

PESQUISA SOCIOECONÔMICA EM TERRITÓRIOS DE VULNERABILIDADE SOCIAL NO DISTRITO FEDERAL

Relatório Metodológico II

Definição Espacial da Pesquisa Socioeconômica em Territórios de Vulnerabilidade Social no Distrito Federal

Contrato de Prestação de Serviços Nº. 050/2009

MAIO DE 2010

**DIEESE**
DEPARTAMENTO INTERSINDICAL DE
ESTATÍSTICA E ESTUDOS SOCIOECONÔMICOS

**DESCRIÇÃO DOS PROCESSOS DE SELEÇÃO E DE ESTIMAÇÃO DA
PESQUISA SOCIOECONÔMICA EM TERRITÓRIOS DE VULNERABILIDADE
SOCIAL NO DISTRITO FEDERAL A PARTIR DE PESQUISA DOMICILIAR**

SUMÁRIO

INTRODUÇÃO	04
1. DESCRIÇÃO DO PROCESSO DE SELEÇÃO DA AMOSTRA DE DOMICÍLIOS	05
1.1 DELIMITAÇÃO E QUANTIFICAÇÃO DO UNIVERSO ALVO DE DOMICÍLIOS DA PESQUISA	05
2. SELEÇÃO DA AMOSTRA DE SETORES DO UNIVERSO 1	08
2.1 FORMAÇÃO DE SETORES DENTRO DAS RAs	08
2.2. SELEÇÃO DE SETORES DENTRO DAS RAs	09
2.2.1 SELEÇÃO DE DOMICÍLIOS A SEREM ENTREVISTADOS DENTRO DA RA	10
3. SELEÇÃO DE UNIDADES DOMICILIARES DENTRO DO UNIVERSO 2	12
4. SELEÇÃO DO PROCESSO DE ESTIMAÇÃO	12
ANEXOS	14

INTRODUÇÃO

A Secretaria de Desenvolvimento Social e Transferência de Renda (SEDEST) contratou em 2009 os serviços técnicos do Departamento Intersindical de Estatística e Estudos Sócio-Econômicos (DIEESE) com o propósito de desenvolver e executar Pesquisa Socioeconômica domiciliar detalhada destinada a identificar os Territórios de Vulnerabilidade Social existentes no Distrito Federal e, em cada um deles, quantificar e caracterizar os perfís demográfico, educacional, de ocupação e de rendimentos de suas respectivas populações. As informações levantadas devem contribuir à melhor compreensão da relação existente entre a vulnerabilidade social e o mercado de trabalho, possibilitando a elaboração de propostas que propiciem a inclusão social através da ampliação das oportunidades de trabalho e da melhoria dos rendimentos.

Nesse estudo, o conceito de Vulnerabilidade Social adota a definição proposta no Programa Nacional de Assistência de 2004 (PNAS/2004), onde a população vulnerável é delimitada, sempre no contexto familiar, mediante 9 combinações excludentes de atributos relativos, principalmente, à renda *per capita*, tamanho, tipo, chefia e composição da família. Trata-se, na prática, de um domínio de estudo que só pode ser conhecido e dimensionado através de consulta domiciliar ampla, única forma de se calcular, inclusive, as Taxas de Vulnerabilidade Social (TVS) vigentes em cada um dos Territórios de interesse da Pesquisa.

No âmbito geográfico, dois tipos de territórios de interesse da SEDEST foram considerados: por uma parte, as Regiões Administrativas (RAs), divisões de natureza político-administrativa, reconhecidas no DF; por outra, as Áreas de Vulnerabilidade do Distrito Federal, com extensões geralmente mais limitadas, onde a Assistência Social do DF desenvolve ações sociais específicas de ajuda aos grupos populacionais mais pobres. Ainda que de antemão tenha-se optado por deixar de fora do estudo algumas RAs (Plano Piloto, Lago Sul e Norte, entre outras), em virtude dos elevados e generalizados níveis socioeconômicos que apresentam, o detalhamento geográfico e social dos resultados alvejados no estudo acabou configurando um levantamento de grande envergadura. Dessa forma, visando garantir a boa confiabilidade das estimativas para cada território, a amostra selecionada foi dimensionada em 20.000 domicílios, os quais devem ser entrevistados, em 2010, no prazo de 4 meses.

No presente documento descreve-se o desenho de amostragem probabilístico a ser utilizado na Pesquisa, detalhando os critérios e fontes de dados adotados para definir o Universo do estudo e os procedimentos empregados na seleção da amostra de domicílios correspondente aos dois tipos de territórios acima mencionados. Em seguida, detalha-se o processo de estimação através do qual devem ser preparados os resultados expandidos da Pesquisa (totais, médias, percentagens e razões) para cada território de estimação.

1. . DESCRIÇÃO DO PROCESSO DE SELEÇÃO DA AMOSTRA DE DOMICÍLIOS

1.1 DELIMITAÇÃO E QUANTIFICAÇÃO DO UNIVERSO-ALVO DE DOMICÍLIOS DA PESQUISA

Conforme comentado na Proposta “Pesquisa Socioeconômica em Territórios de Vulnerabilidade Social no Distrito Federal”, preparada pelo DIEESE, não existe na atualidade um registro ou levantamento completo dos Territórios de Vulnerabilidade Social existentes no DF, situação que dificulta a definição e conseqüente dimensionamento do Universo-Alvo da Pesquisa. Por esse motivo, o mesmo documento propõe que na 1ª etapa da Pesquisa sejam consultadas várias fontes no intuito de se construir esse Universo de territórios, com a correspondente localização de cada um deles e os respectivos totais de domicílios e famílias. Entre essas fontes, o citado documento sugere consultar o Cadastro Único de Programas Sociais relativo ao Distrito Federal e o Cadastro de Usuários da CEB que consomem até 80 kW/ mês. Embora o georreferenciamento desses registros possa fornecer nuvens ou manchas de domicílios que apresentam carências com intensidades espaciais diferenciadas, é difícil delimitar com exatidão a extensão geográfica desses territórios menores, até porque espacialmente não existe um critério quantitativo previamente definido que faça a distinção entre territórios vulneráveis e não vulneráveis. Assim, por exemplo, o exame do mapeamento das unidades domiciliares do Cadastro Único correspondente às RAs Ceilândia, Samambaia, Recanto das Emas e muitas outras mostra manchas homogeneamente espalhadas nos correspondentes limites territoriais, tornando difícil qualquer tentativa de reconhecer territórios menores onde a vulnerabilidade possa ser mais acentuada. Por outra parte, qualquer esforço complementar mediante visitas prévias ao próprio levantamento de campo tornaria essa identificação territorial muito demorada e custosa.

Em virtude dessas dificuldades, foi proposto que o próprio levantamento de campo da Pesquisa seria o mecanismo através do qual esses territórios menores poderiam ser mais bem localizados. Para tanto, o Universo geográfico da Pesquisa passou a ser definido da seguinte forma:

1. **Universo 1** - Domicílios pertencentes às RAs do DF que apresentam percentagens elevadas de incidência de beneficiários do Programa Bolsa-Família. Nessa delimitação territorial mais ampla, as informações do georreferenciamento fornecidas pelo DIEESE foram importantes para desconsiderar de imediato algumas RAs. Dessa forma, foram descartados do Universo-Alvo da Pesquisa todos os domicílios localizados no Plano Piloto, Lago Sul e Lago Norte, Cruzeiro, Sudoeste e Octogonal, Park Way, Jardim Botânico e Águas Claras. Portanto, o Universo de domicílios das RAs ficou configurado da seguinte forma:

RA I: Brasília, incluindo apenas o Campamento da Telebrasília, Vila Planalto e Granja do Torto e o Varjão (1)

RA II: Gama

RA III: Taguatinga

RA IV: Brazlândia

RA V: Sobradinho

RA VI: Planaltina

RA VII: Paranoá

RA VIII: Núcleo Bandeirante

RA IX: Ceilândia

RA X: Guará

RA XII: Samambaia

RA XIII: Santa Maria

RA XIV: São Sebastião

RA XV: Recanto das Emas

RA XVII: Riacho Fundo

RA XIX: Candangolândia

RA XXI: Riacho Fundo II

RA XXV: SCIA- Estrutural

RA XXVI: Sobradinho II

RA XXIX: Itapoã

RA XXX: Vicente Pires

(¹) Muito embora o Varjão seja atualmente uma RA (XXIII), para efeitos da seleção da amostra ela foi adscrita à RA I (Brasília) em virtude do limitado número de domicílios nela existentes.

A contagem rápida de estruturas domiciliares efetuada através do Satélite do GOOGLE para as RAs acima relacionadas é da ordem de 500 mil (ver detalhamento no Anexo I). Adverte-se que, nessa contagem, são geralmente omitidas as unidades domiciliares localizadas em edifícios residenciais, motivo pelo qual a dimensão do Universo da Pesquisa pode ter sido reduzida. Entretanto, essa omissão, que afeta principalmente as RAs de maior tamanho (especialmente Taguatinga), acaba deixando de fora da amostra muitas unidades domiciliares que, se fossem entrevistadas, seriam inquestionavelmente enquadradas como sendo não vulneráveis.

2. **Universo 2** - Domicílios pertencentes às Áreas de Vulnerabilidade Social relacionadas pela SEDEST que não tenham sido consideradas nos territórios das respectivas RAs. Essas exclusões podem ocorrer devido à desatualização ou por falta de abrangência dos mapas obtidos através do GOOGLE (referidos a out. 2008) ou da cartografia disponibilizada pela SEDEST (correspondente a 2007-2008). Acredita-se que o universo dessas Áreas Vulneráveis seja no máximo da ordem de 75, sendo que a maior parte delas reúne poucos domicílios. A consideração em separado dessas áreas menores permite garantir que em cada uma delas será selecionada uma amostra independente de domicílios que, mesmo sendo de pequeno porte, pode bem representar o respectivo miniuniverso de domicílios vulneráveis. A localização exata e a listagem domiciliar de cada uma destas áreas, algumas das quais apresentam grande dispersão espacial por corresponderem a zonas rurais, será feita simultaneamente com a listagem de setores selecionados nas respectivas RAs, desconhecendo-se na atualidade o total de domicílios nelas existentes.

Cabe lembrar que a relação de Áreas de Vulnerabilidade preparada pela SEDEST (ver anexo II) inclui muitos outros territórios de interesse da Pesquisa que foram previamente considerados no Universo I e, portanto, setorializados dentro das respectivas RAs. Trata-

se, em geral, de áreas que congregam muitos domicílios e que certamente terão um ou mais setores selecionados na amostra desse Universo 1. Acredita-se, portanto, que o tamanho da amostra de domicílios a serem pesquisados nestas Áreas Vulneráveis será bem maior que o pesquisado no Universo 2. Esse é, por exemplo, o caso das RAs Brazlândia, Samambaia, Itapoã e Gama, onde a totalidade (ou quase) dos setores ou quadras que as integram foram considerados Áreas de Vulnerabilidade pela SEDEST.

Em suma, a variante metodológica proposta propugna que, da mesma forma que nas Áreas de Vulnerabilidade menores identificadas pela SEDEST no Universo 2, ou nas de maior tamanho identificadas de antemão pela mesma Secretaria nas RAs no Universo 1, em cada uma das quais haverá setores selecionados onde serão feitas as entrevistas domiciliares para representar esses territórios previamente identificados, também os restantes setores selecionados nas RAs, geralmente mais numerosos, devem ter capacidade de reconhecer outras áreas vulneráveis que a priori não foram listados pela SEDEST. Afinal, esses outros setores também estão localizados em bairros, quadras, setores, zonas ou pontos cardeais de cada RA que podem servir de referência para calcular taxas de vulnerabilidade e, assim, definir Outras Áreas de Vulnerabilidade Social que não foram antecipadamente reconhecidas pela SEDEST. Nessas condições, o próprio levantamento de campo da Pesquisa seria o mecanismo através do qual esses outros territórios vulneráveis menores seriam identificados.

A alternativa proposta, além de satisfazer o objetivo principal do estudo, qual seja o de identificar, dimensionar e caracterizar os territórios de vulnerabilidade social no DF, abre a possibilidade de que esse quadro de carências sociais seja examinado em um contexto domiciliar mais amplo, referido à totalidade dos domicílios das RAs selecionadas e do DF e também no âmbito da vulnerabilidade domiciliar total, incluindo os domicílios vulneráveis localizados fora das Áreas de Vulnerabilidade Social da SEDEST e dos territórios posteriormente identificados nas RAs através da Pesquisa.

2. SELEÇÃO DA AMOSTRA DE SETORES NO UNIVERSO 1.

Em virtude da inexistência de uma listagem atualizada de domicílios que abranjam a totalidade do território de cada uma das 21 RAs escolhidas, a seleção da amostra de domicílios

correspondente ao Universo 1 deve ser efetivada em duas etapas: na primeira, propõe-se selecionar aleatoriamente alguns dos setores em que previamente foi dividida a totalidade do território da RA; na segunda, devem ser selecionados, também de forma aleatória, alguns domicílios constantes na listagem atualizada de unidades domiciliares existentes nos setores previamente selecionados na etapa anterior. A seguir, detalham-se os procedimentos de seleção da amostra a serem empregados em ambas as etapas.

2.1 FORMAÇÃO DE SETORES DENTRO DAS RAS

O território de cada uma das RAs acima relacionadas deve ser subdividido em setores menores seguindo os seguintes critérios:

- a. O setor deve reunir um total de 100 a 200 domicílios contíguos integrantes de um ou mais quarteirões ou áreas adjacentes;
- b. O total de domicílios de cada setor deve ser estimado através da contagem aproximada de estruturas domiciliares obtidas através da utilização da versão Satélite mais recente do GOOGLE. Por dificuldades de leitura dessas cartografias, os edifícios residenciais são geralmente contados como uma única estrutura domiciliar;
- c. O setor deve ter limites claros que possam ser facilmente identificados em campo por parte dos listadores, supervisores e entrevistadores, evitando, sempre que possível, o uso de linhas imaginárias para sua delimitação;
- d. Áreas vazias ou institucionais, localizadas em quadras ou quarteirões, onde exista suspeita da existência de domicílios particulares devem ser incluídas na setorialização formando parte de um único setor de natureza residencial;
- e. Áreas dedicadas exclusivamente a outras finalidades não-residenciais (centros de saúde, administração, centros deportivos, centros industriais ou comerciais, parques etc.) devem ser consideradas na setorialização, atribuindo-lhes um total de domicílios igual a 0;
- f. Áreas Vulneráveis que fazem parte da relação preparada pela SUDEST, mesmo que sejam perfeitamente localizadas pelo GOOGLE, devem ser consideradas no Universo 2 sempre que exista a suspeita de que, em virtude de seu limitado número de domicílios, elas possam ficar excluídas da Pesquisa.

2.2 SELEÇÃO DE SETORES DENTRO DAS RAS

A seleção de setores dentro de uma RA deve ser realizada de forma aleatória com probabilidade proporcional ao número estimado de domicílios de cada setor. Para tanto, deve ser adotado o seguinte procedimento:

- i. Cada um dos setores formados dentro de uma RA deve ser identificado com um único número de ordem, em sequência unitária desde 001 a n;
- ii. Em uma planilha EXCEL, onde consta o número de ordem (1ª coluna) e a respectiva estimativa de domicílios ou Medida de Tamanho-MDT (2ª coluna), deve ser criada uma coluna adicional (3ª coluna) onde se registre o Total de Domicílios Acumulado dos setores:

Nº de ordem	Nº de domic.(MDT)	Total de Dom. Acum.
001	120	120
002	110	230
....
n	150	\sum MDT

- iii. Embora, inicialmente, tenha se pensado em recomendar a seleção de um número fixo de 50 setores (\underline{a}) em cada RA, esse plano foi logo modificado em função da grande diferença do número de domicílios existente entre as RAs. Após algumas simulações, cujo propósito principal foi o de evitar a dupla seleção de domicílios em qualquer setor, optou-se pela seguinte alternativa (ver Anexo I):

RAs de menor tamanho: $\underline{a} = 30$ a 40 setores

RAs de tamanho intermediário: $\underline{a} = 50$ setores

RAs de grande tamanho: $\underline{a} = 58$ a 60 setores

- iv. Uma vez definido o total de setores a serem selecionados (\underline{a}) numa particular RA α , deve ser calculado o intervalo de seleção $F\alpha = \sum MDT\alpha / \underline{a}$ e, a seguir, seleciona-se um número aleatório de início (X), com valor entre 1 e $F\alpha$;
- v. A sequência de números aleatórios selecionados nessa RA α deve ser calculada de forma sistemática por agregação sucessiva do intervalo $F\alpha$, ou seja, X; X+ $F\alpha$; X+2* $F\alpha$ + X+3* $F\alpha$;

- vi. A identificação dos \underline{a} setores selecionados será feita comparando-se o total de domicílios acumulado com cada um dos números aleatórios calculados em “v”. A identificação de um setor ocorre quando o total de domicílios acumulado supera imediatamente o correspondente número aleatório selecionado;
- vii. A numeração dos setores selecionados (chamados de Unidades Primárias de Amostragem- UPAs) deve ser efetivada através de um código de quatro dígitos. Os dois primeiros dígitos identificam a RA, enquanto que os dois seguintes se referem ao número da seleção da UPA dentro da RA. Assim, o número 0905, por exemplo, corresponde à 5ª UPA selecionada na RA 09, ou seja, Ceilândia.

O procedimento acima descrito configura uma seleção de UPAs com probabilidade proporcional ao tamanho (PPT). Nesse modalidade de seleção, a probabilidade de seleção de qualquer UPA i na RA α é igual a:

$$P(UPA_{\alpha i}) = MDT_{\alpha i} / (F_{\alpha}),$$

onde $F_{\alpha} = \sum MDT_{\alpha} / \underline{a}$ corresponde ao intervalo de seleção de UPAs na RA α .

Assim, se em uma RA α o total estimado de domicílios é $\sum MDT_{\alpha} = 15.000$ e o total de setores (UPAs) a serem selecionados é $\underline{a} = 50$, o intervalo de seleção $F_{\alpha} = 15.000/50=300$. Dessa forma, se o tamanho de um setor selecionado α ($UPA_{\alpha i}$) foi estimado em $MDT_{\alpha i} = 100$ domicílios, então a probabilidade de seleção dessa UPA αi será igual a:

$$P(UPA_{\alpha i}) = 100/300 = 1/3$$

Visando viabilizar esse processo de seleção de UPAs dentro das RAs é preciso que cada RA de interesse reúna um certo número mínimo de domicílios (de 4.500 a 5.000), já que de outra forma o intervalo de seleção F_{α} resultante pode ser muito pequeno, o que poderia ocasionar a dupla seleção de domicílios dentro da UPA, situação que se quer evitar. Esse seria o caso da RA Varjão e das outras três Áreas de Vulnerabilidade Social listadas pela SEDEST na RA Brasília, ou seja, Granja do Torto, Campamento da Telebrásília e Vila Planalto. Por esse motivo, sugere-se que a RA Brasília seja definida reunindo esses quatro territórios, medida que garante a seleção de varias UPAs em cada uma dessas áreas.

A modalidade acima descrita deve resultar na seleção de quase 1.000 UPAs na totalidade das 21 RAs de interesse da Pesquisa (Ver Anexo 1).

2.2.1 SELEÇÃO DE DOMICÍLIOS A SEREM ENTREVISTADOS DENTRO DA RA

A segunda etapa do modelo probabilístico proposto consiste na seleção aleatória de domicílios dentro das RAs. Nesse sentido, duas alternativas podem ser privilegiadas: a primeira, relativa à seleção de um número fixo \underline{b} de entrevistas em cada UPA previamente listada. Já que o total de entrevistas a serem realizadas em ambos universos é da ordem de $\underline{n} = 20.000$ domicílios espalhadas em aproximadamente $\underline{a} = 1.000$ UPAs selecionadas nas 21 RAs e 75 ADVs, esse número fixo \underline{b} de entrevistas por UPA deve ser da ordem de 19 ou 20 domicílios ($\underline{n} = \underline{a} * \underline{b}$); a segunda, referida ao uso de uma fração de seleção domiciliar variável de UPA a UPA na RA α ($f_b = \underline{b} / MDT\alpha$) que garanta a igualdade de probabilidade final f (e, conseqüentemente, o mesmo peso) para cada domicílio dentro da RA $\{ f = (MDT\alpha / \sum MDT\alpha / \underline{a}) * (\underline{b} / MDT\alpha) = \underline{a} * \underline{b} / \sum MDT\alpha \}$. Nesse caso, o número de domicílios entrevistados será variável de UPA a UPA, sendo que a média de entrevistas por UPA (\underline{b}) deve ser da ordem de 19 a 20 domicílios. Muito embora o primeiro método não garanta o mesmo peso para todo domicílio selecionado na RA, situação que deve ser levada em consideração no processo de estimação da Pesquisa, essa primeira modalidade de seleção parece mais recomendável pois facilita a operacionalização da pesquisa em campo, bem como o controle sobre o total dos domicílios entrevistados, evitando que ele seja muito superior ou inferior ao previsto no orçamento da Pesquisa ($\underline{n} = 20.000$ domicílios).

A seguir, descreve-se o procedimento de seleção a ser utilizado na seleção de domicílios (Unidades Secundárias de Amostragem-USAs) dentro das UPAs previamente selecionadas na 1ª etapa da amostragem:

- i. Cada UPA selecionada, e numerada com um código de 4 posições, deve ser listada em campo por um agente listador seguindo o mesmo procedimento de listagem utilizado na PED/DF. Alternativamente, quando se trate de UPAs cuja área já se encontra consolidada e construção fechada, pode-se fazer uso dos registros mais atualizados de endereços fornecidos pela CEB. Em qualquer uma dessas duas alternativas, cada um dos domicílios deve ocupar uma única linha da listagem e levar um número sequencial que vai de 1 a $B\alpha i$ (último domicílio listado na UPA i da RA α);

- ii. Haja vista que em cada UPA será selecionado um número fixo de domicílios (\underline{b} =19 ou 20) e que o total de domicílios listados (B_{ai}) varia de UPA a UPA, em cada uma delas deverá ser calculado um Intervalo de seleção de domicílios ($F_{bi} = B_{ai} / \underline{b}$) e um número aleatório inicial entre 1 e F_{bi} que identifica o primeiro domicílio X_b selecionado na UPA. A sequência sistemática de números aleatórios que identificam os domicílios listados nessa UPA será: X_b ; X_b+F_b ; X_b+2*F_b etc. Dessa forma, em cada UPA será garantida a seleção de domicílios (19 ou 20) prevista no modelo de amostragem ora descrito.
- iii. A probabilidade de seleção de domicílios na UPA i (2ª etapa) será igual a : $P (USA_{\beta}) = \underline{b}/B_i$ e, dessa forma, a probabilidade final de seleção de um domicílio β na RA_{α} será : $\{ P USA_{\alpha\beta} = P (UPA_{\alpha}) * P (USA_{\beta}) = \{ (MDT_{\alpha} / (F_{\alpha}) * (\underline{b} / B_i) \}$, configurando assim um esquema de desigual probabilidade de seleção de domicílios entre UPAs da mesma RA. Essa desigualdade que deve ser considerada na preparação dos pesos de cada unidade domiciliar;
- iv. Cada domicílio selecionado deverá ter um código único de identificação de 6 dígitos, onde os 4 primeiros correspondem à UPA e os dois últimos ao número da seleção domiciliar. Assim, o código 090515 corresponde ao décimo quinto (15º) domicílio selecionado na UPA 05 da RA 09.

3. SELEÇÃO DE UNIDADES DOMICILIARES DENTRO DO UNIVERSO 2

O processo de seleção de domicílios nas Áreas de Vulnerabilidade Social pertencentes ao Universo 2 pode e deve ser bem mais simples que o acima descrito. Em virtude de se tratar de áreas que congregam menos domicílios, essas áreas podem ser listadas sem recorrer à setorialização e à seleção prévia de setores ou UPAs, bastando apenas definir claramente os limites geográficos delas. Nesse caso, a seleção de domicílios deve ser feita em uma única etapa seguindo o mesmo esquema sistemático descrito em 2.2.1, recomendando-se ainda que também neste caso seja selecionado um total de 20 domicílios em cada Área de Vulnerabilidade. Calcula-se, então, que nesse Universo 2, onde aparecem relacionadas um total de 75 áreas, devam ser entrevistados um máximo de 1.500 domicílios.

Embora a seleção aleatória de domicílios em uma única etapa tenha a vantagem de reduzir os erros de amostragem quando comparada com uma amostra do mesmo tamanho que foi

selecionada através de um modelo de duas etapas, nessas áreas menores os erros de amostragem das estimativas e indicadores de interesse da Pesquisa podem ser mais expressivos que os correspondentes ao Universo das RAs ou subdivisões menores das mesmas, uma vez que no Universo 2 o tamanho da amostra será equivalente a apenas uma única UPA (20 domicílios).

A numeração das Áreas de Vulnerabilidade incluídas no Universo 2 também deve obedecer a um código de 4 dígitos, onde os dois primeiros serão definidos como 00 (para indicar de que se trata do Universo 2) e os dois seguintes corresponderão a cada uma das Áreas de Vulnerabilidade, numeradas em sequência unitária que vai de 01 a 75. Por sua vez, os domicílios nelas entrevistados deverão ser identificados agregando-se mais dois dígitos, que correspondem ao número de seleção do domicílio. Assim, o número 000511 identifica o décimo primeiro (11º) domicílio selecionado na listagem da Área 05 do Universo 2.

Para o Universo 2, a probabilidade de seleção de um domicílio na Área de Vulnerabilidade Social (AVS) α é $P(AVS\alpha) = \underline{h} / B\alpha$, onde $\underline{h} = 20$ e $B\alpha =$ total de domicílios listados na $AVS\alpha$.

Dessa forma, qualquer domicílio selecionado na amostra dos Universos 1 ou 2 terá uma probabilidade final conhecida, $P(USA \alpha)$ ou $P(AVS\alpha)$.

4. DESCRIÇÃO DO PROCESSO DE ESTIMAÇÃO

A preparação das estimativas de totais, médias ou percentagens relativas aos Universos 1 e 2 deve fazer uso dos fatores de expansão ou pesos (F') calculados para cada domicílio entrevistado, valores esses que devem constar na base de dados da Pesquisa. Esses fatores de expansão F' correspondem ao inverso da probabilidade final de seleção de cada domicílio - $P(USA \alpha\beta)$ ou $P(AVS\alpha)$ -, previamente ajustados pelas percentagens de perdas de domicílios ocupados não entrevistados (recusas, ausência de moradores no domicílio ou incompletas) que tenham ocorrido na respectiva UPA ou AVS. Assim, se a probabilidade final de seleção de um domicílio na UPA i da $RA\alpha$ é igual a $P(USA \alpha\beta) = 1/4 * 20/140 = 1/28$, então, $F = 28$. Porém, se 2 dos 20 domicílios selecionados nessa UPA não foram entrevistados devido a recusa (10%), o fator de expansão deve ser ajustado para $F' = F/0,9 = 28/0,9 \approx 31$.

Levando-se em consideração que o processo de entrevistas ocasionará tamanhos efetivos da amostra diferentes dos previstos para as RAs, UPAs e outros sub-conjuntos de interesse no estudo, os estimadores de totais, médias e percentagens mais apropriados assumem a forma de razões do tipo $r = y / x$, onde o x representa uma variável de contagem com valores 0 ou 1. Dessa forma, os estimadores expandidos propostos assumem as seguintes fórmulas:

3.1 Médias (percentagens):

- UPA β (ou AVT) :

$$r_{\beta} = y / x, \text{ onde: } y = \sum F'_{\approx\beta} * y_{\beta j} \text{ e } x = \sum F'_{\approx\beta} * x_{\beta j}, \text{ sendo}$$

$y_{\beta j}$ = valor da variável y para o domicílio j da UPA β ; $x_{\beta j}$ = valor unitário para o mesmo domicílio e $F'_{\approx\beta}$ = ponderador ou peso final ajustado para cada domicílio entrevistado na UPA β

- RA α (ou conjunto de UPAs):

$$r'_{\alpha} = y' / x', \text{ onde: } y' = \sum \sum F'_{\approx\beta} * y_{\beta j} \text{ e } x' = \sum \sum F'_{\approx\beta} * x_{\beta j},$$

3.2 Totais Expandidos

- UPA β (ou AVT) :

$$T_{\beta} = \sum F'_{\approx\beta} * y_{\beta j}$$

- RA α (ou conjunto de UPAs):

$$T'_{\alpha} = \sum \sum F'_{\approx\beta} * y_{\beta j}$$

ANEXOS

ANEXO I

RESUMO DE SELEÇÕES DE 1a E 2a ETAPA						
	Estimativa	No seleções	Tamanho	Tamanho		
R.Administrativas	Universo 1	1a.etapa	amostra	amostra	Tamanho	Tamanho
		(a)	domic.	com média	com média	domic.
			(media 20)	19 nas RAs	maiores	(media 19)
Brasília	6.580	30	600	600	600	570
Ceilândia	101.100	60	1.200	1.140	1.140	1.140
Samambaia	53.990	60	1.200	1.140	1.140	1.140
Santa Maria	33.620	50	1.000	1.000	1.000	950
Gama	26.570	50	1.000	1.000	1.000	950
Guará	21.960	50	1.000	1.000	1.000	950
São Sebastião	17.680	50	1.000	1.000	1.000	950
SCIA-Estrutural	5.440	31	620	620	620	589
V.Pires	14.640	50	1.000	1.000	1.000	950
Itapoã	15.990	50	1.000	1.000	1.000	950
Recanto das Emas	35.920	60	1.200	1.140	1.140	1.140
Brazlândia	10.720	50	1.000	1.000	1.000	950
N.Bandeirante	4.200	30	600	600	600	570
Candangolândia	4.590	30	600	600	600	570
Paranoá	9.990	50	1.000	1.000	1.000	950
Planaltina	36.520	54	1.080	1.080	1.080	1.026
Riacho Fundo I	8.050	40	800	800	800	760
Riacho Fundo II	7.080	30	600	600	600	570
Sobradinho II	19.980	50	1.000	1.000	1.000	950
Sobradinho I	11.860	50	1.000	1.000	1.000	950
Taguatinga	56.270	60	1.200	1.140	1.140	1.140
Total	502.750	985	19.700	19.460	19.460	18.715

ANEXO II

**OUTRAS ÁREAS
VULNERÁVEIS**

(Devem ser perfeitamente localizadas e listadas para seleção independente sempre que não apareçam incluídas nos mapas)

Samambaia

Expansão

Ensino Fundamental I

Invasão da Boca da Mata

Santa Maria

Nenhuma

Riacho Fundo

Colônia Agrícola

Núcleo Rural Vargem da Benção

Brasília

Invasões dos Setores de Clubes

Córrego do Palha - Zona Rural

Candangolândia

Nenhuma

Recanto das Emas

Águas Quentes

Complementar SCIA - Estrutural

Nenhuma

Paranoá

Núcleo Rural Jardim II

PAD/DF

Três Conquistas

Sobradinho dos Melos

Susurano

Lamarão Jardim II

Cariru

Boquerão

Altiplano Leste

Gama

Condomínio Residencial Mansões Paraíso

Ponte da Terra

Ponte Alta de Cima

Brazlândia

Incra

Incra 06, 07, 08 e 09

Cascalheira

Rodeador

Assentamento do Betinho

Alexandre Guzmão

Capãozinho

Chapadinha

Assentamento dos Sem Terra

Engenho

Vendinha

Barreiro

Queimadeo

Itapoã

Nenhuma

Planaltina

Vila Pacheco

Rio Preto

Tabatinga

Vila Pombal

Estância 03

Sobradinho II

Fercal

Vila Rabelo

Vila Besavi

Acampamento Sem Terra

Núcleo Rural Arrozal

Sobradinho

Lago Oeste

Alto da Boa Vista

DNOCS

N.Bandeirante

Vargem Bonita

Matadouro

Vila Cauhy

Vicente Pires

Nenhuma

Taguatinga

QNL-Chaparral

Areal

Invasão São José

Invasão XXVI de Setembro

Invasão da M Norte

Ceilândia

Setor de Chácaras

Sol Nascente

Pantanal

Vila Carroceiros

Condomínio Por do Sol

Guará

Invasão do Grêmio

Invasão do Batalhão da Polícia

São Sebastião

Morro do Preá

Condomínio Itaipu

Nova Betânia

Morro Azul

Residencial do Bosque

Jardim Botânico

Área de Córregos Mato Grande

Santo Antônio da Papuda

Bela Vista

Residencial Oeste