

8PE

**PESQUISAS POR AMOSTRA DE
DOMICÍLIOS**

A PERDA DA PONDERAÇÃO DA AMOSTRA

Luiz Nery da Costa

Rio de Janeiro, julho de 1995

J. M. G. A.
Programa 1995

314.6
C837c

SET

537/96

Í N D I C E

PAG.

1.	INTRODUÇÃO	1
2.	CONSIDERAÇÕES INICIAIS	1
2.1	Países que utilizam pesquisas domiciliares	1
2.2	Modelos de Amostragem usados	1
2.3	Marcos de Amostragem	2
2.4	Manutenção e atualização dos Marcos de Amostragem	2
2.4.1	Projetos Novas Construções	3
2.4.2	Contagem Rápida	3
2.5	Desenhos das Amostras	3
3.	FATORES QUE CONTRIBUEM PARA A PERDA DE PONDERAÇÃO DA AMOSTRA	5
3.1	Opção pelas medidas de tamanho usadas nas probabilidades de seleção	6
3.2	Problemas operacionais de coleta ou controle deficiente	11
4.	A IMPORTÂNCIA DOS PROJETOS NOVAS CONSTRUÇÕES (PNC)	12
4.1	Considerações Gerais	12
4.2	Estimativas de totais de domicílios usando ou não os PNC	13
4.3	Estimativas da variabilidade da expansão dos setores normais da PNAD	17
4.4	Distribuição dos tamanhos dos PNC segundo os Municípios e Grande Vitória	19
5.	USO DO ESTIMADOR DE RAZÃO NA EXPANSÃO DOS RESULTADOS DAS PESQUISAS DOMICILIARES	20
6.	CONCLUSÕES	22
7	REFERÊNCIAS BIBLIOGRÁFICAS	23
	ANEXO 1	24

1. INTRODUÇÃO

As pesquisas por amostras probabilísticas de domicílios, geralmente, são planejadas a partir dos Marcos de Amostragem propiciados pelos Censos Demográficos (de população e de domicílios). Tais pesquisas são realizadas com periodicidade anual, ou em intervalos maiores de tempo, assim como são utilizadas para a obtenção de estimativas de características da população, de domicílios, etc., tanto para o conjunto do País, como para as Unidades da Federação, Regiões Metropolitanas, cidades, etc.

A concepção do desenho da amostra, inclusive com a definição das respectivas probabilidades de seleção, bem como a medida que a época de realização de cada pesquisa se distancia do ano de referência do Marco de Amostragem, acrescidas de dificuldades operacionais de coleta conduzem a vários fatores que contribuem para que as estimativas provenientes dos estimadores naturais apresentem maiores variabilidades, devido a perda de ponderação da amostra.

Este documento procura abordar o presente tema — **Perda de Ponderação da Amostra** —, apresentando comentários e sugestões para que os desenhos das amostras e a definição dos estimadores das estimativas sejam mais adequados à realidade da época da pesquisa. Esta abordagem é feita em função das observações procedidas nas pesquisas realizadas no Brasil, principalmente a Pesquisa Nacional por Amostra de Domicílios — PNAD, e a Pesquisa de Orçamentos familiares — POF-87, ambas desenvolvidas pelo IBGE.

2. CONSIDERAÇÕES INICIAIS

A seguir, são elaboradas algumas considerações para a melhor compreensão do tema.

2.1 Países que utilizam pesquisas domiciliares

Na América do Sul diversos países usam amostras de domicílios para pesquisarem o emprego, desemprego, características de domicílios, etc. Entre esses países citamos: Argentina, Bolívia, Brasil, Chile, Paraguai, Peru, Venezuela, Uruguai e, na Europa, pelo menos, a Espanha.

2.2 Modelos de amostragem usados

Os desenhos das amostras adotados são desenvolvidos em duas ou três etapas. Nos países com maiores dimensões territoriais e população são empregadas amostras em três estágios, como por exemplo: a PNAD, no Brasil, e a amostra, na Venezuela e na Espanha. Nestes casos, a opção pelo tipo de amostras em três estágios possivelmente ocorre por razões de custos da coleta. Tais desenhos das amostras poderiam ser em dois estágios, conseqüentemente, as amostras poderiam apresentar menores variabilidades, pois quanto maior o

número dos estágios, maiores são as variabilidades e componentes dos erros de amostragem.

Para cada estágio ou etapa da amostra corresponde uma unidade de pesquisa diferente, conseqüentemente populações distintas. Assim, nas amostras em três etapas há necessidade de se dispor de três populações: a população das unidades primárias, a população das unidades secundárias e a população das unidades terciárias ou finais, (domicílios), enquanto na amostra bietápica são definidas as populações de unidades primárias e das unidades finais.

2.3 Marcos de Amostragem

O universo formado por todas as unidades de amostragem constitui o Marco de Amostragem.(1)

Por ocasião dos Censos Demográficos são organizados e definidos as áreas ou níveis de detalhamento geográfico dos resultados do Censo. Nesta ocasião são concebidas e organizadas as populações de unidades de amostragem necessárias às pesquisas domiciliares por amostragem.

A concepção do marco de amostragem compreende: a finalidade a que se destina, a definição das unidades de amostragem, a cobertura ou abrangência geográfica; o sistema de informação disponível, inclusive para os cálculos das probabilidades de seleção em cada estágio, etc. e, por fim, a forma de atualização do marco. Em decorrência desta concepção diversos problemas operacionais surgem durante a vida útil do marco de amostragem.

2.4 Manutenção e atualização dos Marcos de Amostragem

Após a organização do Marco de Amostragem surgem várias modificações nas suas unidades de amostragem. Essas alterações crescem com o aumento do tempo decorrido da sua elaboração, ou seja, o envelhecimento do Marco.

Quando as unidades de amostragem são formadas por divisões político-administrativas, como por exemplo os Municípios, Distritos Administrativos, etc., as alterações na população dessas unidades são menos freqüentes e são acompanhadas através das legislações que regem as unidades administrativas.

Para as demais unidades de amostragem, o acompanhamento é mais difícil, trabalhoso e caro. Citamos, a seguir, dois tipos de acompanhamento.

2.4.1 Projeto de Novas Construções

O levantamento dos Projetos de Novas Construções — PNC — tem por finalidade complementar o cadastro de unidades domiciliares a fim de acompanhar e controlar o crescimento significativo do número de domicílios nos setores selecionados dos Municípios de Amostra. (2)

A PNAD define como PNC o conjunto de 30 ou mais domicílios que estejam em condições de habitação e edificados, após a construção do marco de amostragem, em todos os setores dos Municípios selecionados para a amostra. Esses projetos provêm de construções regulares, realizadas por pessoas físicas ou empresas de construção, e de construções irregulares que geralmente não atendem formalidades legais, por exemplo: loteamentos clandestinos, favelas, etc.

Para cada PNC, faz-se um mapa ou croqui da sua localização, identificando-se em que setor do marco de amostragem ele está situado. A localização e a quantidade de domicílios são confirmados no campo: se existem até 800 domicílios tem-se um só PNC, se mais de 800 domicílios tem-se tantos PNC quantas sejam as subáreas formadas por cerca de 100 domicílios.

Os projetos PNC, elaborados após a formação do marco de amostragem, dão origem a outra população de unidades de amostragem, consideradas, posteriormente, no desenho da amostra. Esta modalidade de atualização do marco, além do Brasil, também é adotada pela Oficina Central de Estatística e Informática da República da Venezuela.(3)

2.4.2 Contagem Rápida

Atualização do marco inicial da malha viária da parte urbana das unidades de amostragem é realizada contando-se os novos domicílios e quarteirões surgidos após o Censo. Algumas vezes, tais quarteirões são classificados de acordo com o tipo predominante das características externas dos domicílios, por exemplo: em nível A (domicílios destinados a classe alta), nível B (com domicílios destinados a classe média) e nível C (domicílios destinados a classe baixa ou popular). Essa forma de atualização é adotada pelo Instituto Nacional de Estatística e Informática dos Governos do Chile e do Peru.

2.5 Desenhos das Amostras

A amostra de PNAD, destinada à obtenção de estimativas em nível de Unidade da Federação, é trietápica.

Na primeira etapa são formados os estratos de municípios não auto representativos (menos populosos). Desses, dois municípios, são selecionados para a amostra com probabilidades proporcionais às respectivas populações, no marco, por estrato. Em seguida, são identificados os municípios auto-representativos (mais populosos) ou que pertençam às Regiões Metropolitanas

brasileiras.

As unidades secundárias de amostragem são formadas pelos setores censitários do marco, pertencentes aos municípios que compõem a amostra do primeiro estágio. Os setores são selecionados com probabilidade proporcional ao número de domicílios que possuam no marco.

As unidades de amostragem do terceiro estágio são constituídas pelos domicílios dos setores selecionados para a amostra. Neste estágio os domicílios são selecionados, em cada setor, com a equiprobabilidade.

A amostra da PNAD abrange os setores normais, formados por ocasião da definição do marco, e pelas unidades definidas pelo PNC, surgidas após a definição do marco.

O desenho da amostra da POF-87, destinada à obtenção de características de orçamentos familiares da população residente nas Regiões Metropolitanas (RM) do Brasil, possui dois estágios. Cada Município da RM é auto representativo e os setores censitários constituem as unidades primárias de amostragem, as quais são selecionadas com probabilidades proporcionais ao número de domicílios.

Os domicílios existentes em cada setor selecionado para a amostra do primeiro estágio constituem as unidades finais de amostragem, as quais são selecionadas com equiprobabilidade em cada setor.

A amostra de POF-87 não considerou a existência dos PNC por razões diversas, entre elas: as dificuldades operacionais de coleta na construção e manutenção dos PNC, a perda da autoponderação da amostra, a defasagem entre a época da coleta da POF e a da construção do marco de amostragem, etc.

Na amostra da POF, ao se abandonar os PNC, procede-se à listagem completa dos setores da amostra de modo a se comparar o tamanho do setor, na época da listagem, com o do marco de amostragem.

As razões indicadas anteriormente, e esta última solução, por si só não resolveriam o problema, pois são medidas paliativas à situação discutida. (4) e (5)

É oportuno salientar que os desenhos das amostras são autoponderados de modo que se tem:

$$f = P_h \cdot P_{hj/i} \cdot P_{hk/ij}$$

onde:

f - fração geral de amostragem

P_{hi} - probabilidade de seleção do município i do estrato h, (h = 1, 2L)

$P_{hj/i}$ - probabilidade de seleção do j - esimo setor, do i- esimo município do estrato h.

$P_{hk/ij}$ - probabilidade de seleção do k-esimo domicílio do j- esimo setor do i-esimo município do estrato h.

Nos municípios autorepresentados ou nos modelos de amostragem em dois estágios, tem-se:

$$f = 1 \cdot P_{hj/i} \cdot P_{hk/ij}, \text{ pois } P_{hi} = 1$$

3. FATORES QUE CONTRIBUEM PARA A PERDA DE PONDERAÇÃO DA AMOSTRA

As pesquisas domiciliares contínuas ou não, a serem levantadas após a realização do Censo Demográfico, (portanto após a construção do Marco de Amostragem), se defrontam com problemas decorrentes do envelhecimento do Marco de Amostragem em virtude da defasagem entre a época de execução da coleta da pesquisa e a época do levantamento do marco. Quanto maior a defasagem maior é o envelhecimento, conseqüentemente, maiores são as possibilidades de modificações nos diversos tamanhos das unidades de amostragem adotadas.

Além disso, problemas operacionais de coleta, principalmente quanto à atualização do marco, contribuem para a diminuição da qualidade da amostra. Tem-se como conseqüência maiores variabilidades e diminuição na representatividade da amostra.

Assim, algumas definições de medida de tamanho das unidades de amostragem, acrescidas das decisões do uso ou não dos PNC, e os elementos que contribuem para o envelhecimento do marco, aumento da variabilidade e diminuição da representatividade da amostra, definimos como fatores contribuidores da perda de ponderação da amostra.

Adiante, procede-se ao exame de alguns deles, surgidos por ocasião do planejamento das pesquisas, ou por ocasião das listagens — na realidade durante o intervalo entre a coleta de cada pesquisa e a construção do marco.

3.1 Opção pelas medidas de tamanho usadas nas probabilidades de seleção

Na PNAD de 1980, a unidade primária de amostragem é o município, sendo este selecionado com probabilidade proporcional a uma medida de tamanho. Definiu-se como unidade de medida a população do município. Uma outra unidade de medida de tamanho que poderia ser usada seria o total de domicílios.

Como a fração geral de amostragem resulta do produto das três probabilidades de seleção, este resultado seria mais exato se, no primeiro estágio, fosse utilizado o total de domicílios como unidade de medida de seleção, que também é usada como unidade de medida de seleção no segundo estágio.

Os exemplos citados a seguir quase sempre referem-se ao Estado do Espírito Santo e aos Municípios da Grande Vitória - Cariacica, Serra, Viana, Vila Velha e Vitória. A razão de escolha desse espaço geográfico é devida ao fato do autor deste documento vir realizando diversas pesquisas e estudos neste Estado e nesta região.

Tabela 1. População, domicílios particulares total e ocupados segundo os municípios do estrato 2

Municípios (a)	População (b)	Total de Domicílios Particulares (c)	Domicílios Ocupados (d)	Média (e) = (b)/(d)
Total	231.100	55.848	45.602	5,1
Barra de S. Francisco	51.611	12.434	10.498	4,9
Boa Esperança	11.137	2.524	2.081	5,4
Ecoporanga	31.382	8.457	6.283	5,0
Montanha(1)	17.449	4.217	3.376	5,2
Mucurici	11.543	2.934	2.354	4,9
Nova Venécia(1)	46.669	10.821	8.841	5,3
Pinheiros	20.001	5.004	3.968	5,0
S. Gabriel da Palha	42.308	9.457	8.201	5,2

Fonte: Sinopse preliminar do Censo-80.

(1) Municípios selecionados para a amostra.

Neste estrato foram selecionados os municípios de Montanha e Nova Venécia. As probabilidades de seleção de cada município, usando como medida de tamanho a população, são um pouco diferentes daquelas calculadas a partir do total de domicílios.

Por outro lado, as médias de pessoas por domicílio não são iguais nos municípios de cada estrato, resultando, então, em variações nas probabilidades de seleção do primeiro estágio, embora essas variações não sejam grandes.

Assim, a escolha pela população do município, como unidade de medida de seleção das unidades primárias, constitui o primeiro fator que contribui para a perda de ponderação da amostra.

A unidade de amostragem do segundo estágio é o setor. Na PNAD definiu-se, como unidade de medida de probabilidade de seleção do setor, o total de Domicílios Particulares (DT) que corresponde a soma dos diversos tipos de domicílios, no censo, isto é, o estoque de domicílios, como destacado a seguir:

$$DT = PO + PV + DUO + DF$$

onde:

DT - Total de domicílios

DO - Domicílios ocupados

DV - Domicílios vagos. Estão prontos mas não ocupados na época da coleta do Censo

DUO - Domicílios de uso ocasional, geralmente ocupados ocasionalmente, por pessoas não moradoras, (casas de veraneio, de férias, de lazer, etc.)

DF - Domicílios fechados. São domicílios ocupados que na época do Censo os seus moradores nele não se encontravam, por período superior a duração da coleta do setor.

Esta definição, como unidade de medida para os setores, parece ser a mais adequada pois a maioria dos domicílios vagos, após o censo, torna-se ocupada. Como as pesquisas domiciliares são realizadas, no mínimo um ano após o Censo, então esta unidade de medida se aproxima mais do total de domicílios existentes por ocasião da pesquisa domiciliária.

Na amostra da PNAD os municípios autorepresentados ou das Regiões Metropolitanas brasileiras, os municípios tem probabilidade 1 de inclusão na amostra. Os setores passam a constituir as unidades primárias de amostragem; na realidade eqüivalem às unidades secundárias, em se tratando de modelos de amostragem em 3 etapas. Neste caso, a definição da unidade de medida de tamanho da unidade primária nada afeta.

Tabela 2. Domicílios por tipo, segundo o Estado do Espírito Santo e Grande Vitória: 1980 e 1991

Especificação	Domicílios particulares				
	Total	Ocupados	Uso ocasional	Fechados	Vagos
Censo-80					
Espírito Santo	516.243	420.550	14.389	1.959	79.345
Grande Vitória	186.428	152.908	2.842	910	29.768
Censo-91					
Espírito Santo	755.914	620.705	37.815	11.219	86.175
Grande Vitória	306.561	261.572	8.626	8.320	28.043

Fonte: Sinopses dos Censos Demográficos de 1980 e 1991.

Na Tabela 2 nota-se a influência dos domicílios vagos no total de domicílios, bem como a pequena participação dos domicílios fechados, principalmente em 1980. Cabe registrar que o grande número de DF no CD-91 indica problemas de coleta.

Na POF-87 a medida de tamanho usada para a seleção dos setores foi o total de domicílios particulares permanentes ocupados e fechados, ou seja:

$$DT = DP + DF$$

Portanto, não inclui os domicílios vagos (DV) e os de uso ocasional (DUO). Os domicílios (DV + DUO) representam cerca de 12% e 11% do estoque de domicílios particulares existentes em 1991 nas Regiões Metropolitanas do Rio de Janeiro e de São Paulo.

Sem dúvida essa parcela de domicílios, não considerada pela POF, influi sensivelmente nas probabilidades de seleção dos setores. Por conseguinte, constitui um outro fator que contribui para a perda de ponderação da amostra.

Como o IBGE é responsável por ambas as pesquisas, PNAD e POF, cabe a pergunta: por que são utilizados critérios diferentes na unidade de medida das probabilidades dos setores? O autor deste documento supõe que a razão dessa diferença seja devido ao fato da POF estratificar os setores por faixas de renda, bem como porque nos domicílios (DV + DUO) não existem informações sobre o rendimento.

A fim de exemplificar a influência dos (DV + DF + DUO) no total de domicílios (DT), e no cálculo das probabilidades dos setores da Amostra da PNAD, para municípios na Grande Vitória, são apresentadas algumas observações.

Tabela 3. Total de domicílios particulares e ocupados no Censo de 80 e na PNAD-80, segundo os municípios

Municípios	Censo 1980		Amostra PNAD(1980)	
	DT	DO	DT	DO
Total:	186.428	152.908	9.942	8.073
Cariacica	46.554	40.221	2.663	2.346
Serra	27.430	17.500	1.474	723
Viana	6.104	4.918		
Vila Velha	53.326	44.183	2.808	2.463
Vitória	53.014	46.080	2.997	2.540

De acordo com os dados da **tabela 3**, nota-se que no Censo de 1980 e na amostra de 30 setores PNAD-80, cerca de 81% do total de domicílios são ocupados, portanto aproximadamente 19% dos DT representam os domicílios (DV + DUO + DF).

A influência dos (DV + DUO + DF) nas probabilidades dos setores é de cerca de 19%, em média. Cabe destacar que os domicílios DF de um modo geral são parcelas desprezíveis; na grande Vitória eles totalizavam 910 domicílios, ou seja, 0,5% do total.

Algumas informações contidas na **Tabela 1**, em anexo, — que contem a evolução durante 80/90 do número de domicílios, nos setores selecionados para a amostra da PNAD —, possibilita o exame da influência dos critérios da medida da probabilidade de seleção, já mencionados. Dos setores indicados destacamos alguns deles e, para eles, calculamos as respectivas probabilidades de seleção, de modo aproximado, considerando os dois critérios.

3.2 Problemas operacionais de coleta ou controle deficiente

Um outro aspecto que contribui para o aumento da variabilidade da perda de qualidade da amostra é o crescimento no número de domicílios dos setores após a construção do Marco Amostral.

Quando os setores selecionados para a amostra atingem o número de 800 domicílios, eles são subdivididos em subsetores, ou áreas menores, das quais são selecionados dois subsetores para a amostra, por exemplo: setor 193 de Vitória. Este artifício acarreta aumento na variabilidade, pois é mais uma etapa de seleção.

As regras do PNC estabelecem que todos os conglomerados de 30 ou mais novas habitações dão origem a um PNC. Entretanto, esta regra não é bem obedecida pois alguns setores da amostra da PNAD revelam este fato. Citamos como exemplo os setores 008 de Vila Velha e os setores 174 e 193 de Vitória que, em algum ano do período 80/86, cresceram mais de 30 domicílios. Também, os setores 111 de Cariacica, 014 de Serra, 033 e 154 de Vila Velha e, novamente o setor 174 de Vitória, apresentaram, após 1986, em um dos anos crescimentos superiores ao limite considerado. Todos os exemplos citados podem ser observados na Tabela 1, do anexo.

No último estágio da amostra, os domicílios são selecionados, em cada setor, obedecendo a fração geral da amostragem, portanto para cada setor é calculado um intervalo de seleção dos domicílios, com base no Marco Amostral.

Nos setores da amostra, da PNAD, onde ocorreram grandes acréscimos de domicílios, resulta um aumento considerável no número de domicílios selecionados para a amostra de domicílios. Esse é o caso, por exemplo, dos setores 111 de Cariacica, 033 de Vila Velha e 174 de Vitória, sendo que, neste último, foram selecionados cerca de 38 domicílios em 1990, os quais teriam representado 1/5 da amostra, em Vitória.

Considerando que as características de domicílios, famílias e pessoas são mais homogêneas em cada setor do que entre eles, isto pode ser confirmado pelo coeficiente de correlação intraclasse. Resulta, então, um acréscimo na homogeneização da amostra, contribuindo assim para a perda de qualidade da amostra.

Sem dúvida, o crescimento acentuado no número de domicílios dos setores da amostra, sem qualquer medida de controle, ou com falhas operacionais de coleta, gera mais um fator que contribui para a diminuição da qualidade da amostra.

4. A IMPORTÂNCIA DOS PROJETOS DE NOVAS CONSTRUÇÕES (PNC)

4.1 Considerações Gerais

A política habitacional brasileira, durante os decênios 1970/80 e 1980/90, promoveu e financiou milhões de novas habitações, a maioria em grandes conjuntos habitacionais destinados às populações de renda menor.

Geralmente esses conjuntos habitacionais foram e continuam sendo edificados em espaços geográficos de dimensões maiores do que os pequenos projetos de construção, 30 habitações.

Os grandes PNC, em número pequeno, geralmente estão localizados em setores censitários de baixa densidade habitacional, situados nas periferias das cidades, enquanto os de porte menor estão situados em locais mais próximos de centros das cidades.

Destacamos pelo menos dois pontos a considerar quanto às novas habitações, em relação à amostragem:

- a) Os PNC possibilitam a atualização do Marco de Amostragem, além disso, permitem controlar a variação dos tamanhos dos setores, reduzindo a variabilidade da amostra de setores.

Para se dar uma idéia da dimensão desses PNC citamos: no Município de Serra, nos setores 061, 066, 010 e 011 foram edificados respectivamente 4.924, 5.732, 3.310 e 3.080 domicílios; em Vila Velha setores 056 e 059 foram construídos 1.984 domicílios; em Cariacica, setor 014, foram construídos 3.979 domicílios; em Vitória, no setor 197 e 077 foram edificados 1.260 e 6437 novas habitações; isto na década de 1980.

- b) As habitações de cada PNC são muito parecidas, encerram grande homogeneidade. A maioria dos domicílios está situada em grandes conjuntos habitacionais destinados à população de menor poder aquisitivo.

O uso ou não dos PNC como parte integrante do desenho de amostras deve levar em conta os aspectos abordados anteriormente.

A PNAD utiliza os PNC tanto para atualizar o Marco de Amostragem e controlar as variações dos tamanhos dos setores, quanto para não deixar aumentar a homogeneidade dos domicílios na amostra; o que seria desejável.

A POF-87 não considerou as informações dos PNC devido a possíveis duplicações de novas habitações e deficiências operacionais de coleta. Procurou, então, efetuar a listagem completa dos setores selecionados de modo a se comparar o tamanho atual com o usado na seleção da amostra.

De fato, algumas falhas foram cometidas no controle dos tamanhos das listagens das PNAD da década de 1980, de modo que alguns setores da amostra cresceram substancialmente quando deveriam estar controlados, como foi citado anteriormente em 3.2.

Não é demais salientar o fato do PNC, localizado no setor 077 de Vitória, ter considerado até 1988 cerca de 630 novas habitações, quando na realidade existiam cerca de 4.200 domicílios, segundo a Pesquisa de Mercado Habitacional de Vitória, promovida pelo Instituto de Desenvolvimento Industrial do Espírito Santo (IDEIES) e planejada pelo autor deste documento.(6)

Vale dizer que a área localizada neste setor (077) como novas habitações é conhecida como Lixão (depósito de lixo), uma região de paupérrimos domicílios construídos em manguezais. Hoje, esta área está totalmente urbanizada.

Por outro lado, quando não se utiliza os PNC, como foi o caso da POF-87, a possibilidade dos setores selecionados para a amostra incluírem os grandes conglomerados (500 ou mais habitações novas) é pequena, pois eles são em número reduzido (veja-se a tabela 5, mais adiante. Como consequência, parte da população de renda menor não estaria representada na amostra. Acresce, ainda, que provavelmente quando um desses grandes conglomerados estivessem contido em um setor selecionado, vários artifícios seriam necessários, como por exemplo, a criação de outro estágio ou de subamostras para evitar grandes listagens e atenuar a homogeneidade da amostra de domicílios na última etapa.

A existência de falhas operacionais de controle da coleta e manutenção das PNAD não invalidam a idéia do uso dos PNC, mas sim a necessidade de aperfeiçoá-lo, pois o uso dos PNC pode não ser a melhor solução mas é pior não utilizá-los.

4.2 Estimativas de totais de domicílios usando ou não os PNC

A metodologia da PNAD e a da POF-87 conduzem a dois estimadores naturais e diferentes. Qual deles se aproxima melhor do valor verdadeiro desconhecido do total de domicílios, para um determinado ano?

A fim de responder a questão formulada e verificar a repercussão dos grandes crescimentos no número de domicílios e dos setores da amostra da PNAD, foi realizada uma experiência com dados reais da PNAD, nos anos de 1986 a 1990, para os Municípios da Grande Vitória.

O estimador natural do total de domicílios da PNAD é formado pela soma de duas parcelas: uma corresponde a expansão dos resultados do total de domicílios dos setores normais, exclusive os PNC neles localizados e a outra é o total de domicílios cadastrados nos PNC, até o ano da estimativa. Para um particular ano, tem-se:

$$\hat{D}_h^{(1)} = \hat{D}_h + {}_2\hat{D}_h$$

onde:

$\hat{D}_h^{(1)}$ = estimativa do total de domicílios do Município h; em que (1) significa PNAD

\hat{D}_h = estimador natural do total de domicílios, dos setores normais da PNAD, do Município h e

D_h = total de domicílios dos PNC do Município h.

tem-se:

$${}_1\hat{D}_h = \sum_{j=1}^{n_h} \frac{1}{f_{hj}} \cdot {}_1D_{hj} = \sum_{j=1}^{n_h} \frac{I_{hj}}{20 D_{hj}} \cdot {}_1D_{hj}$$

sendo:

n_h - número de setores na amostra do Município h

f_{hj} - fração de amostragem do setor j do Município h

I_{hj} - intervalo de seleção do setor j do Município h

${}_{20}D_{hj}$ - total de domicílios, em 1980, do j-ésimo setor do Município h

${}_1D_{hj}$ - total de domicílios em cada um dos anos de 1986 a 1990 do j-ésimo setor do Município h e

$${}_2D_h = \sum_{j=1}^{n_h} {}_2D_{hj}$$

onde

${}_2D_{hj}$ - total de domicílios no j-ésimo PNC do Município h.

O estimador natural do total de domicílios segundo o critério da POF-87 corresponde à expansão dos resultados dos totais de domicílios dos setores da PNAD, inclusive de todos os domicílios neles localizados. Ou seja:

$$\hat{D}_h^{(2)} = \sum \frac{1}{f_{hj}} \cdot D_{hj}^{(2)}$$

onde:

$\hat{D}_h^{(2)}$ - estimativa do total de domicílios do Município h; onde (2) representa o critério POF-87 e

$D_{hi}^{(2)}$ - total de domicílios do setor j do Município h, inclusive projetos PNC do setor.

É oportuno observar que utilizou-se os dados do universo de cada PNC, embora a PNAD utilize em cada ano uma amostra dos projetos para representar os PNC.

Os resultados das estimativas elaboradas por ambos os critérios estão reunidos na tabela 5, a seguir.

Tabela 5. Estimativas dos totais de domicílios pelos critérios da PNAD e da POF-87 segundo os municípios - 86 a 90
continua.

Especificação	Projetos PNC		Expansão setores normais da PNAD	Expansão	
	Universo	Expansão para o Critério POF-87		Critério da PNAD	Critério da POF-87
(a)	(b)	(c)	(d)	(e)=(b)+(d)	(f)=(c)+(d)
<u>Grande Vitória</u>					
CD-80	-	-	186.274	-	-
1980	-	-	187.180	187.180	187.180
1986	40.776	25.167	222.388	263.164	247.555
1987	41.320	25.167	228.008	269.328	253.175
Expansão 1988	46.040	25.167	233.411	279.451	258.578
1989	47.356	25.167	241.668	289.024	266.835
1990	47.768	25.167	247.380	295.148	272.547
CD - 1991	-	-	307.260	-	-
<u>Cariacica</u>					
CD-80	-	-	46.400	-	-
1980	-	-	46.886	46.866	46.866
1986	5.589	-	53.707	59.296	53.707
1987	5.589	-	54.571	60.160	54.571
Expansão 1988	5.726	-	54.876	60.602	54.876
1989	5.969	-	57.656	63.625	57.656
1990	5.969	-	60.897	66.866	60.897
CD-91	-	-	74.986	-	-
<u>Serra</u>					
CD-80	-	-	27.430	-	-
1980	-	-	29.430	29.430	29.430
1986	16.739	16.544	34.319	51.058	50.863
Expansão 1987	17.147	16.544	33.772	50.919	50.316
1988	21.389	16.544	33.636	55.025	50.180
1989	21.389	16.544	34.455	55.844	50.999
1990	31.606	16.544	35.068	56.674	51.612
CD-91	-	-	64.948	-	-

conclusão

Especificação	Projetos PNC		Expansão setores normais da PNAD	Expansão	
	Universo	Expansão para o Critério POF-87		Critério da PNAD	Critério da POF-87
(a)	(b)	(c)	(d)	(e)=(b)+(d)	(f)=(c)+(d)
<u>Viarã</u>					
CD-1980	-	-	6.104	-	-
1980	-	-	5.962	5.962	5.962
1986	2.270	-	4.752	7.022	4.752
1987	2.270	-	4.752	7.022	4.752
1988	2.270	-	4.752	7.022	4.752
1989	2.270	-	5.192	7.462	5.192
1990	2.270	-	5.126	3.396	5.126
CD-1991	-	-	11.547	-	-
<u>Vila Velha</u>					
CD-1980	-	-	53.326	-	-
1980	-	-	53.090	53.090	53.090
1986	5.456	2.904	57.569	63.025	60.473
1987	5.504	2.904	62.945	68.449	65.849
1988	5.504	2.904	65.954	71.458	68.858
1989	5.580	2.904	69.748	75.328	72.652
1990	5.580	2.904	71.011	76.591	73.915
CD-1991	-	-	77.588	-	-
<u>Vitória</u>					
CD-1980	-	-	53.014	-	-
1980	-	-	52.612	52.612	52.612
1986	10.722	5.719	72.041	82.763	77.760
1987	10.810	5.719	71.968	82.778	77.687
1988	11.151	5.719	74.193	85.344	79.912
1989	12.148	5.719	74.617	86.765	80.336
1990	12.343	5.719	75.278	87.621	80.997
CD-1991	-	-	78.151	-	-

A amostra da PNAD é representativa, com certo nível de precisão, para cada Unidade da Federação e Região Metropolitana. As estimativas para níveis mais detalhados, como por exemplo, a Grande Vitória e seus municípios, devem apresentar maiores variabilidades. Não obstante, o uso das estimativas para esses níveis é realizado como exemplificação do assunto.

A análise dos resultados da tabela 5 para a Grande Vitória, possibilita várias observações. A estimativa referente ao ano de 1980 é muito próxima do resultado do CD-80, indicando que a amostra é bem representativa para o ano do Marco Amostral.

A estimativa do total de domicílios dos projetos PNC, para o critério da POF-87, coluna (c), representa apenas 53% do Universo dos PNC. Essa diferença afeta, para menos, a expansão da amostra pelo critério da POF-87 coluna (f).

A expansão da amostra pelo critério da PNAD é superior cerca de 8% ao resultado apresentado pelo critério da POF-87. A comparação de ambas as estimativas para o ano de 1990 com o resultado do censo demográfico de 91, considerando a defasagem de 1 ano dos dados, indica que o critério da PNAD é muito mais próximo do censo do que a estimativa pelo critério POF-87.

Cabe registrar que na Pesquisa POF-87 os estimadores naturais da população subestimaram os resultados, o mesmo ocorre agora na Grande Vitória.

Depreende-se dessas observações que o critério de expansão da PNAD é mais representativo do que o critério da POF-87, apesar dos problemas operacionais da coleta e perda de ponderação da amostra decorrente do envelhecimento do marco amostral.

Os resultados em nível de Município indicam, também, que as estimativas elaboradas pelo critério da PNAD são superiores às estimativas pelo critério da POF-87 e mais se aproximam dos resultados do CD-91, exceto para Vitória. As estimativas para Vitória, referentes aos setores normais, estão bastante influenciadas pelo grande crescimento do setor 174 que em 1980 possuía o total de 291 domicílios enquanto em 1990 atingiu a 756 domicílios. Aliás esse setor, já em 1986, revelava grande acréscimo com os seus 661 domicílios.

4.3 Estimativas da variabilidade da expansão dos setores normais da PNAD

Tem-se afirmado que, com o passar dos anos, a amostra da PNAD perde eficiência, tendo em vista os problemas operacionais existentes de coleta e controle nos tamanhos dos setores, os quais refletem na variabilidade.

Assim, procurou-se avaliar o coeficiente de variação da estimativa do total de domicílios apenas para os resultados provenientes da expansão dos setores normais.

O estimador da variância do total estimado de domicílios dos setores normais e dados por:

$$V\{\hat{D}_h\} = \sum_{j=1}^{m_h} \frac{m_h^2}{m_h(m_h-1)} \cdot \sum_{j=1}^{m_h} \left(\hat{D}_{hj} - \frac{1}{m_h} \sum_{j=1}^{m_h} \hat{D}_{hj} \right)^2$$

e o coeficiente de variação é dado por

$$CV\{\hat{D}_h\} = \sqrt{\frac{V\{\hat{D}_h\}}{\hat{D}_h}}$$

Para a Grande Vitória tem-se:

$$V\{\hat{D}\} = \sum_{h=1}^5 V\{\hat{D}_h\}$$

onde é o estimador da variância do total de domicílios da Grande Vitória. Analogamente o coeficiente da variação é dado por:

$$CV\{\hat{D}_h\} = \sqrt{\frac{V\{\hat{D}\}}{\sum_{h=1}^5 \hat{D}_h}}$$

Os resultados dos CV referentes a 1980 e 1990 estão reunidos na tabela a seguir.

Tabela 6. Estimativas dos CV para os anos 1980, 1986 e 1990 segundo os Municípios e Grande Vitória

Especificação	Estimativas do CV %		
	1980	1986	1990
Grande Vitória	0,35	3,80	5,10
Cariacica	0,80	2,20	7,60
Serra + Viana	0,46	6,80	9,00
Vila Velha	0,84	6,90	10,30
Vitória	0,46	9,70	11,40

Como era esperado, as estimativas dos CV referentes ao ano de

1980 são baixos, pois é o ano do marco amostral, onde os totais dos domicílios dos setores da amostra são os mesmos do Marco.

Em 1986 as estimativas de totais dos domicílios dos Municípios já apresentaram maiores variabilidades, resultando CV mais elevados, principalmente para Vitória, pelas razões já expostas.

Os mais elevados CV referem-se ao ano de 1990, principalmente o referente ao Município de Vitória.

Nota-se, portanto, que a medida em que o ano de referência da PNAD se afasta do ano do Marco Amostral, maior é a variabilidade, em decorrência do envelhecimento da amostra e falhas nos controles sobre o crescimento do número de domicílios dos setores da amostra.

4.4 Distribuição dos tamanhos dos PNC segundo os Municípios e Grande Vitória

Durante a década de 80 diversos projetos de novas construções foram identificados e cadastrados. Alguns pequenos, outros muito grande. Esses dados estão reunidos por classes de tamanho, em número de domicílios, na tabela a seguir.

Tabela 7. Total de projetos e de domicílios no universo dos PNC por classes de tamanho dos projetos segundo os Municípios

Municípios	Total		Até 99 Domicílios		100 a 199		200 a 499		500 a 999		1000 a mais	
	N	DOM	N	DOM	N	DOM	N	DOM	N	DOM	N	DOM
Total	140	47.768	85	4.303	16	2.309	18	5.709	9	6.482	12	28.965
Carliacica	13	5.969	6	316	2	306	3	722	1	696	1	3.929
Serra	23	21.616	6	285	2	237	5	1.579	2	1.479	8	18.026
Viana	1	1.170	-	-	-	-	-	-	-	-	1	2.270
Vila Velha	19	5.580	9	371	2	299	4	1.834	3	2.184	1	1.152
Vitória	84	12.343	64	5.331	10	1.467	6	1.834	3	2.123	1 ⁽¹⁾	3.588 ⁽¹⁾
		3						1.834		2.123		
							4		3			

Fonte: Relação dos PNC e informações da Pesquisa de Mercado Habitacional de 1988.

(1) Área do Lixão, etc.

Cerca de 60% dos PNC tem até 99 domicílios, com média de 51 domicílios. Esses são os pequenos projetos que contribuem com menos de 10% do total de domicílios. Observa-se que a maioria desses projetos está localizada no Município de Vitória, capital do Estado do Espírito Santo.

Os maiores projetos, (com mais de 100 domicílios cada um), representam menos de 10% dos projetos mas contribuem com mais de 60% do total de domicílios. A maior parte deles está localizada no Município de Serra.

O cadastramento dos PNC, identificação no campo e controle das novas unidades, principalmente nos setores normais, constituem uma tarefa trabalhosa e cuidadosa. Em princípio cada projeto requer o mesmo cuidado independentemente da sua dimensão.

Cabe uma pergunta. Por que o tamanho mínimo de um PNC é igual a 30 domicílios?

Por outro lado a variação dos tamanhos dos setores, no marco amostral é muito grande: a média é de 250 domicílios e o máximo pode atingir várias vezes essa média. Além disso, o controle dos PNC, nos setores da amostra da PNAD, revelou problemas; alguns projetos não foram identificados nesses setores, conforme já foi exposto.

Considerando esses pontos, acredita-se que o tamanho mínimo dos PNC poderia ser alterado. Por exemplo, em se adotando esse mínimo como 100 domicílios, os trabalhos que o PNC exige pode ser reduzido em mais de 50%. Com esta redução, seria possível aumentar os rigores e controles do PNC, principalmente nas áreas da população favelada, mais difíceis de trabalho, como no caso do Lixão, em Vitória.

5. USO DO ESTIMADOR DE RAZÃO NA EXPANSÃO DOS RESULTADOS DAS PESQUISAS DOMICILIÁRIAS

Diante das dificuldades existentes na condução das pesquisas domiciliárias contínuas ou periódicas, — em decorrência da perda de ponderação da amostra; dos deficientes controles nas operações de coleta, na atualização insuficiente do marco de amostragem, etc. —, e acrescido o fato de que os resultados das pesquisas devem ser coerentes com os dados oficiais das projeções de população, o IBGE optou pelo estimador de razão e o utiliza há muitos anos na expansão dos resultados da PNAD, PME, POF, etc.

O estimador de razão adotado emprega como variável auxiliar a projeção da população residente, para uma data específica, para o nível geográfico de estimação, obtido independentemente da amostra.

Esse estimador é definido através de:

$$\hat{y}_h = \frac{\hat{y}}{\hat{x}} \cdot x_p$$

onde:

\hat{Y}_n - é o estimador de razão para a característica Y ajustado à projeção da população residente;

\hat{Y} - é o estimador natural (segundo o desenho da amostra), do total da característica Y;

\hat{X} - é o estimador natural da população residente, para o marco amostral e

X_p - é o resultado de projeção independente da população residente

A precisão das estimativas, expressas pelo coeficiente de variação das estimativas da amostra é examinada agora. A variância do estimador de razão definido anteriormente depende da variabilidade dos dados da amostra e da componente do processo de projeção da população. Como o cálculo da variabilidade da projeção é impraticável, ele é desprezado; então a variabilidade do estimador usado passa a ser definido apenas pela variabilidade da amostra.

A variância relativa de \hat{Y} é dada por:

$$V_n \{ \hat{Y}_n \} = V_n \{ \hat{Y} \} + V_n \{ \hat{X} \} - 2 c_n (\hat{Y}, \hat{X})$$

onde:

$V_n \{ \hat{Y} \}$ - estimador da variância relativa do estimador natural de Y;

$V_n \{ \hat{X} \}$ - estimador da variância relativa da população residente, do marco amostral e

$c_n (\hat{Y}, \hat{X})$ - é o estimador da covariância relativa entre os estimadores naturais das características Y e de X

O estimador do coeficiente de variação do estimador de razão para o total da característica Y é dado por:

$$CV \{ \hat{Y}_n \} = \sqrt{V_n \{ \hat{Y}_n \}}$$

As expressões acima são calculadas pelo método "Ultimate Cluster".

Não é demais observar que, na prática, as projeções de população podem apresentar grandes variações, principalmente para espaços geográficos onde ocorram grandes movimentos migratórios, entretanto as variabilidades relativas dos dados da amostra são pequenas e bem menores.

Para maiores detalhes do cálculo dos erros amostrais, sugere-se a leitura do documento **Pesquisa de Orçamentos Familiares(4)**.

6. CONCLUSÕES

1. O uso da medida de tamanho, total de domicílios, para medir a probabilidade de seleção dos Municípios parece ser melhor do que o total de pessoas.
2. A opção pela medida de tamanho, total de domicílios, na seleção dos setores, produz menos subestimação do que o critério da POF.
3. É necessário melhorar os controles de coleta quanto às variações dos tamanhos dos setores a fim de melhorar a qualidade de amostra.
4. O aumento do tamanho mínimo de 30 domicílios, como critério de PNC, reduz substancialmente os trabalhos e controles dos PNC, sem afetar a sua principal finalidade.
5. A utilização das PNC para atualizar e controlar a variação dos tamanhos dos setores contribuem para aumentar a representatividade da amostra.
6. O uso do estimador de razão, com base em projeções independentes de população atenuam os problemas da perda de ponderação da amostra, mas a precisão dos resultados depende da precisão das projeções.

7. REFERÊNCIAS BIBLIOGRÁFICAS

1. COSTA, Luiz Nery da. Marcos de amostragem. Santiago: CIENES, 1994, mimeo.
2. FUNDAÇÃO INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Levantamento de projetos de novas construções. Rio de Janeiro, 1993.
3. OFICINA CENTRAL DE ESTATÍSTICA E INFORMÁTICA. Encuesta de hogares por muestreo. Venezuela, 1984.
4. FUNDAÇÃO INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Pesquisas de orçamentos familiares; Aspectos de amostragem, V.3. Rio de Janeiro, 1990.
5. BIANCHINI, Zélia Magalhães. Projeções da População e o Ajuste das Estimativas das Pesquisas Domiciliares: Solução ou Problema? IBGE: Rio de Janeiro, 1989.
6. COSTA, Luiz Nery da. Pesquisa habitacional na área da grande Vitória em 1988. Instituto de Desenvolvimento Industrial do Espírito Santo (IDEIES): Espírito Santo, 1989.

ANEXO 1

Tabela 1 - Domicílios por ano, segundo os setores selecionados nos Municípios da Grande Vitória - PNAD-80

Municípios e setores	1980		1986	1987	1988		1989	1990
	DT	DO	DT	DT	DT	DO	DT	DT
Cariacica								
004	428	391	436	434	435	404	443	439
024	312	298	354	354	357	323	364	366
032	435	377	430	435	436	402	444	455
056	218	196	232	240	239	214	231	237
078	342	320	478	488	493	442	512	545
097	318	277	383	385	385	359	400	411
111	401	288	394	436	445	400	595	704
131	209	199	280	281	282	216	275	295
Serra								
014	280	243	335	336	335	306	362	407
020	85	-	124	117	117	105	121	123
043	543	144	544	551	549	531	558	566
067	179	-	183	179	173	160	172	166
083	387	336	447	445	452	397	446	422
Viana								
022	271	?	216	216	216	...	236	233
Vila Velha								
008	204	195	216	212	210	195	211	209
033	309	249	501	554	622	459	682	726
065	371	333	391	427	434	402	438	434
087	268	251	287	320	328	301	345	350
111	356	256	380	407	416	401	453	454
132	263	230	234	244	252	237	265	270
154	508	453	595	583	625	559	667	642
179	341	324	321	339	352	333	363	363
208	188	172	240	240	254	238	275	293
Vitória								
014	276	216	455	384	454	242	431	443
049	228	222	251	247	248	229	253	251
075	172	150	164	178	183	166	188	184
104	383	347	455	459	453	422	454	452
130	452	425	537	559	567	524	583	602
152	296	254	450	446	444	337	443	442
174	291	210	661	682	705	665	734	756
193	541	392	770	794	817	664	89	91
(1)							96	96
208	358	324	368	374	366	336	355	352

Fonte: IBGE - PNAD-80

1) A partir de 89 são subáreas.