

Ministério do Planejamento, Orçamento e Gestão
Instituto Brasileiro de Geografia e Estatística – IBGE
Diretoria de Pesquisas
Departamento de Metodologia

Textos para discussão
Diretoria de Pesquisas
número 8

Principais Aspectos de Amostragem das Pesquisas Domiciliares do IBGE – revisão 2002

Zélia Magalhães Bianchini
Sonia Albieri

Rio de Janeiro
2002

Instituto Brasileiro de Geografia e Estatística - IBGE
Av. Franklin Roosevelt, 166 - Centro - 20021-120 - Rio de Janeiro, RJ - Brasil

ISSN 1518-675X **Textos para discussão. Diretoria de Pesquisas**

Divulga estudos e outros trabalhos técnicos desenvolvidos pelo IBGE ou em conjunto com outras instituições, bem como resultantes de consultorias técnicas e traduções consideradas relevantes para disseminação pelo Instituto. A série está subdividida por unidade organizacional e os textos são de responsabilidade de cada área específica.

ISBN 85-240-3065-8

© **IBGE. 2003**

Impressão

Gráfica Digital/Centro de Documentação e Disseminação de Informações - CDDI/IBGE, em 2003.

Capa

Gerência de Criação/CDDI

Bianchini, Zélia Magalhães

Principais Aspectos de Amostragem das Pesquisas Domiciliares do IBGE –
revisão 2002 / Zélia Magalhães Bianchini, Sonia Albieri. - Rio de Janeiro : IBGE,
Departamento de Metodologia, 2003.

27 p. - (Textos para discussão. Diretoria de Pesquisas, ISSN 1518-675X ; n. 8)

ISBN 85-240-3065-8

1. Levantamentos domiciliares - Métodos estatísticos. 2. Amostragem
(Estatística). I. Albieri, Sonia. II. IBGE. Departamento de Metodologia. III. Título.
IV. Série.

Gerência de Biblioteca e Acervos Especiais CDU 519.2:314.6
RJ/2003-1 EST

Impresso no Brasil / Printed in Brazil

Sumário

Sumário.....	3
Apresentação	5
Introdução	7
Pesquisa Nacional por Amostra de Domicílios (PNAD)	8
Pesquisa Mensal de Emprego (PME).....	10
Pesquisa de Orçamentos Familiares (POF)	12
Pesquisa sobre Padrões de Vida (PPV)	16
Pesquisa de Economia Informal Urbana (ECINF).....	18
A Amostra para a Coleta de Dados do Questionário Detalhado do Censo Demográfico	20
Resumo e Conclusões	21
Referências bibliográficas	26

Apresentação

Desde 1960, o IBGE vem usando amostragem probabilística na realização de suas principais pesquisas domiciliares. Este artigo descreve alguns aspectos metodológicos das principais pesquisas domiciliares por amostra realizadas pelo IBGE, a saber: a Pesquisa Nacional por Amostra de Domicílios (PNAD, iniciada em 1967); a Pesquisa Mensal de Emprego (PME, desde 1980); a Pesquisa de Orçamentos Familiares (POF, com três edições, em 1987/88 e em 1995/96 e em processo de coleta a de 2002/03); a Pesquisa sobre Padrões de Vida (PPV, realizada em 1996/97); a Pesquisa de Economia Informal Urbana (ECINF, realizada em 1997) e a amostra para a coleta do questionário detalhado do Censo Demográfico (desde 1960).

O artigo aborda as semelhanças e as diferenças principais entre as pesquisas, no que se refere a objetivos, população-alvo, abrangência geográfica, plano amostral, incluindo aspectos tais como: cadastros, estratificação, conglomeração, número de estágios, tamanho da amostra, seleção da amostra, taxas de não-resposta e tratamentos adotados, estimação, avaliação de erros amostrais.

Este documento é uma atualização do artigo “Uma Revisão dos Principais Aspectos dos Planos Amostrais das Pesquisas Domiciliares Realizadas pelo IBGE”, das mesmas autoras, publicado nessa mesma série, Textos para Discussão n° 91, de setembro de 1998.

Sonia Albieri
Chefe do Departamento de Metodologia

Introdução

O IBGE realiza várias pesquisas domiciliares com periodicidades diferentes (mensais, anuais, ou não definidas), com graus variados de abrangência geográfica (nacional, regional, regiões metropolitanas, apenas áreas urbanas, alguns municípios de capitais de estado) e complexidade, de acordo não só com os recursos disponíveis mas também com a área temática e objetivos de cada pesquisa. Este artigo revisa alguns aspectos metodológicos das principais pesquisas domiciliares por amostra realizadas pelo IBGE, a saber: a Pesquisa Nacional por Amostra de Domicílios (PNAD, iniciada em 1967); a Pesquisa Mensal de Emprego (PME, desde 1980); a Pesquisa de Orçamentos Familiares (POF, com três edições, em 1987/88, em 1995/96 e em processo de coleta a de 2002/03); a Pesquisa sobre Padrões de Vida (PPV, realizada em 1996/97); e a Pesquisa de Economia Informal Urbana (ECINF, realizada em 1997). As descrições atualizadas referem-se principalmente à PNAD, que teve mais 3 edições, a PME, que por ser mensal teve seus dados atualizados para o mês de agosto de 2002, o Censo Demográfico que teve mais uma realização em 2000 e a POF, que está em campo coletando uma nova rodada da pesquisa e para qual foram feitas várias modificações não só no plano amostral.

Todas essas pesquisas adotam amostras probabilísticas de domicílios. Os desenhos amostrais das pesquisas listadas acima possuem várias semelhanças, as quais incluem amostragem de conglomerados em dois estágios (setores censitários + domicílios) ou mesmo três estágios (municípios + setores censitários + domicílios) com estratificação das unidades primárias de amostragem (UPAs). As UPAs (municípios ou setores censitários) são selecionadas com probabilidade proporcional a uma medida de tamanho – ppt (dentro de cada estrato). Os dados dos Censos Demográficos são freqüentemente usados para construir medidas de tamanho. Os setores censitários possuem em média 300 domicílios na área urbana e 200 domicílios na área rural. Para cada setor selecionado para a amostra de cada pesquisa é preparada uma listagem de todos os seus domicílios com o objetivo de preparar um cadastro atualizado para a seleção dos domicílios no último estágio de seleção. A precisão das estimativas é medida através dos coeficientes de variação (CVs) calculados para um conjunto de estimativas. O método do conglomerado primário (*ultimate cluster*) de

Hansen et al (1953) é o que vem sendo usado na maioria das pesquisas para estimar a variância das estimativas de interesse.

As seções seguintes apresentam aspectos gerais de cada pesquisa aqui considerada, bem como uma breve descrição das principais diferenças entre as pesquisas, com relação a objetivos, população alvo, abrangência geográfica e plano amostral, incluindo aspectos tais como: cadastro, estratificação, conglomeração, número de estágios de seleção, tamanho da amostra, procedimento de seleção utilizado, taxas de não-resposta e tratamentos adotados, método de estimação e avaliação da precisão das estimativas.

Além dessas pesquisas, o sistema de pesquisas domiciliares inclui o Censo Demográfico, realizado a cada 10 anos, que usa amostragem para obter dados sobre características selecionadas de pessoas, famílias e domicílios. O plano amostral e a metodologia de estimação adotados na pesquisa amostral do Censo são particularmente diferentes daqueles utilizados nas pesquisas realizadas no período intercensitário. Na seção 7 é apresentada uma breve descrição da amostra usada para a coleta do questionário de amostra dos últimos Censos Demográficos.

Vale ressaltar que este artigo não tem por objetivo descrever com profundidade os planos amostrais das pesquisas. Os detalhes metodológicos em geral e específicos de amostragem de cada pesquisa podem ser encontrados nas referências relacionadas no final do artigo.

Pesquisa Nacional por Amostra de Domicílios (PNAD)

A PNAD foi implantada no País, gradativamente, a partir de 1967. O plano amostral da pesquisa foi inspirado no Plano Atlântida do *U.S. Bureau of the Census*. Desde então, essa pesquisa tornou-se a pesquisa domiciliar anual mais importante no Brasil. É uma pesquisa com múltiplos propósitos que investiga características econômicas e sociais, principalmente aquelas relativas à situação da força de trabalho. Possui abrangência nacional, com exceção da área rural da região Norte. Além disso, ocasionalmente são aplicados questionários suplementares na mesma amostra e ao mesmo tempo em que se coleta a pesquisa tradicional, com o objetivo de investigar tópicos específicos, tais como: saúde; educação; mão-de-obra e previdência; fecundidade feminina; situação do menor; anticoncepção, acesso a serviços de saúde, suplementação alimentar e associativismo; participação político-social e estoque de aparelhos utilizadores de energia; mobilidade social; trabalho das crianças de 5 a 9 anos; trabalho das crianças e adolescentes de 5 a 17 anos.

A pesquisa possui um plano amostral autoponderado¹ com três estágios de seleção, estratificado, popularizado na década de 60 (ver Kish, 1965), para uso em pesquisas

¹ Em um plano amostral autoponderado a probabilidade de uma unidade elementar qualquer pertencer à amostra é constante e igual à fração de amostragem geral, para um determinado nível geográfico.

domiciliares. As unidades primárias de seleção são os municípios, que são estratificados por tamanho (população), e selecionados sistematicamente com probabilidade proporcional ao tamanho. No segundo estágio, os setores censitários são selecionados também de forma sistemática e com pmt, sendo que nesse caso o tamanho é medido pelo número de domicílios. Uma amostra sistemática simples de domicílios é então selecionada no terceiro estágio.

O plano amostral adotado para a PNAD considera uma estratificação das unidades primárias (municípios), definida separadamente em cada unidade da federação, com seleção de duas unidades por estrato. Municípios pertencentes à mesma microrregião geográfica foram agrupados em estratos com aproximadamente o mesmo tamanho. Os dados de população provenientes do Censo Demográfico foram as medidas de tamanho usadas para os procedimentos de estratificação e de seleção dos municípios. Os municípios grandes em termos populacionais e aqueles pertencentes às regiões metropolitanas foram tratados cada um como um estrato e portanto incluídos na amostra com certeza, e denominados auto-representativos. Os demais municípios selecionados em cada estrato são denominados não-auto-representativos e em cada um foram selecionados 5 setores.

O plano amostral é caracterizado por fração amostral fixa para cada região metropolitana e para o restante da unidade da federação. Os municípios e os setores selecionados são mantidos na amostra até que estejam disponíveis os novos dados do Censo Demográfico, quando então são selecionadas novas unidades para a amostra. No momento em que foi feita a seleção de setores, o número de domicílios por setor para a amostra foi fixado e constante para todos os municípios. Quando a seleção da amostra foi atualizada com as novas definições de setores e com as medidas de tamanho baseadas nos dados do Censo Demográfico de 1991, o número de domicílios por setor na amostra foi fixado em 13. Com a realização do Censo Demográfico de 2000, além de novas informações sobre a população e os domicílios, houve uma redefinição da base de setores censitários. O território brasileiro foi dividido em um número bem maior de setores censitários do que havia em 1991. Assim, foi feita nova seleção da amostra, inclusive unidades primárias e foi feito também um ajuste nos tamanhos das amostras de setores para dar conta dessa nova partição. O número de domicílios por setor na amostra foi fixado em 16.

A cada ano, em cada setor selecionado para a amostra, é preparada (ou atualizada) no campo uma listagem de domicílios, produzindo um cadastro atualizado para a seleção dos domicílios. Uma característica importante dessa operação de listagem refere-se ao Cadastro de Novas Construções, que é preparado de forma a conter os projetos responsáveis por alterações sérias nos tamanhos dos setores. O levantamento das novas construções é feito nos municípios da amostra, tanto nos setores da amostra como naqueles não selecionados para a amostra. Uma área de novas construções é excluída da área do setor original e é tratada em separado no momento da seleção de domicílios, que nesse caso, é feita de acordo com a fração amostral da área.

Como a seleção de domicílios em cada setor selecionado para a amostra é feita sistematicamente, para garantir a autoponderação da amostra, o intervalo de seleção de domicílios permanece fixo de ano para ano. Esse procedimento acarreta um aumento anual no número de domicílios na amostra. Em um dado setor da amostra, o número de domicílios selecionados depende do tamanho atualizado do setor, em número de domicílios, atualização essa dada pela operação de listagem. Para ilustrar, na pesquisa de 1997, foram selecionados 109.541 domicílios sendo 16,4 domicílios em média por setor (veja tabela 2), enquanto que no dimensionamento da amostra com os dados de 1991, esse número médio foi fixado em 13. Para o ano 2001, foram selecionados 126.858 domicílios sendo 17,4 domicílios em média por setor.

O procedimento de estimação adotado na PNAD é baseado em estimação de razão com ajuste de população. Inicialmente os pesos associados aos domicílios são obtidos considerando o plano amostral adotado sem tratamento para não-resposta. Então, esses pesos são ajustados (multiplicados) por um fator que é calculado independentemente para cada região metropolitana e para o restante de cada unidade da federação. Esse fator é a razão entre a estimativa independente de população² para o período de referência da pesquisa e a estimativa de população proveniente da amostra. Os pesos associados às pessoas são aqueles calculados para os domicílios onde moram.

A precisão das estimativas provenientes da amostra é medida pelo coeficiente de variação (CV) calculado para um conjunto selecionado de estimativas de totais de variáveis categóricas. A variância do estimador de razão adotado é calculada pelo método denominado conglomerado primário (Hansen et al., 1953). As estimativas dos coeficientes de variação para variáveis categóricas são divulgadas através do ajuste de um modelo de regressão do tipo *generalized variance functions* (Wolter, 1985). Ou seja, para cada domínio de publicação dos resultados, são ajustados modelos de regressão para explicar os CVs das estimativas de total como função das próprias estimativas. Os coeficientes dos modelos de regressão ajustados são publicados juntamente com os resultados da pesquisa a cada ano. Desde 1984, os resultados são publicados nos níveis nacional, grandes regiões, unidades da federação e nove regiões metropolitanas.

Pesquisa Mensal de Emprego (PME)

A Pesquisa Mensal de Emprego foi iniciada em 1980. É uma pesquisa por amostra domiciliar realizada para fornecer estimativas do nível e das variações no emprego, desemprego e de outras características da força de trabalho. É realizada em seis regiões metropolitanas, a saber: Rio de Janeiro, São Paulo, Belo Horizonte, Recife, Salvador e Porto Alegre. Os resultados são produzidos para cada uma dessas regiões, bem como para o agregado das seis regiões, agregado esse que em 2002 representava cerca de 25% da população brasileira. Para cada uma dessas regiões, a amostra foi selecionada em 2

² As estimativas independentes de população são preparadas pelo próprio IBGE anualmente, para diversos níveis geográficos. O menor nível geográfico é o município.

estágios de forma independente. Em cada região metropolitana, os municípios constituem estratos de setores, os quais por sua vez são as unidades primárias de seleção dessa pesquisa. Os setores são selecionados sistematicamente com probabilidade proporcional a uma medida de tamanho (número de domicílios). Em cada região metropolitana, foi usada alocação proporcional dos setores ao longo dos estratos (municípios). O número de domicílios a serem selecionados por setor foi fixado em 20, e tal como na PNAD, após a atualização da listagem de domicílios em cada setor da amostra, esse número aumenta de forma a manter a autoponderação. O Cadastro de Novas Construções preparado para a PNAD também é utilizado na PME. O tamanho da amostra referente ao mês de agosto de 2002 é apresentado na tabela 2, e o número médio de domicílios selecionados por setor nesse mês é de 26,8.

Como a pesquisa é repetida a cada mês, foi estabelecido um esquema de rotação para a amostra a fim de evitar problemas de falta de cooperação dos entrevistados, que normalmente ocorrem em pesquisas por painéis fixos de domicílios. O procedimento de rotação foi definido de tal forma que cada domicílio permanece na amostra por 4 meses consecutivos, é retirado da amostra por 8 meses, retorna para mais 4 meses de pesquisa e, então, é retirado definitivamente da amostra. Esse procedimento resulta em uma superposição de 75% da amostra a cada mês e de 50% a cada ano. Há também um processo de substituição dos setores que, por força do sistema de rotação da amostra, tiveram todos os seus domicílios incluídos na amostra.

O procedimento de estimação é semelhante ao da PNAD. É baseado no estimador de razão com ajuste para a população estimada para a data de referência³. Para todas as estimativas de indicadores ou de totais calculados e divulgados a cada mês, são levadas em consideração apenas as informações da amostra daquele mês, ou seja, o aspecto longitudinal da amostra não é aproveitado na definição do estimador usado. Os coeficientes de variação são calculados para um conjunto selecionado de estimativas, usando o método do conglomerado primário.

As grandes transformações ocorridas no mercado de trabalho brasileiro desde a implantação da PME impuseram uma revisão tanto na parte conceitual como na forma de indagação visando possibilitar a captação mais adequada das características do trabalhador e de sua inserção no sistema produtivo. O novo questionário foi coletado de outubro de 2001 a dezembro de 2002, com duplicação do levantamento da pesquisa, como recomenda a boa prática estatística para que as diferenças entre as estimativas dos dois levantamentos pudessem ser avaliadas por um determinado período e explicada / entendida pelos usuários. Este novo questionário está totalmente atualizado em relação às últimas recomendações internacionais, incluindo as da última Conferência da Organização Internacional do Trabalho – OIT de 1998⁴. Outra grande inovação incorporada na nova

³ As estimativas independentes de população são preparadas pelo próprio IBGE anualmente, para diversos níveis geográficos. O menor nível geográfico é o município.

⁴ Veja *Pesquisa Mensal de Emprego. Série Relatórios Metodológicos, volume 23*.

pesquisa é a implementação da captura dos dados com entrevista pessoal assistida por computador portátil e com programa específico de entrada de dados.

A lista de indicadores tradicionais foi ampliada para possibilitar uma melhor análise das condições de ocupação e inatividade. A pesquisa abrangia as pessoas de dez anos ou mais de idade mas disponibilizava os indicadores referentes às de quinze anos ou mais. A idade mínima para a nova divulgação é dez anos ou mais.

Quando da divulgação dos indicadores da nova pesquisa, em dezembro de 2002, o IBGE apresentou uma análise comparativa dos resultados obtidos das duas formas de investigação no período de outubro de 2001 a outubro de 2002⁵. A partir de janeiro de 2003 apenas a nova pesquisa estará em campo.

No que diz respeito aos aspectos de amostragem, cabe registrar que nessa revisão não houve mudanças no plano amostral, tendo sido incorporado o tratamento da não resposta no procedimento de estimação e passado a adotar o SUDAAN no processo de produção dos resultados na obtenção das estimativas e correspondentes medidas de precisão (estimativas do erro padrão).

Pesquisa de Orçamentos Familiares (POF)

O objetivo principal da pesquisa é obter informações sobre os padrões de gastos das famílias para a construção dos novos pesos para os itens que compõem os diversos índices de preços que o IBGE calcula a cada mês. A primeira pesquisa sobre despesas familiares e renda no Brasil foi o ENDEF - Estudo Nacional da Despesa Familiar, realizada em 1974/75, que investigou também aspectos relativos à nutrição e antropometria, com plano amostral nos moldes do da PNAD. A Pesquisa de Orçamentos Familiares realizada em 1995/1996 foi planejada para representar a população urbana de 9 regiões metropolitanas, do Distrito Federal e do município de Goiânia. Essa pesquisa foi basicamente uma repetição da pesquisa realizada em 1987/1988, com alguma atualização metodológica.

Foram usados cinco tipos de questionários a saber: questionário do domicílio, incluindo características do domicílio e características demográficas dos moradores do domicílio; questionário de despesa coletiva, para as despesas com o domicílio e com bens duráveis; caderneta de despesas coletivas diárias, para despesas com alimentos, material de limpeza e higiene pessoal; questionário de despesas individuais, para despesas pessoais com alimentação fora do domicílio, transporte, fumo, livros, cuidados com a saúde, educação, roupas e outras; e questionário de renda individual, para todos os tipos de renda, incluindo salário, aluguel, rendimentos de aplicações financeiras e outras, além de taxas e deduções tais como imposto de renda.. A coleta dos dados foi realizada no período de outubro/95 a setembro/96 de forma a captar os padrões sazonais de renda e despesas. As

⁵ Veja *Pesquisa Mensal de Emprego. Outubro 2001 - Outubro 2002*, que contém uma análise sobre o impacto das alterações metodológicas da pesquisa nos resultados.

despesas foram obtidas através de entrevistas, usando diversos períodos de referência tais como: semana, mês, trimestre e semestre. Porém, as despesas com itens menos prováveis de serem lembrados exatamente pelos entrevistados (pequenas compras, despesas com alimentação e outras despesas de uso coletivo) foram obtidas através do registro diário em uma caderneta, feito preferencialmente pelo próprio informante, durante 7 dias, no caso da POF95/96 e de 14 dias no caso da POF 87/88.

Cabe destacar vários aspectos inovadores que foram incorporados no planejamento da amostra da POF, já na sua primeira aplicação. A respeito ver IBGE (1992), sobre a metodologia da POF 87/88, e Bianchini e Vieira (1998), sobre a metodologia da POF 95/96.

Em cada área da pesquisa, o plano amostral considera uma amostra em dois estágios de seleção com estratificação da unidade primária. A unidade primária de seleção, o setor censitário, foi estratificado em duas etapas: estratos geográficos (núcleo e periferia) e renda média do chefe do domicílio no setor. Os setores foram selecionados sistematicamente com probabilidade proporcional ao tamanho (medido em número de domicílios particulares ocupados). Em cada setor da amostra, os domicílios foram selecionados através de amostragem aleatória simples sem reposição. Pela primeira vez, em pesquisas do IBGE, a seleção de domicílios a pesquisar em cada setor foi feita sem empregar amostragem sistemática, com o sorteio aleatório efetuado por computador.

Para a POF 95/96, o tamanho da amostra em cada área foi determinado a partir de uma precisão especificada para estimar a renda total do chefe ($CV=0,05$), com o número de domicílios a serem selecionados em cada setor da amostra fixado em 10, usando as informações investigadas no questionário básico do Censo Demográfico de 1991. A amostra de setores foi dividida em 4 subamostras, uma para cada trimestre de coleta. A alocação dos setores nas subamostras foi aleatória, preservando a estratificação adotada, de tal forma que todos os estratos estão representados em todos os trimestres.

Antes do início da coleta, em cada setor selecionado para a amostra, foi preparada uma listagem completa dos domicílios de forma a construir um cadastro atualizado para a seleção das unidades de segundo estágio. Entretanto, o número de domicílios selecionados em cada setor foi aumentado para 13 para compensar a seleção de domicílios vagos, fechados e possíveis recusas. Para setores com taxas de crescimento acima de certos níveis, o número de domicílios selecionados foi novamente aumentado, de acordo com patamares de crescimento, de forma a reduzir a variância dos pesos, mas o número máximo de domicílios selecionados por setor foi 28. Uma novidade em relação à tradição das demais pesquisas domiciliares foi a eliminação do requisito da autoponderação. A amostra da POF 96/97 ficou com 19.816 domicílios selecionados (veja tabela 2).

O procedimento de estimação usado nessa POF 96/97 é baseado no estimador de razão com calibração⁶ na população residente em domicílios particulares urbanos dada pela Contagem de População de 1996. Os pesos associados a cada domicílio da amostra foram

⁶ Calibração é o processo pelo qual os fatores de expansão para a amostra são determinados de forma a buscar a consistência das estimativas a partir da amostra com os totais conhecidos da população.

obtidos usando o estimador natural derivado do plano amostral, com tratamento para não-resposta, multiplicado por um fator que foi calculado independentemente para cada área da pesquisa. Esse fator é a razão entre a população residente em domicílios particulares permanentes urbanos dada pela Contagem de População de 1996 e a estimativa da população correspondente proveniente da amostra. A data de referência da Contagem de População foi 01.08.96, que é próxima da data de referência definida para a POF, 15.09.96. A data de referência da POF foi usada para a correção da inflação, tornando todos os valores das despesas e receitas a preços constantes. Para o município da capital de cada Região Metropolitana, exceto Belém, foi calculado um outro ajuste para calibrar a população do município, gerando um conjunto de pesos à parte para ser usado apenas na obtenção de estimativas no nível do município.

A precisão das estimativas amostrais foi medida pelo coeficiente de variação calculados para um conjunto de estimativas de total, usando o método do conglomerado primário. A publicação dos coeficientes de variação para variáveis categóricas foi feita usando a função generalizada de variância (generalized variance functions, Wolter, 1985). Ou seja, para cada domínio de publicação, foi ajustado um modelo de regressão para explicar os CVs das estimativas de total como função do valor das próprias estimativas. Os coeficientes do modelo de regressão ajustado são publicados juntamente com as estimativas da pesquisa, bem como os próprios CVs para uma seleção de estimativas relativas a variáveis contínuas relacionadas com valores de despesas e de rendimentos.

Na POF2002/2003, em coleta desde julho/2002, foram introduzidas várias inovações em relação às pesquisas anteriores, onde se destacam: a ampliação da abrangência geográfica que passou a ser de âmbito nacional, incluindo também a área rural; uma significativa ampliação temática, com a investigação de gastos e recebimentos monetários e não monetários, além do aprofundamento de questões relacionadas a condições de vida, ampliando a gama de aplicações de seus resultados; e a implementação da informatização das operações de campo com a utilização do palmtop na operação de Listagem e do notebook na entrevista. O plano amostral da POF 2002/2003 também passou por várias adaptações em relação ao das pesquisas anteriores. Foram definidos os níveis geográficos, segundo os quais seriam determinados os tamanhos de amostra e para os quais teria uma precisão fixa, a saber: a área urbana de cada UF e a área rural de cada grande Região.

O plano amostral continua sendo o de uma amostra em dois estágios de seleção com estratificação da unidade primária. Em função da abrangência nacional da pesquisa, foi introduzido um maior detalhamento na estratificação geográfica da pesquisa, para permitir a estimação, embora não controlada por medida de precisão, para outros níveis geográficos de interesse.

Considerou-se a estratificação geográfica dentro da parte urbana e rural de cada uma das UFs, separadamente, a saber: parte urbana do município da Capital; parte urbana

do resto da região metropolitana da UF, quando houver⁷; resto urbano da UF; e rural da UF. Então, em cada estrato geográfico fez-se a estratificação estatística pela variável anos médio de estudo dos chefes dos domicílios do setor censitário, usando as informações da Contagem de População de 1996⁸.

Para a POF 2002/2003, o número de setores na amostra foi determinado, independentemente, para cada nível geográfico controlado de estimação (urbano de cada UF e rural de cada Região), a partir do plano amostral especificado, com precisão fixa (CV) para estimar a variável “média de anos de estudo dos chefes dos domicílios”, usando as informações investigadas no questionário da Contagem de População de 1996⁹, e também com o número fixo de domicílios a ser entrevistado para cada setor da amostra.

Foram testados valores alternativos dos parâmetros fixados (CV e nº de domicílios por setor), para cada nível geográfico controlado de estimação. Em função da operacionalidade e do custo da coleta, os parâmetros escolhidos variam por UF e por Região. O quadro 1 apresenta os parâmetros fixados escolhidos para cada área geográfica, de acordo com a especificação do nível geográfico de estimação considerado.

A seleção dos setores foi sistemática com probabilidade proporcional ao número de domicílios particulares ocupados por setor, considerando como cadastro os setores da base geográfica do Censo 2000. Ajustes já definidos no nº de domicílios a selecionar por setor após a listagem, passando de 10 para 13 na área urbana, e na área rural: de 16 para 20 nas regiões Norte e Nordeste e de 20 para 25 nas regiões Sudeste, Sul e Centro-Oeste. Dessa forma, o número esperado de domicílios a serem selecionados para a POF 2002/2003 é de 57.010, sendo 44.200 na área urbana e 12.810 na área rural. Pesquisa em campo até junho/2003.

⁷ Foram consideradas apenas as 9 RMs, onde foram realizadas as POFs anteriores, a fim de garantir a comparabilidade temporal da pesquisa.

⁸ As informações do rendimento investigadas no Censo Demográfico de 2000 não estavam disponíveis na ocasião do planejamento da amostra da POF 2002/2003. A decisão para utilizar a variável “média de anos de estudo dos chefes dos domicílios”, para o dimensionamento da amostra foi baseada no fato de que as informações disponíveis mais recentes eram da Contagem de População de 1996, que não investigou rendimento, mas investigou os anos de estudo do chefe

⁹ Para tomar tal decisão foi realizado um estudo realizado com as informações do Censo Demográfico de 91, onde foram comparados cálculos dos tamanhos de amostra conglomerada estratificada, quando se utiliza como variável de desenho o “rendimento do chefe do domicílio” e “anos de estudo do chefe do domicílio”.

Quadro1 – Parâmetros fixados para o cálculo do tamanho da amostra da POF 2002/2003 e nível geográfico de estimação por área geográfica

<i>Área geográfica</i>	<i>Nº de entrevistas esperadas por setor</i>	<i>Nível de precisão fixado CV</i>	<i>Nível geográfico controlado de estimação</i>
UF das regiões Sudeste, Sul e Centro-Oeste	10	0,02	Urbano de cada UF
Roraima, Amazonas, Pará, e cada UF da Região Nordeste		0,03	
Acre, Rondônia , Amapá e Tocantins		0,04	
Regiões Norte e Nordeste	16	0,04	Rural de cada Grande Região
Regiões sudeste, Sul e Centro-Oeste	20	0,03	
<i>Nível geográfico controlado de estimação</i>	<i>Nº de entrevistas esperadas por setor</i>	<i>Nível de precisão fixado CV</i>	<i>Área geográfica</i>
Urbano de cada UF	10	0,02	UF das regiões Sudeste, Sul e Centro-Oeste
		0,03	Roraima, Amazonas, Pará, e cada UF da Região Nordeste
		0,04	Acre, Rondônia , Amapá e Tocantins
Rural de cada Grande Região	16	0,04	Regiões Norte e Nordeste
	20	0,03	Regiões Sudeste, Sul e Centro-Oeste

Pesquisa sobre Padrões de Vida (PPV)

A Pesquisa sobre Padrões de Vida realizada em 1996/1997 foi uma pesquisa piloto baseada no Living Standard Measurement Study (LSMS) estabelecido pelo Banco Mundial em 1980 para desenvolver métodos para a coleta e análise de dados sobre padrões de vida em países em desenvolvimento. As pesquisas do tipo LSMS foram realizadas em vários países, com vários objetivos analíticos, tais como medir a distribuição do bem estar e o nível de pobreza dos domicílios, para entender como os domicílios reagem a programas econômicos e ambientais de governo, e permitir análise complexas da relação entre vários aspectos do bem estar do domicílio. (veja Grosh e Muñoz, 1996 e Caillaux, 1998).

Essa pesquisa caracteriza-se por: “inclusão de temas socioeconômicos, estudados de forma integrada em um mesmo domicílio, um ano de permanência no campo (março de

1996 a março de 1997) de forma a captar fenômenos sazonais e, manutenção de um controle rigoroso tanto na aplicação do questionário como na entrada de dados e no processo de crítica das informações” (Caillaux, 1998).

Para atingir esses objetivos, foram considerados vários aspectos inovadores no planejamento da pesquisa. Os domicílios foram entrevistados duas vezes, com intervalo de duas semanas entre as entrevistas. O questionário foi quase todo pré-codificado de forma que os dados puderam ser digitados diretamente em microcomputadores imediatamente após a entrevista, de forma descentralizada no campo. O programa de entrada de dados realiza críticas de consistência e de validade das respostas. Os dados inconsistentes ou errados eram marcados pelo programa de tal forma que as respostas de certas questões puderam ser verificadas durante a segunda entrevista. Os informantes foram instruídos para registrarem suas despesas durante as duas semanas que antecederam a segunda visita.

A pesquisa foi planejada para investigar uma diversidade de temas sociais e econômicos, a saber: características do domicílio, características básicas demográficas, migração, saúde, mão-de-obra, fecundidade, rendimentos, investimentos e créditos, despesas com bens duráveis, despesas com alimentação, empreendimentos domiciliares, agricultura, avaliação do padrão de vida e antropometria.

O plano amostral inclui estratificação dos setores censitários (as UPAs - unidades primárias de amostragem) em 10 estratos geográficos: de Fortaleza, Recife, Salvador, o restante da área urbana da Região Nordeste, o restante da área rural da região Nordeste, as regiões metropolitanas de São Paulo, Rio de Janeiro, Belo Horizonte, o restante da área urbana da Região Sudeste, o restante da área rural da região Sudeste. Em cada estrato geográfico, foram definidos outros três estratos com base no rendimento médio domiciliar mensal do chefe do domicílio, usando os dados do Censo Demográfico de 1991. O tamanho da amostra em número de setores em cada estrato foi determinado por alocação proporcional. Os setores foram selecionados com probabilidade proporcional ao número de domicílios, com reposição, em cada um dos trinta estratos definidos. A amostra de setores foi dividida em quatro subamostras, uma para cada trimestre de coleta. A alocação da amostra nas subamostras foi aleatória e de tal forma a preservar a estratificação, ou seja, todos os estratos estão representados em todos os trimestres. Em cada setor selecionado para a amostra, foi realizada uma operação denominada listagem de domicílios, com o objetivo de obter um cadastro atualizado para a seleção das unidades de segundo estágio. A partir dessa listagem, os domicílios foram selecionados através de amostragem aleatória simples sem reposição. A operação de listagem foi realizada por partes, o mais próximo possível do início da coleta de cada trimestre. Nos setores pertencentes aos estratos de região metropolitana ou de área urbana, foram selecionados 8 domicílios por setor. Nos estratos de setores rurais esse número foi fixado em 16.

O tamanho da amostra foi de 554 setores, 278 na região Nordeste e 276 na região Sudeste, correspondendo a um tamanho total de amostra de 4 944 domicílios. Com o objetivo de compensar a não-resposta por motivos de recusa, domicílios vagos ou fechados,

foi selecionada uma segunda amostra de domicílios nos mesmos moldes da primeira. Os domicílios com não-resposta foram então substituídos por domicílios provenientes dessa segunda amostra. Esse procedimento foi adotado em função do reduzido tamanho amostra, que tornava indesejável perder informação. Apesar disso, foram perdidos quatro questionários do primeiro trimestre de pesquisa. Essa perda foi compensada por reponderação dos domicílios informantes pelo inverso da taxa de resposta. Com o objetivo de evitar um aumento na taxa de não entrevistas devidas à recusa dos informantes, no processo de seleção da amostra de setores foram adotados procedimentos para evitar a seleção de setores já selecionados para a PNAD, a PME e a POF, uma vez que essas pesquisas foram realizadas na mesma época da PPV e possuem planos amostrais semelhantes.

A estimação de totais e dos erros amostrais associados foi realizada usando o estimador natural derivado do plano amostral adotado. A precisão das estimativas foi medida pelo coeficiente de variação calculado para todos os indicadores usados para a análise e disseminação dos resultados, usando o método do conglomerado primário (veja Albieri e Bianchini, 1997).

Pesquisa de Economia Informal Urbana (ECINF)

Em 1997, o IBGE realizou sua primeira pesquisa domiciliar por amostragem em nível nacional para identificar atividades econômicas desenvolvidas nos domicílios ou em pequenas unidades produtivas, de forma a medir o papel e a dimensão dessas atividades na economia brasileira, através da identificação dos proprietários de negócios informais e da investigação das características de funcionamento das unidades produtivas (ver Jorge, 1995).

A população objetivo inclui as pessoas residentes na área urbana que trabalhavam por conta própria ou como empregadores com até cinco empregados, em pelo menos uma situação de trabalho de atividades não-agrícolas. Os trabalhadores domésticos foram excluídos da população objetivo.

O objetivo da pesquisa foi produzir estimativas para cada um dos 26 estados, o Distrito Federal, cada uma das 10 regiões metropolitanas e o município de Goiânia.

Foram usados dois tipos de questionários na coleta dos dados: o primeiro para obter informação sobre as características dos domicílios e das pessoas moradoras, com o objetivo de identificar as pessoas engajadas em unidades produtivas do setor informal, através das características do trabalho; o segundo questionário foi usado para investigar as características das unidades produtivas dos setores informal e seus proprietários.

O plano amostral considera dois estágios de seleção e estratificação das unidades de seleção. As unidades primárias de amostragem (UPAs) foram os setores urbanos estratificados primeiramente pela localização geográfica. Em cada estado, foram definidos

dois ou três estratos geográficos dependendo se o estado possui ou não região metropolitana, o município da capital do estado, os demais municípios da região metropolitana, o restante do estado. Foi definido um segundo nível de estratificação das UPAs, dentro de cada estrato geográfico, de acordo com a renda média domiciliar do setor, obtida dos dados do Censo Demográfico de 1991. A seleção das unidades primárias em cada estrato foi feita através de amostragem sistemática com probabilidade proporcional ao tamanho (medida em número de domicílios ocupados). As unidades de segundo estágio foram os domicílios com moradores classificados como conta própria ou empregador com até cinco empregados. Os domicílios foram posteriormente classificados de acordo com o grupo de atividades e selecionados sistematicamente a partir da lista atualizada de domicílios realizada nos setores selecionados para a amostra.

Com o objetivo de evitar um aumento na taxa de não entrevistas devidas à recusa, no processo de seleção dos setores foi adotado um procedimento para eliminar as coincidências com os setores selecionados para a PNAD e PME, uma vez que essas duas pesquisas seriam realizadas na mesma época que a de economia informal.

O tamanho da amostra em cada área foi determinado a partir de um coeficiente de variação especificado para estimar o número de proprietários (conta próprias e empregadores) na economia informal (CV=5%, excepcionalmente, por motivos de custo, CV=6% para cada área da Região Norte). O número de domicílios a serem selecionados por setor foi fixado em 16. O tamanho da amostra é apresentado na tabela 2. (Veja Almeida e Bianchini, 1998).

No que se refere a aspectos de planejamento amostral, a pesquisa de economia informal difere das pesquisas domiciliares tradicionais do IBGE. Isto porque é preciso lidar com uma população rara, heterogênea e mais difícil de ser detectada. Todos esses fatores contribuem para aumentar a complexidade do plano amostral, da seleção da amostra, dos procedimentos de estimação e principalmente da preparação do cadastro de unidades amostrais de interesse, ou seja, da listagem de domicílios, a qual requer a realização de uma pesquisa em todos os domicílios do setor para identificar as atividades desenvolvidas pelos moradores de cada domicílio. (Veja Kalton e Anderson, 1986).

A listagem dos domicílios da ECINF foi uma operação com custo elevado, pois além de produzir uma lista completa de endereços das unidades domiciliares, envolveu a realização de entrevista para obter as informações necessárias para identificar a população objetivo, para obter as informações para a segunda estratificação por grupo de atividades que constituíram o objetivo da pesquisa. Os oito grupos de atividades considerados foram: (1) indústria da transformação e extrativa mineral; (2) indústria da construção; (3) comércio de mercadorias; (4) serviços de alojamento e alimentação; (5) serviços de transporte; (6) serviços de reparação, pessoais, domiciliares e de diversões; (7) serviços técnicos e auxiliares; (8) outros serviços.

Em cada setor da amostra, a alocação dos 16 domicílios foi feita proporcionalmente ao número de domicílios existentes em cada grupo de atividades no setor. Além disso,

foram feitos alguns ajustes que ocasionaram um aumento médio de 30% no número de domicílios a serem selecionados por setor.

A estimação foi feita com base no estimador natural que deriva do plano amostral adotado, com tratamento para não-resposta. Os domicílios listados como tendo proprietários do setor informal e que na entrevista não tinham mais essa característica foram excluídos na estimação de totais, mas receberam o valor zero para cada variável de interesse da investigação. Além dos pesos para estimação das características de proprietários de unidades produtivas do setor informal, foi também associado um peso à unidade produtiva do setor informal, que leva em conta o inverso do número de sócios da unidade produtiva. A precisão das estimativas foi medida pelos coeficientes de variação calculados para um conjunto selecionado de estimativas de total. O método do conglomerado primário foi usado para estimar a variância de cada variável.

A Amostra para a Coleta de Dados do Questionário Detalhado do Censo Demográfico

O Censo Demográfico de 1960 foi o primeiro a utilizar amostragem na coleta de dados relativos a um conjunto selecionado de características de pessoas, famílias e domicílios. Foram usados dois tipos de questionários: um questionário pequeno aplicado a todos os domicílios e seus moradores, não selecionados para a amostra (chamado questionário básico); e um questionário longo (chamado questionário da amostra) aplicado a todos os domicílios selecionados para a amostra, bem como seus moradores. Nos Censos de 1960, 1970 e 1980, foi utilizada uma única fração amostral de 25% dos domicílios. Em 1991, uma revisão amostral com grande impacto foi o emprego de duas frações amostrais diferentes de acordo com o tamanho do município, medido em função da projeção de população¹⁰ para a data de referência do Censo de 1992: 20% para os municípios com até 15.000 habitantes e 10% para os demais municípios. Em 2000, a definição das frações amostrais seguiu o mesmo critério de 1991. Assim, nos municípios com população projetada para a data de referência do Censo 2000 (1º de agosto de 2000) até 15.000 habitantes foi usada a fração amostral de 20%; para os demais municípios foi usada a fração de 10%.

O território nacional foi dividido em partições geográficas denominadas setores censitários, de tal forma que seus limites respeitam as divisões internas dos municípios em zonas urbanas e rurais, e em distritos e subdistritos, caso existam. O setor censitário foi planejado de forma a que um entrevistador consiga realizar a operação de coleta no período de realização do Censo. Em 2000, foram definidos 215.811 setores censitários (que, no plano amostral, foram considerados como estratos). Os domicílios particulares foram selecionados por amostragem sistemática em cada setor censitário. As famílias ou pessoas só moradoras em domicílios coletivos (alojamentos estudantis, quartéis, prisões, hospitais, orfanatos, conventos, etc.) foram selecionadas, também de forma sistemática,

¹⁰ As estimativas independentes de população são preparadas pelo próprio IBGE anualmente, para diversos níveis geográficos. O menor nível geográfico é o município.

independentemente da seleção de domicílios particulares, usando a mesma fração amostral definida para o setor a que pertence cada domicílio coletivo.

O procedimento de estimação de totais é aplicado em cada área de ponderação separadamente. Uma área de ponderação é uma unidade geográfica formada por um agrupamento mutuamente exclusivo de setores censitários, para a aplicação dos procedimentos de calibração das estimativas com as informações conhecidas para a população como um todo. Em 1991, o método de Mínimos Quadrados Generalizados em duas etapas (*Generalized Least Squares Estimation Procedure - GLSEP*, Bankier, Rathwell e Majkowski, 1992), denominado no IBGE por MQG2, foi usado na determinação dos pesos e as variáveis auxiliares utilizadas foram definidas dentre aquelas investigadas para 100% da população, no próprio Censo Demográfico. Esse procedimento de estimação de regressão atribui um único peso fracionário a cada domicílio e a cada um de seus moradores, sendo importante destacar essas duas situações novas em relação aos censos anteriores: o peso fracionário e único para domicílios, famílias e pessoas. Em 2000, o procedimento de estimação foi ligeiramente modificado para realizar apenas um ajuste o modelo de regressão em apenas uma etapa, mas manteve as mesmas características do adotado em 1991. A principal diferença em relação a 1991 reside na definição das áreas de ponderação, que em 2000 passou a considerar o critério de contiguidade entre os setores componentes de uma mesma área, além de definir um tamanho mínimo para essa área, em termos do número de domicílios particulares na amostra. Em alguns casos (municípios grandes para os quais se dispunha da malha digitalizada de setores censitários) a definição das áreas de ponderação fez uso de um sistema de informações georreferenciadas e de uma metodologia para agrupamento de setores com restrição de contiguidade e de busca de homogeneidade em relação a um conjunto de características conhecidas a 100%, com dados do próprio censo.

Resumo e Conclusões

As tabelas a seguir apresentam de forma resumida algumas características gerais dos planos amostrais, os tamanhos das amostras, as distribuições da amostra por tipo de entrevista e as taxas de resposta para as várias pesquisas domiciliares, excetuando o Censo Demográfico. No caso da Pesquisa Mensal de Emprego - PME, os dados referem-se à pesquisa antiga, pois a nova ainda estava em fase de implantação.

Tabela 1 - Características gerais dos planos amostrais das várias pesquisas
(continua)

Pesquisa	Abrangência Geográfica	Estágios de seleção	Tipo de estratificação das UPAs
PNAD	Nacional (exceto o Norte rural)	3 (municípios, setores e domicílios)	Geográfica
PME	6 regiões metropolitanas	2 (setores e domicílios)	Geográfica
POF	11 áreas urbanas, em 1987/88 e 1995/96; Nacional em 2002/03	2 (setores e domicílios)	Geográfica e classes de renda do chefe, em 1987/88 e 1995/96; geográfica e classes de anos de estudo do chefe, em 2002/03
PPV	Regiões Nordeste e Sudeste	2 (setores e domicílios)	Geográfica e classes de renda do chefe
ECINF	Nacional, somente áreas urbanas	2 (setores e domicílios)	Geográfica e classes de renda domiciliar

Tabela 1 - Características gerais dos planos amostrais das várias pesquisas
(continuação)

Pesquisa	Auto-ponderação	Seleção de setores	Variável usada como medida de tamanho	Cadastro de seleção de domicílios	Seleção de domicílios
PNAD	sim	sistemática com ppt	número de domicílios	Listagem + Novas construções	sistemática simples
PME	sim	sistemática com ppt	número de domicílios	Listagem + Novas construções	sistemática simples
POF	não	sistemática com ppt	número de domicílios particulares ocupados	Listagem	aleatória simples sem reposição
PPV	não	ppt com reposição	número de domicílios particulares	Listagem	aleatória simples sem reposição
ECINF	não	sistemática com ppt	número de domicílios ocupados	Listagem com entrevista para identificar população objetivo	seleção sistemática

Tabela 1 - Características gerais dos planos amostrais das várias pesquisas
(conclusão)

<i>Pesquisa</i>	<i>Periodicidade</i>	<i>Período de coleta</i>	<i>Aspectos longitudinais</i>
PNAD	anual (desde 1967, exceto em anos de Censo Demográfico)	3 meses	nova seleção da amostra de domicílios (3º estágio)
PME	mensal (desde 1980)	um mês	rotação da amostra de setores e domicílios
POF	1974/75 (ENDEF) 1987/88 1995/96 2002/03	um ano	nova seleção da amostra a cada execução da pesquisa
PPV	1996/97 (pesquisa piloto)	um ano	---
ECINF	1997 (primeira pesquisa)	3 meses	---

Tabela 2 – Tamanho da amostra e distribuição por tipo de entrevista das várias pesquisas

<i>Pesquisa</i>	<i>Número de</i>		<i>Número de domicílios na amostra</i>		
	<i>Setores selecionados</i>	<i>Selecionados</i>	<i>Média por setor</i>	<i>Eleitos</i> (^{**})	<i>Entrevistados</i>
PNAD 1997	6 678	109 541	16,4	91 811	90 006
PNAD 2001	7 273	126 858	17,4	105 657	103 483
PME Abril/98	1 510	40 090	26,5	32 549	30 951
Agosto/2002	1 510	40 348	26,8	31 926	30 624
POF 95/96	1 456	19 816	13,6	17 628	16 014
POF 2002/03	3 984	57 010 ^(***)	14,3 ^(***)		44 248 ^(***)
PPV 96/97 ^(*)	554	4 944	8 (urbanos) e 16 (rurais)	4 944	4 940
ECINF 97	2 340	48 934	20,9	38 099	37 010

^(*) Foi adotado o procedimento de substituição controlada de domicílios durante a coleta.

^(**) Os domicílios eleitos foram definidos de acordo com a população alvo de cada pesquisa, e considera o total de entrevistas realizadas mais as entrevistas não realizadas do tipo fechada ou recusa.

^(***) Número esperado de domicílios.

Tabela 3 - Taxas de resposta das várias pesquisas

<i>Pesquisa</i>	<i>Taxas</i>		
	<i>Entrevistados / selecionados (%)</i>	<i>Eleitos / selecionados (%)</i>	<i>Entrevistados / eleitos (%)</i>
PNAD 97	82,2	83,8	98,0
PNAD 2001	81,6	83,3	98,0
PME Abril/98	77,2	81,2	95,1
PME Agosto/2002	75,9	79,1	95,9
POF 95/96	80,8	89,0	90,8
POF 2002/03	em coleta	em coleta	em coleta
PPV 96/97 ^(*)	99,9	100,0	99,9
ECINF 97	75,6	77,9	97,1

^(*) Foi adotado o procedimento de substituição controlada de domicílios durante a coleta.

Tabela 4 - Procedimentos de estimação das pesquisas por amostra

<i>Pesquisa</i>	<i>Estimador</i>	<i>Variável de calibração</i>	<i>Tratamento de não- resposta</i>	<i>Publicação de erros amostrais</i>
PNAD 2001	razão	Projeção de População	não	ajuste de modelo de regressão para variáveis categóricas
PME Abril/98 PME Agosto/2002	razão	Projeção de População	não	calculados para um conjunto de estimativas
POF 95/96	razão	Contagem de População de 1996	sim	seleção de variáveis contínuas e ajuste para variáveis categóricas
PPV 96/97	natural do desenho	Não	sim	calculados para todos os indicadores divulgados
ECINF 97	natural do desenho	Não	sim	seleção de variáveis

As taxas de resposta representam a qualidade das pesquisas, pelo menos sob a ótica da aceitação por parte dos informantes, e são consideradas satisfatórias, em função da complexidade de cada pesquisa.

O IBGE vem realizando várias pesquisas domiciliares com diferentes níveis de complexidade em função das diferenças nos objetivos, na população-alvo, na abrangência

geográfica, no nível de precisão e na periodicidade definidas para cada uma. Apesar das muitas semelhanças entre os planos amostrais das pesquisas citadas, as diferenças apontadas também revelam mudanças e aperfeiçoamentos metodológicos que foram possíveis de serem implantados ao tempo em que cada pesquisa foi planejada ou realizada. Essas modificações (aperfeiçoamentos) incluem a incorporação de estratificação de acordo com o nível de renda do setor, ou na ausência dessa informação, a utilização de uma proxy, como foi o caso da POF 2002/03, que usou os anos de estudo do chefe para a estratificação, a previsão de perda de unidades da amostra por não-resposta, o tratamento da não-resposta, a redução do número de domicílios selecionados por setor, a redução do número de estágios de seleção e a eliminação do requisito de autoponderação.

Essa revisão serviu de base para as discussões que vêm sendo realizadas com vistas à reformulação das pesquisas domiciliares sobre mercado de trabalho realizadas pelo IBGE, a anual e a mensal. O objetivo é, em um futuro próximo, o ano de 2005, por em campo somente uma pesquisa de emprego, contínua, com abrangência nacional, para dar resultados tanto para os indicadores de conjuntura de emprego como para medidas do tipo estrutural sobre o mercado de trabalho.

A reformulação dos planos amostrais das pesquisas de forma a reduzir custos através de amostras menores conjugada com novos e aperfeiçoados procedimentos de estimação colocam-se como um desafio para o IBGE.

Como aperfeiçoamentos possíveis, vale destacar: a incorporação de efeitos espaciais ou de vizinhança na estratificação dos setores¹¹; a consolidação de tratamentos para a não-resposta total; a introdução de estimadores que se beneficiem da estrutura longitudinal da amostra, no caso da PME, por exemplo; o uso de sistemas genéricos para estimação de dados de pesquisas com planos amostrais complexos, que proporcionam economia de tempo e qualidade, com facilidades para o cálculo de erros amostrais também. O SUDAAN deixou de ser usado apenas nos estudos internos para fazer parte do processo de produção dos resultados, na obtenção das estimativas e correspondentes medidas de precisão (estimativas do erro padrão), no caso da PME.

Outros estudos também foram realizados e podem ser usados para melhorar as metodologias de planejamento, operação e análise de resultados das pesquisas, a saber: ajustes de modelos hierárquicos de dados de pesquisas por amostragem usando corretamente os fatores de expansão da amostra e as características do desenho amostral (Corrêa, 2001); estudos sobre a estimação de variâncias para séries dessazonalizadas pelo método X-12-Arima, considerando o desenho amostral (Cruz, 2002); estimação de variância para o estimador de diferença entre duas taxas da Pesquisa Mensal de Emprego (Corrêa e Silva, 2001); estudos sobre estratificação considerando o componente espacial para o planejamento das amostras (Freitas, 2002).

¹¹ Estudos sobre estratificação considerando o componente espacial para o planejamento de amostras como a PME estão descritos em Freitas, 2002, uma dissertação de mestrado desenvolvida na Escola Nacional de Ciências Estatísticas do IBGE.

Referências bibliográficas

ALBIERI, S. e BIANCHINI, Z.M. *Aspectos de amostragem relativos à pesquisa domiciliar sobre padrões de vida*. Rio de Janeiro: IBGE, Departamento de Metodologia, 1997. 14p.

ALMEIDA, R.A.P.; BIANCHINI, Z.M. *Sampling aspects of the 1997 Brazilian survey of the urban informal sector*. Proceedings of the Joint IASS/IAOS Conference. 1998.

ALMEIDA, R.A.P.; BIANCHINI, Z.M. *Aspectos de amostragem da Pesquisa de Economia Informal Urbana 97*. Rio de Janeiro: IBGE, 1998. 32p. (Texto para Discussão, nº 89).

BANKIER, M.D.; RATHWELL, S.; MAJKOWSKI, M. *Two step generalized least squares estimations in 1991 Canadian Census*. Ottawa: Statistics Canada, Methodology Branch Working Paper. 1992.

BIANCHINI, Z.M. e ALBIERI, S. Uma Revisão dos Principais Aspectos dos Planos Amostrais das Pesquisas Domiciliares Realizadas pelo IBGE, *Revista Brasileira de Estatística*, v.60, n. 213, p.7-23, jan./jun. 1999.

BIANCHINI, Z.M. e ALBIERI, S. Principales Aspectos de Muestreo de las Encuestas de Hogares del IBGE - revisión 2002. 17p. 2002. [Artigo apresentado no MECOVI - 10° Taller Regional sobre La Practica del Muestreo para el Diseño de las Encuestas de Hogares, 27 a 29 de noviembre, Buenos Aires, Argentina].

BIANCHINI, Z.M.; VIEIRA, M. *Aspectos de amostragem da Pesquisa de Orçamentos Familiares 1995-1996*. Rio de Janeiro: IBGE, 1998. 106p. (Texto para Discussão, nº 93).

CAILLAUX, E.L. *Living standard survey 1996-1997*. Rio de Janeiro: IBGE. 1998. 10p. [Apresentado no Meeting of the Expert Group on Poverty Statistics (Rio Group), 13-15 Maio, 1998].

COCHRAN, W.G. *Sampling Techniques* (3rd ed.). New York : Wiley. 1977.

CORRÊA, S.T. *Modelos Lineares Hierárquicos em Pesquisas por Amostragem Relacionando o Índice de Massa Corporal às Variáveis da Pesquisa sobre Padrões de Vida do IBGE*. Rio de Janeiro: IBGE/ENCE. 2001. [Dissertação de Mestrado].

CORRÊA, S.T.; SILVA, P.L.N. *Estimação de Variância para o Estimador da Diferença entre duas Taxas na Pesquisa Mensal de Emprego*. Rio de Janeiro: IBGE, Departamento de Metodologia, 2001. 26p.

CRUZ, M.M. *Estimação de variâncias para séries dessazonalizadas pelo método X-12-Arima, considerando o desenho amostral*. Rio de Janeiro: IBGE/ENCE. 2002. [Dissertação de Mestrado].

FREITAS, M.P.S. *Estratificação para a amostra de uma pesquisa domiciliar sobre mercado de trabalho*. Rio de Janeiro: IBGE/ENCE. 2002. [Dissertação de Mestrado].

GROSH, M. E.; MUÑOZ, J. *Manual for planning and implementing the LSMS Survey*. Poverty e Human Resources Division, Policy Research Department, The World Bank. 1996.

HANSEN, M.H., HURWITZ, W.N. e MADOW, W.G. *Sample Survey Methods and Theory*, Vol. I e II. New York: Wiley. 1953.

JORGE, A. The survey of the urban informal economy in Brazil. *Proceedings of the International Seminar on Informal Sector Employment Statistics*, 1995. pp. 239-256.

JORGE, A.F. *Pesquisa de Economia Informal Urbana*. Rio de Janeiro: IBGE, 1996. 17p. [Artigo apresentado no Encontro Nacional de Produtores e Usuários de Informações Sociais, Econômicas e Territoriais].

KALTON, G. e ANDERSON, D.W. Sampling rare populations. *The Journal of the Royal Statistical Society A*, 149, part 1, pp 65-82. 1986.

KISH, L. *Survey sampling*. New York: Wiley. 1965

METODOLOGIA da Pesquisa Nacional por Amostra de Domicílios na década de 80. Rio de Janeiro: IBGE (Série Relatórios Metodológicos, vol.1). 1983.

METODOLOGIA da Pesquisa Mensal de Emprego 1980. Rio de Janeiro: IBGE. (Série Relatórios Metodológicos, vol.2). 1983.

PESQUISA Mensal de Emprego. Rio de Janeiro: IBGE (Série Relatórios Metodológicos, vol.23). 72p. 2002.

PESQUISA Mensal de Emprego. Outubro 2001-Outubro 2002. Rio de Janeiro: IBGE, Departamento de Emprego e Rendimento, 78p. 2002.

PESQUISA Orçamentos Familiares. Volume 3. Aspectos de Amostragem. Rio de Janeiro: IBGE, 218p. (Série Relatórios Metodológicos, vol. 10). 1992.

PESSOA, D.G.C.; SILVA, P.L.N. *Análise de dados amostrais complexos*. São Paulo: Associação Brasileira de Estatística, 1998.

SILVA, P.L.N. *Planejamento, estimação e análise de dados em pesquisas por amostragem: desvendando a realidade brasileira com o "telescópio da estatística"*. Rio de Janeiro: IBGE, 1996. 28p. [Artigo apresentado no Encontro Nacional de Produtores e Usuários de Informações Sociais, Econômicas e Territoriais].

WOLTER, K. M. *Introduction to Variance Estimation*. New York: Springer-Verlag. 1985