo, tem-se associado a cada indivíduo um peso  $w_i$ , tal que  $\sum w_i = 5040$ ;

Uma estatística-chave no cálculo de várias características para uma variável qualquer  $Y$  é o seu total ponderado na amostra, dado pela expressão  $T_y = \sum w_i y_i$ , onde  $y_i$  é o valor observado da característica  $Y$  para o  $i$ -ésimo indivíduo da amostra. Por exemplo, um estimador não viesado da média populacional é a média ponderada na amostra  $\sum w_i y_i / \sum w_i$ . O emprego de qualquer um dos dois sistemas de ponderação levaria à mesma estimativa, já o que fator de multiplicação apareceria simultaneamente no numerador e denominador. A grande maioria das estatísticas empregadas nas análises em tela pode ser expressa como sendo do tipo razão  $r = T_y / T_x$ , onde  $T$  são estimadores dos totais das características de interesse  $X$  e  $Y$ , portanto, seriam também estimadores populacionais adequados.

A estimação de proporções é uma aplicação direta do estimador razão, descrito anteriormente. Exemplificando: para estimar a proporção das mulheres no universo da pesquisa que, tendo entre 30 e 60 anos, tiveram mais de três filhos, basta encontrar esse valor na amostra. Ou seja, o quociente  $p_x = T_y / T_x$  onde  $T_x$  = total de mulheres e  $T_y$  = total de mulheres entre 30 e 60 anos e mais de três filhos, ambas estimadas com as recomendações indicadas.

Para encontrar estimativas de totais para essa população baseou-se em dados de projeções populacionais do IBGE. Ainda no exemplo anterior, para estimar o total das mulheres entre 30 e 60 anos de idade com mais de três filhos, seria necessário obter, das projeções do IBGE, o número  $N_x$  projetado de mulheres existentes no universo de referência, e o total estimado será então  $N_x \cdot p_x$ . Dependendo das informações disponíveis sobre as populações populacionais, poder-se-ia também usar métodos mais refinados, levando em conta as estimativas produzidas para os domínios e estratos.

A determinação dos erros amostrais dos estimadores em planos amostrais complexos exige pacotes especiais para o seu cálculo, nesta pesquisa optou-se pelas sub-rotinas específicas do SPSS.

## COMENTÁRIOS

No presente artigo foram brevemente descritos alguns procedimentos metodológicos necessários à estimação de parâmetros populacionais e seus respectivos erros amostrais para estudos sobre diversos aspectos da Pesquisa sobre Comportamento Sexual e Percepções Sobre HIV/Aids da População Brasileira, em 2005.

Para realizar as entrevistas de campo nas 5.040 unidades alocadas pelos diversos domínios da pesquisa, foi necessário recorrer a aproximadamente 30% de unidades substitutas, principalmente, devido a recusas e inexistência de moradores com as características desejadas no domicílio. O comportamento dessas substituições

foi muito semelhante e não implicou vícios relevantes. A taxa de realização de entrevistas nos domicílios titulares da região Norte/Nordeste foi a mais elevada, enquanto que nas outras três regiões foram mais baixas e similares entre si.

Muitos dos trabalhos descritos no presente suplemento comparam os resultados com os observados em 2005 com a pesquisa de 1998. Os resultados do primeiro levantamento sugeriram aos pesquisadores o aprimoramento de seus instrumentos de coleta dos dados, este foi feito de modo a garantir a comparabilidade dos resultados. Quanto ao plano amostral, as principais mudanças introduzidas em 2005 foram: a redefinição dos estratos, visando outras configurações regionais, e a ampliação do número de microrregiões elegíveis para a população-alvo.

Apesar do aumento do número de microrregiões, o mesmo não ocorreu com a população-alvo, moradores com idade entre 16 e 64 anos de idade. Os estratos de 1998 teriam, em 2000, cerca de 72 milhões de pessoas, enquanto que para o planejamento de 2005 este contingente, também em 2000, seria próximo de 79 milhões, correspondendo a um aumento de 7 milhões de pessoas (9,8%). Ambas as pesquisas teriam, portanto, um universo comum de 90% da população de interesse com os dados do censo de 2000.

Na pesquisa de 1998, a população foi dividida em três estratos: Norte-Nordeste (NONO), Centro-Oeste Expandido (COEX) e Sul Expandido (SULX), enquanto que a de 2005 está distribuída em quatro estratos.

Com o propósito de comparar as diferenças entre os dois universos analisados, foram utilizados os dados do censo 2000, isto é, estimam-se como seriam os

**Tabela 6.** Distribuição do número de microrregiões e respectivas populações de 16 a 65 anos. Brasil, 1998 e 2005.

Estrato de 2005	Estrato de 1998				Total Linha	Total estrato 2005
	Fora	NONO	COEX	SULX		
Fora		12			12	0
NONO	16	62			78	78
COSU	37		36	6	79	79
SPAU	18			32	50	50
RSUL	17			35	52	52
Total Coluna	88	74	36	73	271	259
Total Estrato 98	0	74	36	73	183	171

Fora, indicando que não pertencia ao estrato daquele ano, NONO=Norte e Nordeste; COEX=Centro Oeste Expandido; SULX=Sul Expandido; COSU=Centro-Oeste e Sudeste; SPAU=São Paulo e RSUL=Sul.

Fonte: Dados do Censo de 2000



**Tabela 7.** Distribuição da população com idade entre 16 e 64 anos pelos estratos de 1998 e 2005.

Estrato de 2005	Estrato de 1998*				Total Linha*	Total estrato 2005*
Fora		990			990	0
NONO	1.358	18.670			20.027	20.027
COSU	3.376		12.827	8.475	24.679	24.679
SPAU	1.877			20.972	22.849	22.849
RSUL	1.498			10.493	11.991	11.991
Total Coluna	8.109	19.660	12.827	39.941	80.537	79.547
Total Estrato 1998	0	19.660	12.827	39.941	72.428	71.438

\* em 1.000 hab

números em 2000 com as condições impostas no desenho de 1998 e o de 2005. Os resultados constam das Tabelas 6 e 7.

## REFERÊNCIAS

1. Berquó E, Barbosa RM, Grupo de Estudos em População, Sexualidade e Aids. [Introdução]. *Rev Saude Publica*. 2008;42(Supl 1):7-11.
2. Bolfarine H, Bussab WO. Elementos de amostragem. São Paulo: Edgard Blücher; 2005.
3. Kish L. Survey Sampling. New York: John Wiley & Sons; 1965.
4. Marques RM, Berquó ES. Seleção da unidade de informação em estudos de tipo "survey": um método para a construção das tabelas de sorteio. *Rev Bras Estat*. 1976;37(145):81-92.
5. Pessoa DGC, Silva PLN. Análise de dados amostrais complexos. São Paulo: Associação Brasileira de Estatística; 1998.
6. United Nations. Department of Economic and Social Affairs. Statistics Division. Household sample surveys in developing and transition countries. New York; 2005. [Series F, 96]. Disponível em: [http://unstats.un.org/unsd/HHSurveys/pdf/Household\\_surveys.pdf](http://unstats.un.org/unsd/HHSurveys/pdf/Household_surveys.pdf)

Artigo baseado em dados da pesquisa "Comportamento sexual e percepções da população brasileira sobre HIV/Aids", realizada pelo Centro Brasileiro de Análise e Planejamento (CEBRAP), com o apoio do Programa Nacional de DST e Aids do Ministério da Saúde (Processo n.º ED 213427/2004).

Este artigo seguiu o mesmo processo de revisão por pares de qualquer outro manuscrito submetido a este periódico, sendo garantido o anonimato entre autores e revisores. Editores e revisores declaram não haver conflito de interesses que pudessem afetar o processo de julgamento dos artigos. Os autores declaram não haver conflito de interesses.

Observando as Tabelas, constata-se que, em 1998, o universo era constituído por 183 microrregiões. Este aumento se deveu à incorporação de 88 novas microrregiões e à exclusão de 12 antigas. Este aumento ocorre principalmente em decorrência da incorporação de novas microrregiões dos estados mais populosos do Brasil: São Paulo, Minas Gerais, Rio de Janeiro, Rio Grande do Sul e Pernambuco.

Embora as comparações entre as duas pesquisas sejam feitas a partir de populações distintas, diferenças significativas observadas com relação às características estudadas devem ser antes imputadas a mudanças efetivas de comportamento, do que a possíveis diferenças entre as referidas populações. Alegar uma eventual diferença entre resultados das duas pesquisas ao fato de serem diferentes populações só seria possível se o comportamento destes novos 10% da população se apresentasse, em sua grande maioria, no sentido e direção contrária a dos resultados da pesquisa de 1998. Dada a composição destas novas microrregiões, é bastante improvável que isto venha a ocorrer.

## ANEXO

### Probabilidade de Inclusão

#### a. Nomenclatura

(i). Variáveis

P = população entre 16 e 64 anos (IBGE)

D = domicílios particulares no Censo (IBGE)

D' = domicílios particulares contados. (obtidos no campo)

a = número de setores censitários alocados no estratos (plano amostral)

M = número de moradores no domicílio (contados no campo)

c = número efetivo de casas entrevistadas (trabalho de campo)

b = número de microrregiões sorteadas no interior (plano amostral)

(ii) Índices

h = estrato

j = microrregião (MR)

k = setor censitário (SC)

l = domicílio

i = indivíduo

(iii) Convenções utilizadas

$M_{hkl}$  = número de moradores existentes no domicílio l, do SC k, na MR j do estrato h.

A supressão de um índice indica a soma de valores daquela variável para aquele indicador, assim,  $M_{hjk} = \sum M_{hkl}$  indica o total de domicílios no SC k, da MR j, do estrato h.

As mesmas convenções valem para os demais índices e variáveis.

#### b. Estimativas referentes às Capitais

Nas capitais tem-se apenas uma MR, assim não entra o índice j.

Probabilidade esperada

$$P(I_{hkli} \in A) = \frac{a_h \cdot D_{hk}}{D_h} \cdot \frac{9}{D_{hk}} \cdot \frac{1}{M_{hkl}} = \frac{9a_h}{M_{hkl} \cdot D_h}$$

Probabilidade efetiva

$$P(I_{hkli} \in A) = \frac{9a_h}{M_{hkl} D_h} \cdot \frac{D_{hk}}{D_h} \cdot \frac{c_{hk}}{9} = \frac{a_h c_{hk}}{M_{hkl} D_h} \cdot \frac{D_{hk}}{D_h}$$

#### c. Estimativas referências ao Interior

Probabilidade Teórica

$$P(I_{hkli} \in A) = \frac{b_h P_{hj}}{P_h} \cdot \frac{7 \cdot D_{hjk}}{D_{hj}} \cdot \frac{9}{D_{hjk}} \cdot \frac{1}{M_{hkl}} = \frac{63b_h \cdot P_{hj}}{P_h D_{hj} M_{hkl}}$$

Probabilidade efetiva

$$P(I_{hkli} \in A) = \frac{63b_h \cdot P_{hj}}{P_h D_{hj} M_{hkl}} \cdot \frac{D_{hjk}}{D_{hjk}} \cdot \frac{c_{hk}}{9}$$