

# Sobre os processos de seleção e estimação utilizados na Pesquisa Nacional por Amostra de Domicílios

Alfonso Rodríguez Arias\*

O propósito desta breve Nota é a apresentação de alguns comentários a respeito do desenho de amostragem utilizado na Pesquisa Nacional por Amostra de Domicílios - PNAD. Para tanto, esclarece-se que por desenho de amostragem entendemos o duplo processo de seleção de uma amostra probabilística de domicílios, onde a probabilidade de seleção de cada unidade domiciliar do universo é conhecida e diferente de zero, e o subsequente processo de estimação, através do qual os resultados da amostra são ponderados e expandidos para fornecer estimativas relativas ao conjunto das unidades que compõem o universo.

## Sobre o processo de seleção

De forma muito simplificada, o processo de seleção da amostra de domicílios na PNAD obedece a um modelo seqüencial tri-etápico, onde a probabilidade final de seleção de cada unidade (domicílio, família ou pessoa) é igual a uma constante  $f$  dentro de cada domínio de estimação. A PNAD reconhece um total de 36 domínios geográficos, compostos de nove regiões metropolitanas e 27 estados (excluídos os municípios componentes das nove metrópoles).

Na primeira etapa, a seleção de uma particular UPA  $i$  (unidade primária de amostragem), definida previamente como um único município ou conjunto contíguos destes, se processa aleatoriamente com probabilidade proporcional ao tamanho dela dentro do respectivo estrato  $h$  da qual forma parte.

Sublinha-se que cada uma das UPAs de maior tamanho, entre as quais incluem-se as regiões metropolitanas, os municípios-capitais e outros municípios de grande porte, formam um único estrato, sendo, portanto, selecionadas com probabilidade  $P_{hi}=1$  nesta etapa (UPAs auto-representadas). As restantes UPAs de cada domínio foram agrupadas em um número variável de estratos, formados agora por duas ou mais UPAs, em cada um dos quais foi selecionada aleatoriamente uma única UPA. Dessa forma, as UPAs menores integrantes da amostra, selecionadas nos estratos compostos por duas ou mais UPAs, têm probabilidade  $P_{hi} < 1$  nesta etapa (UPAs não auto-representadas).

As estimativas de tamanho utilizadas na conformação das UPAs bem como na definição dos estratos dentro de cada domínio de estimação correspondem ao total da população de cada município por ocasião do Censo

---

\* Consultor do Instituto de Pesquisa Econômica, IPEA.

Demográfico mais recente. Dessa forma, as probabilidades  $P_{hi}$  das UPAs selecionadas na PNAD 1990 foram calculadas com a população de julho de 1980 ao passo que as relativas às PNADs 1992-1996 foram computadas com a população de setembro de 1991.

Na segunda etapa, a seleção de USAs (unidades secundárias de amostragem) dentro de cada UPA previamente selecionada, correspondentes a setores censitários completos, ou parte ou agregados deles, também se processa com probabilidade proporcional ao tamanho da população delas no último censo. Na formação das USAs cuida-se de que o total de domicílios se aproxime de 300 nas áreas urbanas e 250 nas rurais. Nesta etapa, o procedimento de seleção também é aleatório com ordenamento prévio da totalidade das USAs integrantes da UPA, segundo o grau de urbanização e proximidade geográfica. A probabilidade de seleção da USA  $j$  dentro da UPA  $i$  no estrato  $h$ , designada como  $P_{hij}$ , é inferior a 1.

Destaca-se que a PNAD contempla a seleção de novas USAs dentro das UPAs selecionadas valendo-se da manutenção e atualização permanente de um cadastro de agrupamentos de novas construções sempre que eles sejam iguais ou superiores a 30 domicílios.

Por último, na terceira etapa, a seleção de UTAs (unidades terciárias de amostragem) dentro das USAs previamente selecionadas, referidas à unidade domicílio, se leva a cabo através de uma seleção praticada acima da relação atualizada de unidades domiciliares existentes nas USAs selecionadas, onde o intervalo de seleção de domicílios ( $1/P_{hijk}$ ) garante que o produto das probabilidades das três etapas seja igual à fração final  $f$  = constante para cada domicílio localizado dentro do mesmo domínio de estimação. Dessa forma:

$$P_{hi} * P_{hij} * P_{hijk} = f$$

A atualização da relação de domicílios de cada USA selecionada por ocasião de

cada levantamento anual da PNAD evita que unidades domiciliares que já não mais existem no setor censitário sejam selecionadas para formar parte da amostra, da mesma forma que permite que unidades novas que não constavam da relação anterior ou original de domicílios do setor sejam incluídas no levantamento.

Observado através de uma mesma década, o processo de seleção acima resumido reúne duas particularidades importantes: por uma parte, as probabilidades de seleção não se alteram em nenhuma das etapas, sendo que elas são sempre calculadas a partir dos dados de população do último censo. Esse modelo de seleção pode afetar as estimativas na medida em que as proporcionalidades implícitas em cada uma de suas etapas não espelhem adequadamente as mudanças que ocorrem no universo que representam ao longo da década; segundo, o anterior não significa que o processo de seleção original permanece imutável na década. Conforme comentado, a segunda e a terceira etapas de seleção contemplam atualizações, revisões ou complementações que, espera-se, reproduzam as mudanças domiciliares que ocorrem no universo. A adição de setores de novas construções, a retirada de domicílios inexistentes e a incorporação de outros novos dentro do setor selecionado, sempre observando os intervalos de seleção originais, acabam alterando os tamanhos da amostra de unidades de segunda e terceira etapas da PNAD, produzindo, de um modo geral, um aumento paulatino dos tamanhos finais das amostras em cada domínio de estimação ao longo da década.

Entretanto, essas atualizações parecem insuficientes. O esquema de seleção não incorpora mecanismos de atualização ou de revisão das probabilidades de primeira etapa no decorrer da década, situação que pode levar a sub ou superestimar a participação das unidades não auto-representadas na composição da amostra final do correspondente domínio. Por outra parte, na segunda etapa, o critério de incluir no cadastro de novas

construções somente conglomerados iguais ou superiores a 30 domicílios também pode limitar a seleção de novas USAs na amostra do domínio. Por último, também o tamanho final da amostra pode acabar subestimado pela dificuldade de reconhecer novas unidades domiciliares por ocasião da atualização das listagens das UTAs em cada USA selecionada, pela seleção de unidades domiciliares inexistentes e, ainda, pela subenumeração de pessoas dentro do domicílio selecionado.

Em suma, o marco de amostragem construído no início da década com base nos antecedentes do Censo Demográfico mais recente, apesar das atualizações posteriores, apresenta crescentes dificuldades para acompanhar as mudanças demográficas que se processam no decorrer da década. Essas insuficiências poderiam ser atenuadas, por exemplo, através de revisões quinquenais do processo de seleção aproveitando os dados da Contagem da População, em meados da década, em combinação com o uso de procedimentos que maximizem a probabilidade das UPAs selecionadas, no início da década, permanecerem na amostra.

Mais importante, entretanto, é a revisão do processo de estimação. Além de corrigir as insuficiências decorrentes do processo de seleção da amostra acima comentadas, que resultam em uma subestimação do tamanho da amostra na quase totalidade dos domínios, ele também deveria incorporar mecanismos de correção de outros erros próprios do processo de amostragem ou do levantamento de campo.

### **Sobre o processo de estimação**

O modelo de seleção acima descrito, onde o tamanho final da amostra de cada domínio de estimação é, a rigor, uma variável aleatória, recomenda o uso de indicadores de razão do tipo  $r = y / x$ , onde  $x$  representa uma variável de conta.

As estimações de médias, percentagens e totais e as correspondentes medidas

de dispersão a eles associados :  $var(r)$  ou  $var(Nr)$ , que naturalmente deveriam ser obtidas através do uso do inverso da fração final de cada domínio ( $1/f$ ), atuando como fator de ponderação e expansão das unidades incluídas na amostra, lamentavelmente constituem um procedimento raramente utilizado nas pesquisas domiciliares. As dificuldades decorrem, por uma parte, das perdas de entrevista que ocorrem entre os domicílios ocupados por recusas, ninguém em casa ou outras de caráter emergencial ou acidental e, por outra, pelas insuficiências da atualização do marco de amostragem .

Em lugar dele, recorre-se a um procedimento de estimação que incorpora estimativas populacionais independentes para cada domínio de estimação no tempo  $t$  de referência dos dados.

As experiências internacionais demonstram que essas estimativas populacionais, que geralmente correspondem a projeções demográficas intercensais, são preparadas com diversos graus de detalhamento, dependendo da incidência dos erros amostrais bem como de distorções existentes nos próprios dados levantados, tais como a presença mais acentuada de mulheres ou pessoas com idades terminadas em dígito 0 ou 5, ou outros desvios, como é o caso da situação urbano-rural dos domicílios pesquisados.

Na PNAD, o processo de estimação vale-se de uma única projeção do total da população para cada domínio de estimação, sem desagregações por idade-sexo ou por situação do domicílio. Dessa forma, o levantamento reconhece um único fator de correção que é comum para todas as unidades domiciliares do domínio. Esse fator é calculado como quociente entre o total da população projetada para o domínio e o total da população pesquisada no levantamento entre as unidades que respondem à PNAD.

De um modo geral, os novos fatores de expansão superam em muito os respectivos inversos das probabilidades finais de seleção dos domicílios ( $1/f$ ). A diferença entre ambos

engloba dois tipos de correções: as relativas a não-resposta e as correspondentes às omissões/ excessos incorridos na atualização conjunta das três etapas do marco de amostragem.

Destaca-se que o uso de um único fator de correção por domínio nas estimativas da PNAD em nada modifica os erros ou distorções que possam existir na estrutura etária e de sexo levantadas anualmente pela PNAD entre as unidades respondentes do domínio. Mais ainda, esse procedimento de estimação da PNAD mantém inalterada a composição urbano-rural original da amostra ao longo de toda a década, situação que pode comprometer crescentemente as estimativas dos principais agregados demográfico-ocupacionais da PNAD quando desagregados por área de residência. Nesse sentido, é importante anotar que, ao contrário do que ocorre com o sexo ou a idade, a variável área de residência não é objeto de levantamento anual em cada domicílio selecionado na amostra.

As considerações anteriores sugerem a necessidade de revisar o procedimento de estimação da PNAD. A preparação de projeções demográficas anuais para cada domínio de estimação desagregadas, por exemplo, segundo 6-8 grupos etários, dois sexos (H e M) e duas áreas de residência (U e R), com validade para todo o quinquênio, deve permitir: por uma parte, uma melhor aproximação das estimativas demográficas da PNAD com os resultados dos Censos Demográficos e a Contagem da População; por outra, uma melhor comparabilidade dos resultados entre PNADs ao longo desse quinquênio, reduzindo os efeitos dos erros amostrais e não-amostrais nos resultados anuais.

Como acontece em países que utilizam esse procedimento de estimação com projeções de população mais detalhado, as estimativas são preparadas não apenas valendo-se de um único fator de ponderação por domínio, mas através de uma multiplicidade deles de forma a garantir séries históricas coerentes para os mais importantes agregados demográfico-ocupacionais no tempo.

Naturalmente, as estimativas anuais da PNAD para o quinquênio deveriam ser revistas, uma única vez, à luz dos resultados definitivos do Censo Demográfico ou da Contagem de População.

A seguir, apresentam-se alguns resultados que ilustram os problemas acima levantados a propósito dos processos de seleção e estimação utilizados na PNAD.

### **Apresentação de resultados**

A Tabela 1 apresenta as percentagens de perda completa de entrevistas em domicílios ocupados que foram selecionados nas amostras da PNAD no período de 1992-1996.

De um modo geral, as percentagens de perdas por recusa, não em casa e outras de caráter acidental, são baixas, situando-se em patamares inferiores aos usuais neste tipo de levantamento. Em nível nacional, as perdas anuais na presente década nunca foram superiores a 3%, sendo que no conjunto das nove regiões metropolitanas, embora mais elevadas, nunca ultrapassaram 5%. Ainda que em alguns domínios de estimação específicos as percentagens de perdas possam ser pontualmente mais expressivas, como no caso de Rondônia em 1996, Amapá, Bahia e Rio de Janeiro em 1992 e, de forma mais significativa, nas Metrôpoles de Rio de Janeiro em 1992-1995 e São Paulo e o Distrito Federal em 1996, acreditamos que essa não-resposta não chega a comprometer os resultados demográfico-ocupacionais divulgados pelas PNADs.

As percentagens de perda por não-entrevista completa em domicílios ocupados, juntamente com os diferenciais entre os inversos das frações de amostragem e os fatores de expansão finalmente utilizados para cada domínio nos microdados das PNADs, permitem avaliar o grau de omissão implícito no marco de amostragem. De conformidade com os resultados apresentados na Tabela 2, as perdas atribuíveis à desatualização do marco de amostragem são mais elevadas que as comentadas anteriormente a propósito da

**Tabela 1 - Distribuição dos domicílios ocupados não entrevistados, segundo as Unidades da Federação e Regiões Metropolitanas - 1992/1996**

Unidades da Federação e Regiões Metropolitanas	Distribuição dos domicílios ocupados não entrevistados (%)			
	1992	1993	1995	1996
<b>Brasil</b>	<b>2,55</b>	<b>2,03</b>	<b>1,90</b>	<b>2,63</b>
Rondônia	0,98	1,54	2,46	6,14
Acre	1,95	0,96	0,45	0,84
Amazonas	3,58	2,38	0,75	3,38
Roraima	3,42	0,90	3,94	0,75
Pará	2,06	1,87	2,16	2,36
Amapá	5,19	2,53	4,09	3,13
Tocantins	0,00	0,13	0,82	1,60
Maranhão	1,36	1,40	1,06	1,30
Piauí	0,46	0,73	0,52	1,02
Ceará	2,40	2,74	1,39	2,74
Rio Grande do Norte	0,47	1,08	0,83	0,60
Paraíba	0,44	0,51	0,41	0,53
Pernambuco	4,18	1,67	4,56	2,48
Alagoas	4,18	3,68	2,97	2,92
Sergipe	2,86	4,77	0,41	1,01
Bahia	5,71	3,52	2,52	1,89
Minas Gerais	1,70	1,47	1,25	3,38
Espírito Santo	0,82	1,17	0,47	1,39
Rio de Janeiro	6,41	5,03	4,68	2,74
São Paulo	1,76	0,95	1,96	4,69
Paraná	0,60	0,91	0,74	2,22
Santa Catarina	1,75	2,42	0,57	3,41
Rio Grande do Sul	1,97	1,24	0,59	0,48
Mato Grosso do Sul	1,62	1,82	2,21	1,90
Mato Grosso	0,45	0,88	0,46	0,24
Goiás	0,47	1,65	1,10	2,33
Distrito Federal	1,01	2,51	2,32	7,23
<b>Regiões Metropolitanas</b>	<b>4,73</b>	<b>3,18</b>	<b>3,26</b>	<b>3,49</b>
Belém	2,71	1,67	2,60	1,77
Fortaleza	3,43	3,84	1,87	3,38
Recife	6,00	1,76	5,87	2,30
Salvador	10,07	6,63	3,94	1,84
Belo Horizonte	2,19	1,96	1,36	4,55
Região Metropolitana do Rio de Janeiro	8,29	6,78	6,07	3,21
Região Metropolitana de São Paulo	2,91	1,68	3,63	7,90
Curitiba	1,13	1,06	0,92	3,06
Porto Alegre	3,11	1,59	0,72	0,70

Fonte: Microdados das PNADs 1992-1996.

não-resposta. Segundo essa tabela, as perdas deste tipo, em nível nacional, variaram crescentemente entre 4,4% em 1992 e 7,6% em 1996, mostrando um perfil por estados e regiões metropolitanas também discrepante em relação a não-resposta. De um modo geral, as perdas por desatualização do marco de amostragem são mais elevadas nos domínios não-metropolitanos. Em alguns estados, tais como Rondônia, Amazonas, Bahia e Mato Grosso, onde elas se mantêm elevadas em todos os anos, atingem-se patamares próximos ou superiores a 15% em alguns anos, situação que pode dar margem a

distorções nas estimativas na medida em que os indicadores calculados para o conjunto dos domicílios entrevistados não reflitam a situação das unidades que deixaram de ser selecionadas por desatualização em uma ou mais etapas do marco de amostragem. Destaca-se ainda que este problema, geralmente desconsiderado na avaliação dos resultados das pesquisas domiciliares, é crescente na PNAD até 1996, podendo, inclusive, vir a ser bem mais expressivo nos anos finais da década. Observe-se que nas Metrôpoles de São Paulo e Rio de Janeiro este tipo de perda foi mais que duplicada entre 1992 e 1996.

**Tabela 2 - Distribuição da perda de entrevistas ocasionadas pelo marco da amostragem, segundo as Unidades da Federação e Regiões Metropolitanas - 1992/1996**

Unidades da Federação a	Distribuição da perda de entrevistas ocasionadas pelo marco da amostragem (%)				
	Regiões Metropolitanas	1992	1993	1995	1996
<b>Brasil</b>		<b>4,81</b>	<b>6,35</b>	<b>5,72</b>	<b>8,70</b>
Rondônia		9,91	11,59	22,90	28,27
Acre		9,49	10,93	7,19	7,09
Amazonas		19,88	10,31	14,13	11,43
Roraima		5,92	6,70	(-) 0,09	21,08
Pará		4,10	1,49	5,39	10,20
Amapá		1,13	10,46	6,14	26,26
Tocantins		4,00	6,19	2,49	2,34
Maranhão		6,67	6,76	6,59	11,60
Piauí		(-) 2,25	7,80	3,66	6,50
Ceará		5,02	6,98	0,58	3,10
Rio Grande do Norte		2,52	6,24	(-) 0,04	4,57
Paraíba		6,14	7,65	9,15	6,43
Pernambuco		0,81	7,96	4,41	6,49
Alagoas		5,78	9,61	11,39	11,06
Sergipe		5,24	6,65	4,57	1,30
Bahia		12,20	8,83	15,22	17,34
Minas Gerais		5,97	6,61	4,28	6,47
Espírito Santo		(-) 0,43	(-) 1,17	(-) 2,66	0,39
Rio de Janeiro		(-) 0,26	0,94	2,67	0,32
São Paulo		4,53	4,07	4,32	7,38
Paraná		2,45	3,77	3,51	6,32
Santa Catarina		1,47	2,37	3,41	0,80
Mato Grosso do Sul		3,95	7,01	4,96	9,21
Goiás		4,50	7,20	0,55	2,88
<b>Regiões Metropolitanas</b>					
Belém		(-) 0,12	0,30	(-) 0,65	(-) 3,74
Recife		1,99	0,21	(-) 1,64	3,08
Belo Horizonte		(-) 0,24	(-) 1,18	(-) 2,94	(-) 6,46
São Paulo		3,18	4,61	6,14	9,41
Porto Alegre		2,22	3,33	6,73	5,26

Fonte: Microdados das PNADs 1992-1996.

Os resultados da Tabela 3 relativos à evolução das taxas de crescimento populacional entre PNADs consecutivas revelam comportamentos, por idade e sexo, de alguma forma surpreendentes em se tratando de resultados agregados em nível nacional. Embora reconhecendo que as taxas calculadas correspondem a estimativas sujeitas a erros de amostragem que podem explicar, ao menos parcialmente, algumas dessas diferenças, particularmente em se tratando

de amostras anuais independentes, elas parecem longe de definir um padrão coerente de crescimento demográfico por grupos de idade-sexo na presente década. Naturalmente, essas incertas trajetórias demográficas podem, em maior ou menor medida, afetar a comparabilidade de resultados ocupacionais, de rendimentos e outros divulgados pelas PNADs, notadamente quando eles se referem aos menores domínios.

**Tabela 3 - Evolução das taxas de crescimento populacional, segundo o sexo e grupos de idade - 1992/1996**

Sexo e grupos de idade	Evolução das taxas de crescimento populacional (%)		
	1992/1993	1993/1995	1995/1996
<b>Total</b>	<b>2,10</b>	<b>4,15</b>	<b>2,22</b>
Homem	2,05	4,02	1,98
Mulher	2,16	4,28	2,45
<b>De 10 a 14 anos</b>	<b>2,89</b>	<b>0,03</b>	<b>(-) 0,86</b>
Homem	2,50	0,54	(-) 1,55
Mulher	3,30	(-) 0,49	(-) 0,14
<b>De 15 a 19 anos</b>	<b>2,45</b>	<b>4,50</b>	<b>4,73</b>
Homem	3,71	4,78	5,14
Mulher	1,21	4,22	4,32
<b>De 20 a 29 anos</b>	<b>(-) 0,31</b>	<b>0,75</b>	<b>(-) 0,14</b>
Homem	(-) 0,80	0,36	0,88
Mulher	0,16	1,12	(-) 1,10
<b>De 30 a 39 anos</b>	<b>2,67</b>	<b>4,56</b>	<b>2,16</b>
Homem	2,03	4,89	0,87
Mulher	3,26	4,26	3,37
<b>De 40 a 49 anos</b>	<b>2,70</b>	<b>8,88</b>	<b>5,42</b>
Homem	3,12	8,78	4,18
Mulher	2,31	8,97	6,59
<b>De 50 a 59 anos</b>	<b>3,48</b>	<b>7,09</b>	<b>(-) 1,97</b>
Homem	2,82	6,83	2,86
Mulher	4,07	7,33	1,17
<b>De 60 anos ou mais</b>	<b>2,82</b>	<b>7,77</b>	<b>4,26</b>
Homem	3,59	6,35	3,70
Mulher	2,18	8,95	4,72

Fonte: IBGE, Microdados das PNADs 1992-1996.

Finalmente, o Quadro 1 mostra alguns agregados demográfico-ocupacionais selecionados e divulgados pelas PNADs, Censos Demográficos e Contagem de População nas últimas duas décadas. Visando a possibilitar a comparabilidade de resultados entre essas fontes, foram retiradas das duas últimas os dados correspondentes às áreas rurais dos Estados de RO, AC, AM, RR, PA e AP. Por outra parte, para facilitar a comparabilidade dos resultados sobre os ocupados da PNAD nas duas décadas também foram

excluídos os trabalhadores não remunerados com menos de 15 horas de trabalho na semana, bem como os ocupados sem remuneração que trabalharam na construção de sua própria moradia e na produção de alimentos para o próprio consumo ou das pessoas integrantes de sua família. Por último, no caso da PNAD 1990, apresentam-se estimativas correspondentes à divulgação original e à versão corrigida em microdados à luz dos resultados do Censo Demográfico de 1991.

**Quadro 1 - Principais agregados, por forma de divulgação - 1980/1996**

Pesquisa Nacional por Amostra de Domicílios								
Agregados	1981	1982	1989	1990	1992	1993	1995	1996
População	119 692	122 507	144 293	(1) 147 306 (2) 141 580	146 003	148 217	152 375	154 361
Urbana	86 165	87 254	107 240	(1) 109 106 (2) 104 102	113 837	116 067	120 351	122 511
Rural	34 527	35 253	37 053	(1) 36 200 (2) 37 478	32 166	32 130	32 024	31 850
Ocupados agrícolas (3)	13 300	14 139	14 035	(1) 14 181 (2) 13 998	14 989	14 726	14 551	13 464
Trabalhadores em ocupações agropecuárias e nas extrativas vegetal e animal	12 665	13 431	12 945	(1) 13 065 (2) 12 889	14 147	13 981	13 747	12 842
Censo Demográfico e Contagem da População								
Agregados	1980		1991		1996			
População	116 160				143 106		153 136	
Urbana	80 436				110 991		123 082	
Rural	35 723				32 115		30 054	

Fonte: IBGE, várias PNADs.

Nota: Excluída a zona rural da Região Norte.

(1) Refere-se aos dados originalmente divulgados. (2) Dados revisados, divulgados em CD. (3) O conceito de ocupado é o mesmo da PNAD 1990.

**Os resultados dessa tabela revelam que:**

- Há um significativo viés de superestimação da população total anual estimada na PNAD em ambas as décadas. A discrepância é particularmente acentuada na década passada, chegando a acumular uma diferença de 5,7 milhões de pessoas entre os resultados da PNAD 1990 original e a corrigida. Na presente década, a superestimação da população total alcança a 1,2 milhão quando se compara a estimativa da população total da PNAD de setembro de 1996 com a correspondente à Contagem referida a agosto do mesmo ano;
- Muito embora o total da população da versão corrigida da PNAD 1990 possa guardar uma razoável correspondência com a população do Censo Demográfico registrada em 1991, a desagregação dos resultados por área de residência mostra grande discrepância entre ambas. Tudo leva a crer que as estimativas de população urbana e rural da versão revisada da PNAD 1990 não foram corrigidas. A simples comparação dos totais da população urbano e rural da PNAD original com os resultados do Censo Demográfico dão conta de que a superestimação da população da PNAD em 1990 ocorreu na população rural e não na urbana como apontado pela versão corrigida. As estimativas de população urbana e rural das PNADs desta década evidenciam a distorção existente nas estimativas da população rural da PNAD na década dos 80, erro que lamentavelmente não foi corrigido nos microdados da PNAD 1990. Entretanto, ressalta-se que essa distorção afeta não somente os resultados da PNAD 1990 senão também os dos restantes anos, principalmente os correspondentes à segunda metade de ambas as décadas. Esses resultados reforçam a suspeita de que a manutenção do atributo urbano-rural das USAs ao longo da década produz um erro crescente nos resultados das PNADs quando desagregados por área de residência;
- Há uma boa aproximação entre a evolução dos totais da população rural levantada nos Censos Demográficos, a Contagem e as estimativas dos dois primeiros levantamentos da PNAD de cada década. De conjunto, esses resultados evidenciam uma incontestável tendência declinante da população rural nos últimos 16 anos, aproximando-se a 32,5 milhões em setembro de 1990. Muito embora os resultados originais da PNAD 1990 surgiram de uma elevação na relação ocupados agrícolas/ população rural nos anos 80, esse aumento seria insuficiente para mostrar um crescimento dos ocupados agrícolas entre 1981-1990, conforme revelado pelas PNADs originais e corrigidas desse ano; e
- O fato da PNAD 1992 e posteriores revelarem totais de ocupados agrícolas bem superiores aos 12,7 milhões calculados para 1990, é um complicador adicional para conhecer a real evolução ocupacional ocorrida no campo. A forte elevação da relação ocupados agrícolas/ população rural mostradas pelas PNADs 1992-1996 não parece ter correspondência com a realidade. O assunto requer um estudo mais aprofundado, não descartando-se a possibilidade de que esses resultados tenham alguma explicação, ao menos parcial, na mudança, no conteúdo e no roteiro de entrevista adotados pela PNAD a partir de 1992.