TEXTOS PARA DISCUSSÃO, ISSN 0103-6661

REVISÃO DA PNAD

A QUESTÃO AMOSTRAL

MÓDULO II DO ANTEPROJETO

NÚMERO 54

SETEMBRO DE 1991

Lose Colleges Atmetes do Bais

Presidente da República Fernando Collor de Mello

Ministro da Economia, Fazenda e Planejamento Marcílio Marques Moreira

FUNDAÇÃO INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE

Presidente
Eduardo Augusto Guimaráes

Diretor-Geral

José Guilherme Almeida dos Reis

ÓRGÃOS TÉCNICOS SETORIAIS

Diretoria de Pesquisas Lenildo Fernandes Silva

Diretoria de Geociências Mauro Pereira de Mello

Diretoria de Informática

Nuno Duarte da Costa Bittencourt

Centro de Documentação e Disseminação de Informações Nelson de Castro Senra



Fundação Instituto Brasileiro de Geografia e Estatística - IBGE Diretoria de Pesquisas - DPE

### REVISÃO DA PNAD

THE REPORT OF THE PARTY OF THE PROPERTY OF THE PARTY.

# A QUESTÃO AMOSTRAL

## MODULO II DO ANTEPROJETO

ANTONIO CARLOS FERNANDES DE MENEZES
Estatístico, Gerente do Projeto de Reformulação
e Modernização, do DEREN
ROSANGELA ANTUNES PEREIRA ALMEIDA
Estatístico, Chefe de Equipe de Estudos e Análises
Estatísticas do DEREN
ZÉLIA MAGALHÃES BIANCHINI
Estatístico, Analista Especializado do NME

RIO DE JANEIRO

e Analises Estativilles Pedro al al result de l'estativille de l'estativil

#### FUNDAÇÃO INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE

Av. Franklin Roosevelt, 166 - Centro CEP 20 021 - Rio de Janeiro, RJ - Brasil

# DIRETOR DE PESQUISAS LENILDO FERNANDES SILVA DIRETORA-ADJUNTA DE PESQUISAS MÁRCIA BANDEIRA DE MELLO LEITE

CHEFE DO DEPARTAMENTO DE EMPREGO E RENDIMENTO

MARIA MARTHA MALARD MAYER

CHEFE DA DIVISÃO DE ESTUDOS E ANÁLISES

PAULO CESAR DE SOUSA QUINTSLR

CHEFE DE PESQUISA MENSAL

MARILENI SILVA MANSOLDO

CHEFE DE PESQUISA ANUAL

TEREZINHA BATISTA TAVARES COUTINHO

@ 1990,91

#### TEXTOS PARA DISCUSSÃO

Série publicada pela Diretoria de Pesquisas do IBGE, com o objetivo de divulgar ensaios, estudos e outros trabalhos técnicos nas áreas econômica, social e demográfica, elaborados no âmbito da Diretoria.

Edição: Núcleo de Documentação e Disseminação da Diretoria de Pesquisas. (NDI/DPE)

Menezes, Antonio Carlos Fernandes de

Revisão da PNAD: a questão amostral: módulo II do anteprojeto / Antonio Carlos Fernandes de Menezes, Rosângeia Antunes Pereira Almeida, Zélia Magalhães Bianchini. — Rio de Janeiro: IBGE, Diretoria de Pesquisas, 1991.

33 f. — (Textos para Discussão / IBGE, Diretoria de Pesquisas, ISSN 0103-6661; n.54)

#### ISBN 85-240-0429-0

Pesquisa Nacional por Amostra de Domicilios. I. Almeida,
 Rosângela Antunes Pereira. II. Bianchini, Zélia Magalhães. III.
 IBGE. Diretoria de Pesquisas. IV. Título. V. Título: Pesquisa Nacional por Amostra de Domicilios: a questão amostral.
 VI. Série.

IBGE.CDDI. Dep. de Documentação e Biblioteca RI-IBGE/92-02

CDU 314.6(81)

Informações: Biblioteca Setorial da Diretoria de Pesquisas Rua Visconde de Niterói, 1.246, Bloco B, sala 1211-B, Mangueira.

Telefone: (021) 284-8938 ou 284-3322 - ramal 303

Dando sequencia à elaboração do Anteprojeto da Pesquisa Nacional por Amostra de Domicilios - PNAD, apresentamos o Módulo II - A Questão Amostral, que é precedido pelo Módulo I - A Questão Conceitual, publicado na série Textos para Discussão nº 46.

Este módulo anuncia as questões a serem estudadas com o objetivo de definir-se o desenho amostral para a PNAD da década de 90.

O texto, ora apresentado, foi elaborado pelos técnicos Antonio Carlos Fernandes de Menezes (Gerente do Projeto de Reformulação e Modernização), Rosangela Antunes Pereira Almeida (Chefe da Equipe de Estudos e Análises Estatísticas) e Zélia Magalhães Bianchini (técnica do Núcleo de Metodologia - NME - da DPE).

Nas discussões que precederam esta versão final, foram muito importantes as contribuições dos técnicos Marilia Biangolino Chaves e Maria Deolinda Borges Cabral (da Equipe de Estudos e Análises Estatísticas), Pedro Luiz do Nascimento Silva (chefe do Núcleo de Metodologia da DPE) e Sonia Albieri (do Núcleo de Metodologia da DPE).

Com esta divulgação, espera-se a leitura atenta e as criticas e sugestões de todos, especialistas ou aqueles interessados no tema.

Maria Martha Mallard Mayer

Chefe do Departamento de Emprego e Rendimento



#### SUMARIO

1. INTRODUÇÃO	7
2. HISTORICO	. 8
3. DESENHO AMOSTRAL E OBJETIVOS DA PESQUISA	. 9
3.1. A SITUAÇÃO DO DOMICILIO	13
4. VARIAVEIS RELEVANTES PARA O DIMENSIONAMENTO DA AMOSTRA	
5. ESTAGIOS DE SELEÇÃO	16
6. ESTRATIFICAÇÃO	18
7. METODO DE SELEÇÃO	
8. AUTOPONDERAÇÃO	22
9. A QUESTÃO CADASTRAL	24
10. O CONTROLE DA AMOSTRA	27
11. ESTIMAÇÃO	29
12. CONSIDERAÇÕES FINAIS	31

#### 1. INTRODUÇÃO

Este documento tem como objetivo apresentar para debate os tópicos que foram alvo de discussão nas reuniões preliminares sobre A Questão Amostral da Pesquisa Nacional por Amostra de Domicilios - PNAD para a década de 90.

Estas reuniões, realizadas entre técnicos do DEREN e do NME, visavam fazer uma primeira análise crítica do desenho amostral atual da pesquisa, com a finalidade de promover-se uma reformulação, de tal maneira que o novo desenho seja adequado aos objetivos da PNAD revisada, descritos no módulo I - A Questão Conceitual.<sup>1</sup>

A definição de um desenho amostral para a PNAD revisada irá depender de análises futuras, aliando aspectos conceituais / metodológicos a restrições operacionais / orçamentárias.

---- do TBUIL askintido but ne caseness

¹ Ver Rinaldi et al. (1990).

#### 2. HISTORICO

A Pesquisa Nacional por Amostra de Domicilios - PNAD foi implantada no país, gradativamente, a partir de 1967. O plano amostral da pesquisa foi inspirado no Plano Atlantida, do U.S. Bureau of the Census. Trata-se de um modelo de desenho autoponderado, em estágios múltiplos, popularizado na década de 60 para uso em pesquisas domiciliares. Tal plano vem sendo empregado, com variações, em pesquisas do gênero e se encontra largamente descrito em Hansen et alii (1953).

Na década de 70<sup>2</sup>, foram definidas no ambito geográfico da pesquisa, para efeito de divulgação, sete grandes regiões sócio-economicas independentes, que constituíram uma primeira estratificação geográfica. A amostra foi delineada com um desenho estratificado com quatro estágios de seleção, mantendo-se a fração global de amostragem constante a nível de cada região, obtendo-se, ao final, uma amostra probabilistica autoponderada. A partir de 1976, um dos estágios de subamostragem, o subsetor, foi suprimido, passando-se a ter um desenho em três estágios.

Foram estabelecidos, por região, a fração amostral e o tamanho absoluto da amostra, visando obter estimativas do total de pessoas na força de trabalho com coeficientes de variação em torno de 1%. Estes parametros foram fixados, na época da implantação da pesquisa, em função de estudos realizados pelo U.S. Bureau of the Census e pela equipe do Grupo Executivo de Pesquisas Domiciliares do IBGE, assistida por assessores da

<sup>&</sup>lt;sup>2</sup> Ver Metodologia da Pesquisa Nacional por Amostra de Domicilios na década de 70.

United States Agency for International Development (USAID).

Para a década de 80, de posse dos resultados preliminares do Censo Demográfico, elaborou-se um desenho de amostra que, em linhas gerais, mantinha a filosofia do desenho utilizado na década de 70, adotando-se, entretanto, novas frações de amostragem. Gradativamente o nível de divulgação foi sendo extendido até atingir em 1984, o nível atual: Grande Regiões (separou-se a Região Norte da Centro-Oeste), Unidades da Federação (exceto para a Região Norte onde se considera o conjunto Rondonia-Acre-Roraima-Amapá), Regiões Metropolitanas e Distrito Federal.

Em 1986, foi realizado um estudo com a finalidade de reduzir a amostra, face às dificuldades enfrentadas pela equipe de campo, bem como, ao alto custo da pesquisa; a redução efetuada na amostra, entretanto, não provocou alterações na estrutura do desenho amostral.<sup>3</sup>

#### 3. DESENHO AMOSTRAL E OBJETIVOS DA PESQUISA

O primeiro passo no planejamento da amostra de uma pesquisa domiciliar é a explicitação dos objetivos quanto:

- às informações que se quer obter;
  - à população a ser investigada;
- à precisão requerida;
  - ao ambito geográfico;

além, é claro, dos recursos disponíveis para a sua realização.

Tais objetivos devem ser tão específicos, claros e precisos

quanto possível, dai resultando a adequabilidade da ferramenta

Ver Menezes et al.(1986).

utilizada, no caso a amostra, aos propósitos do levantamento.

Ficou evidenciado no módulo I - A Questão Conceitual - que a PNAD exige um desenho de amostra que atenda às particularidades decorrentes do fato da investigação do tema Trabalho ser de natureza complexa demandando, além da medição das variáveis ligadas ao tema - condição de atividade, posição na ocupação, etc - a investigação de outras derivadas do seu aspecto multidisciplinar, tais como: grau de instrução, migração, etc.

Uma amostra com esse objetivo é chamada de "múltiplos propósitos" e seu sucesso depende do grau de correlação existente entre as variáveis a serem investigadas, haja visto que a definição do processo de amostragem, quanto ao desenho a ser utilizado e ao dimensionamento da amostra, não poderá contemplar todas as informações que se deseja obter; muito pelo contrário, é baseado em um conjunto bastante reduzido de variáveis para as quais se vai identificar o melhor desenho dentre os que a teoria da amostragem nos fornece e o tamanho de amostra necessário para obtermos a precisão requerida. Para as demais informações, a amostra será tão eficiente quanto maior for o grau de correlação existente entre estas e as que balizaram o desenho amostral.

A adequabilidade da amostra para estimar parametros para o conjunto de variáveis de interesse está ligada, não só à obtenção de resultados dentro de um nível de precisão considerado suficiente, como também, à sua viabilidade sob o ponto de vista operacional e dos recursos disponíveis. A conduta normalmente utilizada quando se requer, para um desenho de amostra pré-estabelecido, uma maior precisão para as

estimativas - erros de amostragem mais reduzidos - é investigar um número maior de unidades, ou seja, aumentar-se o tamanho da amostra, elevando desta forma o custo da pesquisa, podendo acarretar, inclusive, dificuldades operacionais.

E da maior relevância considerar no planejamento de uma pesquisa por amostra, o que, ao menos em grandes linhas, consistirá no plano tabular objeto de divulgação. Este procedimento evitará que a posterior desagregação das informações alcance níveis incompatíveis com o desenho adotado, impossibilitando o uso e/ou a divulgação completa dos dados, levando a pesquisa, em grande medida, ao fracasso.

Da mesma forma que o nivel de detalhamento excessivo da informação pode trazer prejuizos aos resultados obtidos, a definição da população que se quer abranger também é de fundamental importancia para a determinação do plano amostral.

Numa pesquisa onde se pretende investigar características do mercado de trabalho, fica claro que estamos tratando, não com uma, mas com várias populações-objetivo: a população propriamente dita e suas partições específicas para o tema em estudo. Atender às necessidades de informações sobre cada uma destas partições, mantendo a precisão das estimativas e o tamanho da amostra em níveis aceitáveis, depende fundamentalmente da proporção da população que estas partições representam. Dizendo de outra forma, há que se ter cuidado em não aprofundar a investigação sobre subpopulações rarefeitas, sob pena de elevar-se o custo da pesquisa de forma proibitiva, face à necessidade de aumentar-se substancialmente o tamanho da amostra.

O nivel geográfico minimo de estimação de uma pesquisa

precisa também ser alvo de avaliação. A teoria da amostragem garante a estimação ao nivel minimo de cada estrato. Porém, a essas estimativas podem estar associados erros de amostragem que, em muitas ocasiões, poderão não atender às necessidades dos usuários no que diz respeito à significancia desses resultados. Mais uma vez, a forma de garantir a precisão das estimativas a nível do estrato é pelo aumento do tamanho da amostra.

O requerimento de maior precisão para as estimativas, que acarreta um aumento do tamanho da amostra, conduz à questão chave relativa ao custo do levantamento. Assim, apesar de ser possivel estabelecer para o erro de amostragem um limite tão baixo quanto se queira, sempre existirá a restrição dos recursos disponíveis que determinará até que ponto o crescimento da amostra será viável ou não. Contudo, esta limitação não é conhecida, hoje, de forma clara, isto é, não se conhece a priori o total de recursos disponível para a realização da PNAD, tendo-se apenas a idéia de que o montante atual é considerado bastante elevado; por isso, esforços deverão ser feitos para a redução destes custos. Cabe ressaltar a dificil tarefa que se apresenta, qual seja, a de compatibilizar a necessidade de redução de custos com a ampliação dos objetivos que se pretende atingir com a nova PNAD.

A elaboração de uma função custo, o mais detalhada possível, será de grande utilidade, não só no momento em que se fará a opção pelo desenho de amostra definitivo, como também, para o acompanhamento e/ou futuras reformulações pelas quais o esquema amostral poderá passar.

E interessante lembrar também que, em pesquisas de ambito nacional como a PNAD, o número de unidades investigadas, quer para atender a um maior detalhamento da investigação, quer para atender à necessidade de maior precisão no que já seja objeto de levantamento, determina não só um aumento de recursos financeiros, mas também um requerimento crescente na melhoria das condições de gerenciamento das diversas etapas da pesquisa.

#### 3.1. A SITUAÇÃO DO DOMICILIO

Um reclamo dos usuários vem sendo o fato da PNAD não fornecer resultados confiáveis para as áreas rurais. Não se trata apenas de dizer que o desenho atual privilegia as áreas urbanas, posto que a contribuição proporcional de cada um dos tipos de área (urbana e rural) para a amostra se aproxima bastante da realidade observada no Censo Demográfico de 1980. A restrição que se pode fazer, contudo, é que o desenho atual não foi estabelecido para produzir estimativas, especialmente para a área rural, com um nível de precisão pré-definido.

Qualquer decisão que se tome, com o intuito de fazer com que hoje a amostra seja mais representativa de fenómenos da área rural, implicará num volume de recursos adicionais dirigidos à rede de coleta do IBGE que, devido a todo tipo de carência, encontra-se despreparada inclusive para realizar a PNAD dentro dos moldes atuais.

Este ponto merece também ser tratado sob outra ótica. Seria a situação do domicílio uma dimensão relevante a ser incorporada sob o ponto de vista do uso da informação? Deve-se ter presente que a dicotomia urbano/rural é, em grande medida,

determinada por decisões políticas a nível dos governos municipais. Colocada dessa maneira, pode-se questionar, de imediato, a validade da situação do domicílio como relevante em termos da demanda por informações ou, mais precisamente, das análises desenvolvidas nas diversas áreas do conhecimento.

Uma outra questão diz respeito à capacidade da amostra da PNAD conseguir, ao longo do tempo, ser representativa dessa realidade, ou seja, acompanhar as modificações no contexto urbano/rural. A classificação dos setores segundo a situação do domicílio é determinada à época da realização do Censo Demográfico. Entretanto, à medida que a PNAD se afasta do ano censitário, suas estimativas não incorporam as alterações do ponto de vista da situação do domicílio, trazendo prejuízo direto do lado da oferta ou produção do dado, sendo, também, ôbvios seus efeitos quanto ao uso da informação.

Também sob estes aspectos deve ser avaliada a pertinência da ampliação da abrangência geográfica da PNAD de forma a incorporar a área rural da Região Norte, tornando-a realmente uma pesquisa de ámbito nacional. Ressalte-se ainda que, as dificuldades operacionais em função das características peculiares da área em foco, exigirão um esforço muito maior por parte da rede de coleta do IBGE do que aquele requerido para as demais áreas rurais do país.

#### 4. VARIAVEIS RELEVANTES PARA O DIMENSIONAMENTO DA AMOSTRA

Na PNAD, por se tratar de uma pesquisa que pretende investigar uma vasta gama de variáveis, é praticamente impossível identificar um tamanho de amostra que assegure um

nivel de precisão pré-definido para todas as variáveis. O procedimento usual é o de escolher uma variável para orientar o delineamento da amostra e um conjunto pequeno de variáveis relevantes para as quais se fará uma avaliação prévia da precisão alcançada.

A definição do desenho da amostra consiste na especificação dos niveis de estratificação e/ou conglomeração, dos procedimentos de seleção das unidades em cada estágio e do método de expansão a ser utilizado, o que torna possível identificar o estimador adequado para cálculo da variancia da estimativa de uma dada variável. A partir da fixação de alguns parametros, entre eles, a variância, parte-se para o dimensionamento da amostra.

A escolha das variáveis está intimamente ligada às questões abordadas no módulo I - A Questão Conceitual e deverá, portanto, contar com a participação de toda a equipe envolvida com o projeto. Destaca-se, entretanto, que no planejamento da amostra deve ser levada em conta a demanda pela investigação das relações de trabalho agrícola e do não-agrícola. Isto significa que a definição e a avaliação do plano amostral devem contemplar variáveis que se relacionem com as populações-objetivo Agrícola e Não-Agrícola da PNAD.

A definição do desenho amostral será feita com base em simulações utilizando as informações do Censo Demográfico de 80. Estas simulações têm como objetivo testar variantes de desenhos amostrais para escolha daquele que melhor atenda às necessidades teórico-operacionais da pesquisa. Estabelecido o desenho a ser adotado, e uma vez disponíveis os dados do Censo vindouro, passa-se ao dimensionamento "definitivo" da amostra

que contemplará a nova malha setorial do país. E importante mencionar aqui que, face à necessidade de investigação da questão agrícola, será preciso contar com as informações do boletim da amostra, CD - 1.02, já que, no questionário básico e mesmo na folha de coleta, as informações são insuficientes.

#### 5. ESTAGIOS DE SELEÇÃO

O processo atual de seleção da amostra da PNAD compreende três estágios.

A unidade primaria de amostragem, definida pelo municipio, teve os seguintes aspectos considerados para sua escolha:

- limites físicos identificáveis no campo;
- estatisticas recentes e outras informações que pudessem ser utilizadas nos processos de estratificação, seleção e outras operações; e,
- o tamanho da unidade.

Uma vez definida e caracterizada a unidade primária de amostragem, foi feita sua estratificação. A seleção destas unidades, dentro de cada estrato, realizou-se com probabilidade proporcional à população total, dada pelo Censo Demográfico. Os municípios das Regiões Metropolitanas(com algumas exceções), os das capitais e aqueles cuja população ultrapassou a metade do tamanho médio do estrato constituíram o grupo dos municípios autorepresentativos, formando assim, cada um deles, um estrato isolado dentro da UF. Os demais formaram o grupo dos municípios não autorepresentativos.

Os setores censitários, além de servirem como áreas de controle e execução da coleta do censo, formam um cadastro de

areas do país, de grande aplicação como estágio intermediário de amostragem, principalmente nas pesquisas domiciliares. Como os setores tinham, em sua grande maioria, limites físicos bem definidos e se encontravam suficientemente mapeados, permitindo fácil identificação no campo, foram então definidos como unidades secundárias de amostragem e selecionados com probabilidade proporcional ao número de domicilios existentes na época do censo.

Desde 1976, os domicilios vém sendo selecionados diretamente dos setores incluidos na amostra, a partir de uma listagem obtida ao inicio de cada realização da pesquisa, a fim de manter atualizado o cadastro de seleção destas unidades.

Com relação aos estágios de seleção o que se tenciona verificar é a vantagem em se manter o município como unidade primária de amostragem. Se por um lado, sua manutenção pode levar a um menor espalhamento geográfico da amostra, facilitando, principalmente, o deslocamento para a realização das entrevistas, por outro, gera uma perda de precisão devido ao efeito de conglomeração. No caso da sua eliminação, teriamos como vantagem adicional, a simplificação nos procedimentos para definição do desenho. Além disso, existe a possibilidade de que um investimento numa estratificação eficiente de setores nos leve a uma redução significativa no número de setores na amostra em comparação à situação de manter o município como estágio de seleção.

#### 6. ESTRATIFICAÇÃO

Vários fatores sugerem a adoção de estratificação no plano amostral da PNAD. Em primeiro lugar, dado um tamanho de amostra, a estratificação normalmente assegura a obtenção de estimativas mais precisas que aquelas derivadas de um plano amostral sem estratificação ou, de outra forma, para um dado nível de precisão, a estratificação normalmente assegura a redução do tamanho da amostra em relação a um plano não estratificado.

Uma vantagem adicional a ser considerada é a possibilidade de usar a estratificação para definição dos niveis de estimação, sejam eles geográficos e/ou referentes a subpopulações.

O Módulo I, como já se viu, aponta para a necessidade de investigação das características do trabalho agricola e não-agricola. Para tanto, pretende-se distinguir a população ocupada segundo o tipo de atividade que realiza e, então, formular questões específicas para cada grupo. E fácil imaginar que, para este objetivo ser alcançado de forma satisfatória, é necessário que o plano amostral garanta a representatividade dos dois grupos e, por conseguinte, a precisão das estimativas a serem obtidas. Uma maneira de solucionar esta questão é fazer com que a(s) variável(eis) que caracterizará(ão) a separação dos grupos seja(m) utilizada(s) para uma primeira estratificação da(s) unidade(s) do(s) primeiro(s) estágio(s) (municípios e/ou setores) para posterior definição do tamanho da amostra em cada estrato.

Entretanto, este pode não ser o único nível de estratificação das unidades. Avançar mais sob esta ótica pode mostrar-se necessário; como por exemplo, a adoção de estratificação por áreas geográficas, visando a representatividade espacial da amostra, o que, naturalmente, pode acarretar o espalhamento da amostra e não trazer, como seria esperado, ganhos significativos sob o ponto de vista da precisão das estimativas. As simulações constituir-se-ão num excelente instrumento para a definição de outros níveis.

Atualmente, a PNAD utiliza um modelo em que, em cada Unidade da Federação, somente os municípios foram estratificados. O critério de estratificação foi de pertencerem a uma mesma Microrregião Homogénea. De certa forma, podemos dizer que se trata de um processo de estratificação basicamente geográfica, embora a definição dos limites das microrregiões obedeça a critérios de homogeneidade, os quais, não necessariamente, são os mais apropriados para a pesquisa. Os estratos foram construidos de forma a possuir, na medida do possivel, um mesmo tamanho, relativamente à sua população total, o que fez com que nem sempre se conseguisse manter em um único estrato os municípios pertencentes à mesma microrregião. Além disso, alguns municipios foram considerados como estratos certos, ou seja, passaram a ser incluidos com certeza na amostra, o que, ao menos ao nivel teórico, permitiria obter estimativas com precisão controlada para cada um deles. Entretanto, somente para as Regiões Metropolitanas a amostra foi dimensionada de forma a permitir a divulgação informações. Contudo, cabe avaliar se, para a próxima década, a PNAD deve manter o mesmo nivel de divulgação. E grande a

demanda por informações a nível dos grandes municípios. O atendimento a essa demanda, porém, pode implicar num custo proibitivo. Há que se analisar a relação custo x beneficio com bastante cuidado. Uma alternativa seria considerar os municípios de maior relevência (a questão é qual seria o critério?) como estratos certos ou municípios autorepresentativos, deixando aos usuários a tarefa de avaliar se a precisão obtida com a amostra nesse nível é adequada aos seus propósitos.

Diante da decisão por um desenho estratificado, além da escolha da(s) variável(eis) e níveis de estratificação, mais dois pontos deverão ser considerados:

- 1) a decisão sobre o número de estratos;
- 11) o critério de alocação da amostra, ou seja, a determinação do tamanho da amostra em cada estrato.

#### 7. METODO DE SELEÇÃO

Pelo desenho atual da PNAD a amostra, como já se viu, é selecionada em tres etapas. Os municípios são selecionados de forma independente, dentro de cada estrato, com probabilidade proporcional ao total da população. Em seguida, selecionam-se os setores, também com probabilidade proporcional, mas neste caso a medida de tamanho utilizada é o número de domicílios existentes por ocasião do Censo Demográfico de 1980. Finalmente, são selecionados sistematicamente os domicílios com probabilidades iguais, de forma que, ao final do processo, tem-se uma amostra autoponderada, isto é, em que todo domicílio tem a mesma chance de ser incluído na amostra (chance igual à

fração global de amostragem).

O critério de definição das probabilidades das unidades de 1º e 2º estágios, associado ao processo de estratificação utilizado, leva-nos hoje a dizer que o desenho da amostra "privilegia" os municipios e/ou setores de maior concentração populacional, apresentando, portanto, um alto nivel de urbanização. Assim sendo, os resultados obtidos pela PNAD considerados bastante satisfatórios para as atividades urbanas. o mesmo não acontecendo para as rurais. Entretanto, deve-se salientar que, face à grande variabilidade dos tamanhos dos conglomerados (municípios e setores), o uso de outra forma de atribuição de probabilidades de seleção nesses estágios (iguais, por exemplo) poderia trazer graves prejuizos à precisão das estimativas. Uma alternativa para tratar o problema para as próximas PNADs seria a estratificação dessas unidades segundo o critério a ser estabelecido para a PNAD Agricola e a Não Agricola, o que viria atender às necessidades da pesquisa.

Um outro aspecto a ser abordado diz respeito ao envelhecimento das medidas de tamanho utilizadas para a fixação das probabilidades de seleção. Dado que essas são estabelecidas a partir dos resultados obtidos no Censo Demográfico, à medida que o ano de realização da pesquisa se afasta do ano censitário podem ocorrer modificações que impactarão mais uma vez a precisão das estimativas. Para contornar esta questão, utiliza-se um universo paralelo, denominado Novas Construções, através do qual se pretende manter inalterada, tanto quanto possível, a distribuição das probabilidades de seleção pelo recurso de incluir num cadastro à parte todos os domicílios dos

projetos (a partir de um determinado tamanho) construídos após o Censo Demográfico. Este recurso, contudo, apresenta problemas que serão discutidos num capitulo adiante.

#### 8. AUTOPONDERAÇÃO

O uso de amostras autoponderadas, em que a seleção das unidades dos primeiros estágios é feita com probabilidade proporcional a uma medida de tamanho definida adequadamente, traz como vantagem principal o equilibrio da carga de trabalho de campo. Isto é, mesmo tendo um estrato composto de conglomerados (municípios e/ou setores) de tamanhos variados, se a seleção se processa sob essas condições, o número de unidades a serem investigadas no estrato será o mesmo, independente de qual conglomerado foi selecionado. Ainda mais, se os estratos forem formados tendo o mesmo tamanho, a contribuição para a dimensão final da amostra será homogênea. Isso sem mencionar a vantagem de simplificação dos processos de estimação.

Entretanto, a autoponderação da amostra deve ser avaliada sob mais alguns aspectos, tratados a seguir:

i) no momento do dimensionamento da amostra, estabelece-se o número de domicilios a serem selecionados por setor; com o decorrer da pesquisa, esse número pode ser alterado em função do acrescimo ou decrescimo do número de unidades domiciliares existentes no setor. Diante do crescimento, há duas alternativas a considerar: aumentar o tamanho da amostra, de forma a tentar preservar a autoponderação (a exemplo da PNAD atual), o que pode fazer com que se perca o equilibrio dos

trabalhos de campo sem necessariamente obter uma elevação substancial no nível de precisão das estimativas; ou abandonar a autoponderação e controlar o aumento da amostra a partir do estabelecimento de patamares de crescimento. A questão é: controla-se o número de domicílios selecionados por setor - e nesse caso não teríamos mais o intervalo de seleção de domicílios fixo por setor - o que dificulta sobremaneira nas situações em que se é obrigado a realizar a operação de listagem juntamente com as entrevistas; ou define-se, com base no tamanho alcançado pelo setor, um estágio intermediário de seleção, o subsetor, o que por sua vez, também pode acarretar problemas operacionais, face à dificuldade de estabelecer limites físicos identificáveis no campo, e trazer complexidade para o processo de estimação.

Uma outra questão que abala a autoponderação na fase de seleção dos domicilios é a variabilidade no número de entrevistas efetivamente realizadas. A precisão que se estabeleceu será alcançada a partir da efetiva realização das entrevistas em todos os domicilios selecionados. Contudo, devido à defasagem entre a operação de listagem e a realização das entrevistas, a seleção é feita com base em todos os domicilios listados, incluindo os fechados, vagos e os de uso ocasional, que podem transformar-se em entrevistas realizadas, ou não. Para suprir as perdas provenientes das entrevistas não-realizadas pode-se dar um acrescimo no número de domicilios selecionados de acordo com o número esperado de não-entrevistas; e,

ii) no método de expansão atual da PNAD faz-se uso da

<sup>4</sup> Ver Seleção da amostra para a Pesquisa de Orçamentos Familiares.

simplificação do processo que advém da autoponderação, isto é, as estimativas são obtidas multiplicando a soma dos elementos da amostra por um peso constante, não havendo, portanto, tratamento para a não-resposta proveniente de entrevistas não realizadas. A esta facilidade de obtenção das estimativas se contrapõe alguma distorção nos resultados, haja visto que quase nunca é possível obter uma amostra realmente autoponderada, por exemplo, quando ela exige que seja selecionado um número fracionário de unidades.

Para a próxima década, as dicussões acerca do tema apontam como primeira idéia tentar manter a autoponderação apenas na fase do dimensionamento da amostra, abandonando-a na fase de expansão da amostra. O tratamento alternativo seria através do ajustamento dos fatores de expansão, considerando apenas as entrevistas realizadas. Neste caso, cada setor passaria a ter um peso diferenciado pelo número de entrevistas realizadas.

#### 9. A QUESTÃO CADASTRAL

Um elemento crítico no planejamento e na seleção de uma amostra é a garantia de que a população a ser investigada esteja sendo inteira e corretamente coberta. Na PNAD, isto aponta para a necessidade de se dispor de um cadastro, completo e atualizado, que relacione todos os domicilios existentes na área que se deseja investigar.

Uma vez mantida na PNAD revisada a seleção da amostra em multiplos estágios, tem-se que o cadastro da(s) unidade(s) do(s) primeiro(s) estágio(s) estará estabelecido a partir do

<sup>5</sup> Ver Bianchini-Expansão da amostra e precisão das estimativas da Pesquisa de Orçamentos Familiares.

Censo Demográfico, ou seja, o cadastro dos setores censitários constitue fonte teoricamente segura da cobertura da população.

Já para a seleção dos domicilios é necessário que se elabore, a cada realização da pesquisa, a listagem das unidades domiciliares existentes no setor, de modo a garantir que toda unidade de habitação tenha chance de ser selecionada.

Deve-se ter em mente que a tarefa de relacionar os domicilios precisa ser feita de maneira criteriosa, pois da qualidade deste trabalho vai depender boa parte do sucesso da pesquisa.

Dois aspectos estão sendo alvo de discussão. O primeiro deles diz respeito à imutabilidade da amostra quanto às unidades dos primeiros estágios. Atualmente municípios e setores são mantidos na amostra durante toda a década enquanto a seleção dos domicílios é renovada a cada ano. Pretende-se avaliar vantagens e desvantagens em se promover regularmente uma rotação, principalmente na amostra de setores, com o intuito de evitar a ocorrência de possíveis vícios.

O segundo aspecto relaciona-se com o cadastro de Novas Construções-NC. Conforme já se mencionou, a idéia básica deste cadastro é que ele contenha os projetos responsáveis por alterações sérias nos tamanhos dos setores, os quais se tratados em separado manteriam a distribuição remanescente não muito distinta da utilizada na seleção, preservando, portanto, a precisão do plano amostral.

Atualmente o levantamento de NC é feito em todos os municípios da amostra, cobrindo, então, tanto os setores selecionados quanto os que não fazem parte da amostra. Identificada uma área de NC, ela é excluída da área do setor

original, passando a fazer parte de um "universo paralelo", para o qual a seleção dos domicilios é feita diretamente a partir do municipio e de acordo com a fração global de amostragem, provocando, desta forma, que se tenha um plano amostral diferenciado em relação à "amostra normal".

Se por um lado, a questão teórica fica parcialmente satisfeita (o ideal seria que o levantamento fosse realizado inclusive nos municípios não selecionados), pelo lado operacional cria-se uma situação que se suspeita seja bastante grave; tem-se verificado, através dos anos, a grande quantidade de erros que vém sendo gerados a partir dessa fonte, tanto por duplicidade quanto por omissão de unidades.

Para a PNAD revisada, o tratamento do cadastro de Novas Construções tende a merecer uma grande dose de atenção. Pretende-se verificar se o procedimento ora utilizado vem cumprindo eficientemente o seu papel, isto é, se estará realmente sendo capaz de manter a distribuição dos tamanhos dos setores. Além disso, como minimizar as dificuldades operacionais? Como acompanhar e controlar sua elaboração? Como adequá-lo ao desenho amostral a ser proposto? Existe alguma forma segura de eliminá-lo?

São intimeras as questões a serem examinadas e, no caso da manutenção do cadastro de NC, deve-se procurar alcançar a garantia de que se trata de algo confiável e que, ao contrário de ser uma fonte de erros alheios à amostragem, venha a cumprir sua finalidade, sem os transtornos operacionais hoje existentes.

A possibilidade de se alcançar o nivel de precisão estabelecido para as estimativas depende da capacidade da pesquisa em que sejam realizadas as entrevistas nos domicilios selecionados.

E claro que, como em toda pesquisa por amostra, existe uma relação direta entre o cadastro utilizado e a situação da entrevista. Quanto mais completo e atualizado o cadastro, maiores as chances de se realizarem as entrevistas. Entretanto, face à defasagem entre a época de realização da listagem e das entrevistas, é razoável esperar alguma desatualização desse cadastro, motivo pelo qual dá-se chance de seleção não só aos domicílios ocupados como também aos não-ocupados, na expectativa que essa situação se altere quando da realização das entrevistas, garantindo, assim, chance de seleção para toda a população. O mais freqüente é que, no computo final, o tamanho da amostra seja menor do que aquele definido, acarretando, conseqüentemente, perda de precisão.

Pode-se concluir que os domicilios selecionados são classificados em eleitos e não eleitos para o levantamento. Os domicilios onde a entrevista é realizada, bem como aqueles com não entrevista tipo A (fechado, recusa) são considerados como eleitos. Por outro lado, os domicilios com não entrevista tipo B (uso ocasional, improvisado, etc ) e tipo C (vagos, inexistentes, em obras, ruinas) são classificados como *não eleitos*.

Para tratar a falta de resposta na etapa de coleta dos

dados, provenientes da seleção de domicilios não eleitos e da perda de entrevistas em domicilios eleitos (recusa, fechado), foram discutidas duas alternativas: a substituição dos domicilios que não se conseguiu entrevistar, ou a definição antecipada de um incremento na amostra visando cobrir as eventuais perdas. A substituição de elementos da amostra, contudo, pode acarretar vicios que prejudicarão os resultados da pesquisa. Apresenta-se como melhor opção a seleção dos domicilios da amostra em número maior que aquele definido no dimensionamento, baseado numa proporção esperada de entrevistas não realizadas, embora implicando numa elevação dos custos.

Independente do fato de se procurar manter o tamanho final esperado da amostra, é necessário que se estabeleça procedimentos para controle e acompanhamento da coleta, tais como: estudos comparativos entre as situações verificadas por ocasião da listagem e das entrevistas; avaliação de taxas obtidas a partir da situação de entrevista (taxa de eleitos, taxa de resposta, taxa de recusa, etc). Para isso, uma análise de qual deve ser o conteúdo do Arquivo de Controle da Amostra deverá ser realizada, inclusive avaliando a possibilidade de se ter, total ou parcialmente, as variáveis deste arquivo contidas no arquivo de dados da pesquisa com o intuito de facilitar, não só aos usuários internos, mas, principalmente, aos externos ao IBGE, na utilização da amostra da PNAD para a realização de seus trabalhos.

#### 11. ESTIMAÇÃO

Faz parte da definição do desenho amostral a obtenção de seus estimadores. A questão que se apresenta é sobre o método de expansão a ser utilizado.

Desde a sua implantação, a PNAD, assim como outras pesquisas domiciliares por amostra, vem empregando a técnica de ajustar suas estimativas obtidas através do estimador natural, com base na informação da população projetada por métodos analiticos - é o que se denomina expansão por estimadores de razão. A utilização dessa técnica demonstra, entre outras coisas, a preocupação com o nivel de precisão alcançado pelas estimativas com o decorrer dos anos de realização da pesquisa. Isto porque os planos amostrais de pesquisas deste tipo, de maneira geral, seguem o modelo clássico de pesquisas do género: estratificado, em multiplos estágios, onde o setor e o domicilio constituem as unidades dos altimos estágios e os setores são selecionados com probabilidade proporcional a alguma medida de tamanho. Dessa forma, quão próximos estivermos da realização do Censo, de onde se obtém as medidas de tamanho, mais próximos estaremos do valor estabelecido para o erro de amostragem das variáveis escolhidas. Entretanto, com o passar dos anos, o envelhecimento da base setorial faz com que a medida de tamanho utilizada na seleção dos setores não corresponda mais ao seu tamanho atual, o que pode fazer com que as variancias cresçam consideravelmente. Já vimos que cadastro de Novas Construções é uma das formas pelas quais tenta-se minimizar o problema; uma outra forma é pelo uso do

estimador de razão.

Sobre essa segunda forma, vém sendo discutidas as diferenças observadas quando se comparam os resultados obtidos pela expansão das amostras através de seus estimadores naturais e utilizando o ajuste da população projetada. A título de exemplo, tem-se que, em 1987, para o conjunto das Regiões Metropolitanas, além das Unidades da Federação de Goiás e Distrito Federal, a PNAD teria subestimado a população em 20% em relação à projetada, enquanto que a Pesquisa de Orçamentos Familiares -POF- o percentual de subestimação seria de 15% para as mesmas áreas exceto no caso de Goiás onde só foi considerado o município de Goiánia.

E claro que o envelhecimento da base setorial, o não tratamento da não-resposta (no caso da PNAD), entre outros fatores, podem estar influenciando os resultados das pesquisas e precisam ser cuidadosamente avaliados. Mas, e quanto à precisão associada à projeção da população? O fato das migrações internas não serem incorporadas no modelo para cálculo das projeções de população deve contribuir significativamente para que os valores projetados se distanciem dos valores reais, especialmente em níveis geográficos detalhados.

Em vista destas questões, há que se estudar criteriosamente o processo de expansão a ser utilizado para a próxima década, inclusive tratando a possibilidade de se produzir as estimativas mediante o uso, não só dos estimadores de razão, como também, dos estimadores naturais.

Ver Bianchini(1990).

#### 12. CONSIDERAÇÕES FINAIS

A definição de um desenho amostral para a PNAD revisada será decorrente das análises a serem realizadas através de simulações que serão feitas a partir das informações do Censo Demográfico de 1980.

Cabe adiantar alguns aspectos já mencionados que permitirão explicitar variantes de desenhos amostrais e que devem ser testados nas simulações:

- alternativas de variáveis e niveis de estratificação;
- manutenção ou eliminação do estágio de seleção dos municípios;

Além das simulações, algumas análises com os dados da PNAD da década de 80 auxiliarão nas tomadas de decisão com relação a caminhos que venham a ser propostos para o desenho amostral da pesquisa, como por exemplo:

- manutenção ou eliminação do cadastro de NC;
- definição do procedimento para controle e acompanhamento da amostra.

Por fim, um dos objetivos a serem alcançados com a PNAD revisada deve ser a função custo da pesquisa que, além de auxiliar na definição do desenho amostral, será um importante elemento de acompanhamento e controle para a pesquisa.

#### BIBLIOGRAFIA

ATLANTIDA: A case study in household sample surveys - unit IV - sample design. Washington, D. C.: Bureau of the Census, 1966.

BIANCHINI, Zélia Magalhães. Expansão da amostra e precisão das estimativas da Pesquisa de Orçamentos Familiares. Rio de Janeiro: IBGE, Departamento de Indices de Preços, 1989.

BIANCHINI, Zélia Magalhães. Projeções de população e o ajuste das estimativas das pesquisas domiciliares: solução ou problema? <u>Indicadores IBGE</u>, Rio de Janeiro, v.9, n.2, p.91- 9, 1990.

BIANCHINI, Zélia Magalhães. O tratamento da não-resposta da Pesquisa de Orçamentos Familiares. Rio de Janeiro: IBGE, Departamento de Indices de Preços, 1989.

HANSEN, Morris H.; HURWITZ, William N; MADOW, William G. Sample Suveys Methods and Theory. New York: John Wiley, 1953.

MENEZES, Antonio Carlos Fernandes et al. Pesquisa Nacional por Amostra de Domicilios: redução do número de domicilios na amostra e proposta de novos estimadores. [Rio de Janeiro: IBGE, 1986]

METODOLOGIA da Pesquisa Nacional por Amostra de Domicilos na década de 70. Rio de Janeiro: IBGE, 1981. (Relatórios Metodológicos, v1).

RINALDI, Doris; VIANNA, Marcia Coelho de Segadas; BRITO, Sebastiana Rodrigues de. PNAD - Anteprojeto de Revisão (módulo I): a questão conceitual. Rio de Janeiro: IBGE, Departamento de Emprego e Rendimento, 1990.

SELEÇÃO da amostra para a Pesquisa de Orçamentos Familiares. Rio de Janeiro: IBGE, Coordenadoria de Metodologia, 1986.

#### Textos para Discussão já publicados

- \* Pesquisas Contínuas da Indústria Vol. 1, nº I, jan. 1988
- \* Pesquisas Agropecuárias Contínuas: Metodologia Vol. I, n-2, 1988
- \* Uma Filosofia de Trabalho: As experiências com o SNIPC e com o SINAPI Vol. I, nº 3, março 1988
- \* O Sigilo das Informações Estatísticas: Idéias para reflexão -Vol. I, nº 4, abril 1988
- \* Projeções da População Residente e do Número de Domicílios Particulares Ocupados: 1985-2020 Vol. I,  $n^2$  5, maio 1988
- \* Classificação de Atividades e Produtos, Matérias-Primas e Serviços Industriais: Indústria Extrativa Mineral e de Transformação Vol. 1,  $n^{\circ}$  6, agosto 1988
- \* A Mortalidade Infantil no Brasil nos Anos 80 Vol. I,  $n^{\circ}$  7, setembro 1988
- \* Ensaio sobre o Produto Real da Agropecuária Vol. I, nº 9, setembro 1988
- \* Principais Características das Pesquisas Econômicas, Sociais e Demográficas - Vol. I, número especial, outubro 1988
- \* Novo Sistema de Contas Nacionais, Ano Base 1980 Resultados Provisórios - Vol. I, nº 10, dezembro 1988
- \* Pesquisa de Orçamentos Familiares Metodologia para Obtenção das Informações de Campo - nº 11, janeiro 1989
- \* De Camponesa a Bóia-fria: Transformações do trabalho feminino nº 12, fevereiro 1989
- \* Pesquisas Especiais do Departamento de Agropecuária Metodologia e Resultados nº 13, fevereiro 1989
- \* Brasil Matriz de Insumo-Produto 1980 nº 14, maio 1989
- \* As Informações sobre Fecundidade, Mortalidade e Anticoncepção nas PNADs  $n^{\circ}$  15, maio 1989

- \* As Estatísticas Agropecuárias e a III Conferência Nacional de Estatística  $n^{\circ}$  16, junho 1989
- \* Brasil Sistema de Contas Nacionais Consolidadas nº 17, agosto 1989
- \* Brasil Produto Interno Bruto Real Trimestral Metodologia  $n^{\circ}$  18, agosto 1989
- \* Estatísticas e Indicadores Sociais para a Década de 90 nº 19, setembro 1989
- \* Uma análise do Cotidiano da Pesquisa no DEREN (As Estatísticas do Trabalho)  $n^{\circ}$  20, outubro 1989
- \* Coordenação Estatística Nacional Reflexões sobre o caso brasileiro - nº 21, novembro 1989
- \* Pesquisa Industrial Anual 1982/84 Analise dos Resultados n-22, novembro 1989
- \* O Departamento de Comércio e Serviços e a III Conferência Nacional de Estatística - nº 23, dezembro 1989
- \* Um projeto de Integração para as Estatisticas Industriais n-24, dezembro 1989
- \* Cadastro de Informantes de Pesquisas Econômicas nº 25, janeiro 1990
- \* Ensaios sobre a Produção de Estatística nº 26, janeiro 1990
- \* O Espaço das Pequenas Unidades Produtivas: Uma tentativa de delimitação nº 27, fevereiro 1990
- \* Uma Nova Metodologia para Correção Automática no Censo Demográfico Brasileiro: Experimentação e primeiros resultados nº 28, fevereiro 1990
- \* Notas Técnicas sobre o Planejamento de Testes e Pesquisas Experimentais - nº 29, março 1990
- \* Estatísticas, Estudos e Análises Demográficas Uma visão do Departamento de População  $n^{\circ}$  30, abril 1990
- \* Crítica de Equações de Fechamento de Empresas no Censo Econômico de 1985  $n^{\circ}$  31, maio 1990

- \* Efeito de Conglomeração da Malha Setorial do Censo Demográfico de 1980  $n^{\circ}$  32, maio 1990
- \* A Redução da Amostra e a Utilização de Duas Frações Amostrais no Censo Demográfico de 1990  $n^{\circ}$  33, junho de 1990
- \* Estudos e Pesquisas de Avaliação de Censos Demográficos 1970 a  $1990 n^{\circ}$  34, julho de 1990
- \* A Influência da Migração no Mercado de Trabalho das Capitais do Centro-Oeste - 1980 - nº 35, agosto de 1990
- \* Pesquisas de Conjuntura: Discussão sobre Variáveis a Investigar  $n^{\circ}$  36, setembro de 1990
- \* Um Modelo para Estimar o Nível e o Padrão da Fecundidade por Idade com Base em Parturições Observadas nº 37, outubro de 1990
- \* A Estrutura Operacional de Uma Pesquisa por Amostra  $n^{\circ}$  38, novembro de 1990
- \* Produção Agrícola, Agroindustrial e de Máquinas e Insumos Agrícolas no Anos 80: Novas Evidências Estatísticas - nº 39, dezembro de 1990
- \* A Inflação Medida pelo Índice de Preços ao Consumidor nº 40, janeiro de 1991
- \* A Participação Política Eleitoral no Brasil 1988, Análise Preliminar  $n^{\circ}$  41, fevereiro de 1991
- \* Ensaios sobre Estatísticas do Setor Produtivo nº 42, março de 1991
- \* A Produção Integrada de Estatística e as Contas Nacionais: Agenda para Formulação de um Novo Plano Geral de Informações Estatísticas e Geográficas  $n^{\circ}$  43, março de 1991
- \* Matriz de Fluxos Migratórios Intermunicipais Brasil 1980 nº 44, abril de 1991
- \* Fluxos Migratórios Intrametropolitanos Brasil 1970-1980 nº-45, abril de 1991

- \* A Revisão da PNAD, A Questão Conceitual e Relatório das Contribuições  $n^{\circ}$  46, maio de 1991
- \* A Dimensão Ambiental no Sistema de Contas Nacionais  $n^{\circ}$  47, maio de 1991
- \* Estrutura das Contas Nacionais Brasileiras nº 48, junho de 1991
- \* Mercado do Couro e Resultados da Pesquisa Anual do Couro  $n^{\circ}$  49, junho de 1991
- \* As Estatisticas e o Meio Ambiente nº 50, julho de 1991
- \* Novo Sistema de Contas Nacionais Séries Correntes : 1981-85 Metodologia, Resultados Provisórios e Avaliação do Projeto - nº 51, julho de 1991 ( 2 Volumes : Volume 1-Metodologia, Resultados Provisórios e Avaliação do Projeto; Volume 2-Tabelas)
- \* O Censo Industrial de 1985 Balanço da Experiência  $n^{\circ}$  52, agosto de 1991
- \* Análise da Inflação nº 53, agosto de 1991