

## Notas técnicas

---

### **Histórico da investigação sobre cor ou raça nas pesquisas domiciliares do IBGE**

No primeiro Censo Demográfico realizado no Brasil, em 1872, a classificação por raças estava presente na forma de quatro opções de resposta: branco, preto, pardo e caboclo, esta última dirigida a contabilizar a população indígena do País. “Ressalva-se que tal escolha foi muito apropriada, pois em um levantamento dessa natureza é importante que os termos empregados tenham uso corrente e o mais disseminado possível para proporcionar maior uniformidade e confiabilidade aos dados obtidos” (OSORIO, 2003, p. 18).

Em 1890, ano do segundo Censo Demográfico, foi utilizado o termo mestiço em substituição a pardo, enquanto as outras três categorias continuaram a ser utilizadas. Verifica-se assim a existência de dois critérios simultâneos na classificação da população no mesmo quesito: um que aponta para o registro da “cor” do entrevistado, utilizando as categorias de branco e preto, e outro que remete à ascendência ou origem racial, manifestando-se na utilização do termo mestiço para os produtos das uniões de pretos e brancos e de caboclo para classificar os índios e seus descendentes. “Note-se que a mestiçagem foi promovida aqui à condição de categoria, assumindo claramente o sentido atribuído pelo branqueamento, qual seja o de diluição do sangue negro no cruzamento com os contingentes migratórios, que levaria ao gradual desaparecimento desta população” (CAMARGO, 2010, p. 243).

A seguir, nos Censos de 1900 e 1920, as informações sobre cor ou raça não foram coletadas e, em 1910 e 1930, não foram realizadas operações censitárias no País.

A partir do Censo 1940, as categorias utilizadas para a classificação, agora denominada de cor, sem fazer referência às raças e com critérios de atribuição diferentes compreendem: branco, preto e amarelo, esta última para dar conta da imigração japonesa ocorrida fundamentalmente entre 1908 e 1930. Assim, a cor tornou-se “o suporte para as representações ambíguas que satisfizeram o ideário de nação que visava agregar e não dividir” (NASCIMENTO, 2008, p. 138). A instrução para o preenchimento do quesito, em 1940, foi de que se considerassem apenas as três respostas mencionadas, lançando um traço (–) no espaço correspondente do questionário em qualquer outro caso. Posteriormente, o traço foi codificado como categoria residual, parda, e foi destinada tanto para classificar os que utilizaram outros termos de cor ou raça, quanto para os indígenas, para quem não se proporciona termo de identificação. “Há aí mudança radical de perspectiva, pois o que se valoriza não são mais os tipos raciais originários, mas a ‘cor’, isto é, as tonalidades de pele, sem a antiga referência à continuidade sanguínea” (CAMARGO, 2010, p. 254).

Os Censos 1950 e 1960 reincorporaram o grupo pardo à categorização de cor, como unidade de coleta e análise, sendo os primeiros levantamentos que orientaram explicitamente nas suas instruções de preenchimento a respeitar a resposta da pessoa recenseada, constituindo a primeira referência explícita ao princípio de autodeclaração.

No Censo 1970, mais uma vez a variável foi excluída da pesquisa, sendo que a partir do Censo 1980 o quesito voltou a ser pesquisado, desta vez no questionário da amostra. Em 1991, foi acrescentada a categoria indígena às já mencionadas, após um século de ausência desta identificação, passando a pergunta a ser denominada como de “raça ou cor” e, no Censo 2000, de “cor ou raça”. Em 2010, último censo realizado, repetiram-se as mesmas categorias de classificação da pergunta, que voltou ao questionário básico aplicado à totalidade da população, sendo que, pela primeira vez, as pessoas identificadas como indígenas foram indagadas a respeito de sua etnia e língua falada.

A investigação de cor ou raça também passou a integrar outras pesquisas domiciliares, tais como a Pesquisa Nacional por Amostra de Domicílios - PNAD, em 1987, a Pesquisa de Orçamentos Familiares - POF, em 2002-2003, e a Pesquisa Mensal de Emprego - PME, em 2003.

Além disso, conta-se com os resultados de duas pesquisas específicas anteriores: a PNAD, realizada em 1976<sup>2</sup>, e a PME, em julho de 1998<sup>3</sup>, em que foram incluídas uma

<sup>2</sup> Pesquisa especial que investigou, por meio de uma subamostra, o tema mobilidade social e a característica cor, denominada pela própria pessoa em um quesito aberto e outro fechado com base em uma classificação em quatro categorias: branca, preta, amarela e parda. A pesquisa, de âmbito nacional, era realizada, à época, em todo o território, exceto na área rural das Unidades da Federação das Regiões Norte e Centro-Oeste, sem contar o Distrito Federal, que estava integralmente representado.

<sup>3</sup> Pesquisa suplementar que investigou a característica cor ou raça, denominada pela própria pessoa em um quesito aberto e outro fechado com base em uma classificação em cinco categorias: branca, preta, amarela, parda e indígena. A pesquisa, realizada nas Regiões Metropolitanas de Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo e Porto Alegre, também investigou, através de quesitos aberto e fechado, a origem do entrevistado, sem nenhuma definição explícita deste conceito. Para informações complementares, consultar a publicação: PETRUCCELLI, J. L. A cor denominada: um estudo do suplemento da PME de julho/98. Rio de Janeiro: IBGE, 2000. 54 p. (Textos para discussão. Diretoria de Pesquisas, n. 3).

pergunta aberta, de resposta espontânea, e outra fechada, ou pré-codificada, sobre a cor ou raça do entrevistado.

Na sequência, encontra-se um quadro que resume a evolução das categorias étnico-raciais consideradas nos Censos Demográficos realizados no País.

**Quadro 1 - Categorias raciais nos Censos Demográficos - Brasil - 1972/2010**

1872	1890	1940	1950	1960	1980	1991	2000	2010
<b>População livre</b> (define sua cor)								
Branca	Branca	Branca	Branca	Branca	Branca	Branca	Branca	Branca
Preta	Preta	Preta	Preta	Preta	Preta	Preta	Preta	Preta
Parda	Mestiça		Parda	Parda	Parda	Parda	Parda	Parda
Cabloca	Cabloca							
		Amarela	Amarela	Amarela	Amarela	Amarela	Amarela	Amarela
		(outras respostas foram codificadas como de cor parda)				Índigena	Índigena	Índigena (se indígena: Etnia e língua falada)
<b>População escrava</b>								

Fonte: Recenseamento do Brasil 1872-1920. Rio de Janeiro: Directoria Geral de Estatística, 1872-1930; e IBGE, Censo Demográfico 1940/2010.

## Processo de construção da pesquisa

As ideias iniciais que propiciaram o processo de discussão e que resultaram na atual pesquisa sobre o sistema de classificação étnico-racial brasileiro remontam ao segundo semestre de 1998, quando se concretizou a proposta de inclusão de quatro perguntas sobre cor ou raça e origem na PME realizada em julho daquele ano, ao mesmo tempo em que se realizou um pequeno seminário na Coordenação de População e Indicadores Sociais, vinculada à Diretoria de Pesquisas, com pesquisadores externos.

A percepção da necessidade de estudar, de forma abrangente e aprofundada, o significado, a construção e a utilização destas categorias conduziu ao planejamento de consultas com representantes dos movimentos negro e indígena, pesquisadores, órgãos estaduais e outros usuários das informações derivadas da classificação de cor ou raça. Assim, numa perspectiva de construção em um processo de diálogo, foram realizados seminários e oficinas de trabalho, visando à elaboração conjunta de uma pesquisa que permitisse aprofundar a compreensão das categorias raciais utilizadas e que subsidiasse a elaboração de novas propostas atualizadas de classificação, atendendo às repetidas demandas da sociedade.

Realizaram-se em 2002 e 2003 três reuniões ampliadas na Diretoria de Pesquisas e quatro seminários nas Unidades Estaduais do IBGE de Salvador, Rio de Janeiro e São Paulo, com participação de mais de 50 pessoas, representando 15 organizações do movimento negro, 12 instituições de pesquisa, organizações não governamentais que trabalham com questões relativas às nações indígenas e secretarias estaduais, além de pesquisadores e técnicos do próprio IBGE. É importante destacar que os três estados onde foram realizados seminários comportam quase a metade dos pouco mais de 10 milhões de pessoas que se declararam de cor ou raça preta no País e dos 66 milhões que se declararam de cor ou raça parda, segundo os dados do Censo 2000.

Uma vez definido o conteúdo da pesquisa, em maio de 2005, foram realizados testes de campo do questionário em três cidades: Recife, Rio de Janeiro e Porto Alegre. Em janeiro de 2007, realizou-se uma reunião de trabalho com mais de 20 convidados externos, provenientes de diversas Unidades da Federação, dentre os quais se destacaram lideranças dos movimentos negro e indígena, pesquisadores e especialistas no tema.

Como produto de todas estas reuniões, configurou-se o formato final do questionário, definindo-se a escolha de seis Unidades da Federação - uma em cada Grande Região brasileira, além do Distrito Federal - para aplicação da pesquisa. Esta foi realizada no segundo semestre de 2008, visando alcançar uma significativa cobertura da diversidade étnico-racial do País.

## Objetivos da pesquisa

A PCERP 2008 pode ser caracterizada como uma pesquisa de objetivos múltiplos que visa compreender melhor o atual sistema de classificação da cor ou raça utilizado nas pesquisas domiciliares do IBGE, contribuindo para seu aprimoramento. Seus objetivos específicos são:

1. Ampliar o espectro de compreensão das categorias nas estatísticas oficiais em relação às questões étnico-raciais;
2. Fornecer novos elementos de interpretação para possíveis alternativas de aprimoramento do atual sistema de classificação étnico-racial;
3. Construir uma base empírica que permita subsidiar estudos e análises sobre o tema;
4. Levantar as denominações correntes de cor, raça, etnia e origem de forma mais abrangente e completa, tanto do ponto de vista da composição étnica da população como das diversidades regionais;
5. Identificar as dimensões que definem a construção e o uso desta terminologia; e
6. Correlacionar os níveis de instrução e a posição na ocupação da população entrevistada com os dos pais, segundo os grupos de cor ou raça.

## População-alvo

A população-alvo da PCERP 2008 é constituída pelos moradores de 15 ou mais anos de idade residentes em domicílios particulares permanentes pertencentes à área de abrangência geográfica da pesquisa. Esta área foi definida como sendo o recorte geográfico formado pelas seguintes Unidades da Federação: Amazonas, Paraíba, São Paulo, Rio Grande do Sul, Mato Grosso e Distrito Federal.

## Aspectos da amostragem

Por ser uma pesquisa domiciliar cujo conteúdo faz parte dos temas previstos para o Sistema Integrado de Pesquisas Domiciliares - SIPD, em fase de implementação no IBGE, o plano amostral da PCERP 2008 utilizou a estrutura amostral desenhada para tal sistema.

Nas seções a seguir, serão abordados aspectos relacionados ao plano amostral, tamanho e alocação da amostra e precisão das estimativas das proporções de pessoas em cada uma das cinco categorias de cor ou raça – branca, preta, amarela, parda e indígena - que serviram de base para o dimensionamento da amostra.

## Plano amostral

A principal finalidade do plano amostral para esta pesquisa foi permitir a obtenção de estimativas das proporções de pessoas em cada uma das cinco categorias de cor ou raça no conjunto das seis Unidades da Federação selecionadas. A pesquisa foi domiciliar e o plano amostral empregado foi amostragem conglomerada em três estágios, com estratificação das unidades primárias de amostragem. Os setores censitários formaram as unidades primárias de amostragem; os domicílios foram as unidades de segundo estágio; e os moradores de 15 anos ou mais de idade definiram as unidades de terceiro estágio.

Como parte integrante do SIPD, as unidades primárias de amostragem da pesquisa foram obtidas da amostra mestra, que é a estrutura amostral do sistema e é composta por setores censitários selecionados com um método probabilístico e da qual é possível selecionar uma subamostra para as pesquisas que integram o sistema. Portanto, a estratificação dos setores censitários (unidades primárias de amostragem) nesta pesquisa foi a mesma adotada para a amostra mestra. Para maiores detalhes, ver Freitas, Lila e Azevedo (2007).

No segundo estágio, foi selecionado por amostragem aleatória simples um número fixo de domicílios particulares permanentes em cada setor selecionado no primeiro estágio. Por fim, no terceiro estágio, em cada domicílio selecionado, um morador com 15 anos ou mais de idade foi selecionado para responder ao questionário, também por amostragem aleatória simples.

## Tamanho da amostra

O tamanho da amostra foi determinado considerando o nível de precisão desejado para as estimativas das proporções de pessoas em cada uma das cinco categorias de cor ou raça. Vários cálculos baseados em amostragem aleatória simples foram efetuados a fim de definir o tamanho da amostra, fixando-se vários níveis de precisão para as estimativas das proporções e para os diversos níveis geográficos.

O plano amostral adotado na pesquisa não foi amostragem aleatória simples de pessoas, por isso foi feito um ajuste no tamanho da amostra obtido, considerando os valores do efeito de plano amostral (EPA). Esta medida, EPA, indica o quanto o plano amostral por conglomerados é menos eficiente (maior variância) que a amostragem aleatória simples, razão pela qual o tamanho da amostra obtido para amostragem aleatória simples é aumentado para que se alcance a mesma precisão. O EPA foi calculado levando-se em conta o número de domicílios/pessoas selecionado em cada setor, a variância do estimador por amostragem aleatória simples, e o coeficiente de correlação intraclasse, que mede a heterogeneidade dentro dos setores em relação à variável de interesse. Para maiores detalhes sobre a definição teórica do efeito de plano amostral, ver Lila e Freitas (2006).

As fórmulas utilizadas foram as seguintes:

$$n_{AAS} = \frac{N}{N-1} \cdot P \cdot Q \cdot \frac{I}{CV^2 \cdot P^2 + \frac{P \cdot Q}{N-1}}$$

$$n_{AC} = n_{AAS} \cdot EPA$$

onde:

$n_{AAS}$  é o tamanho da amostra de pessoas sob amostragem aleatória simples;

$N$  é o número total de pessoas;

$P$  é a proporção de interesse;

$$Q = 1 - P;$$

$CV$  é o coeficiente de variação desejado da estimativa de proporção;

$n_{AC}$  é o tamanho da amostra sob amostragem conglomerada; e

$EPA$  é o efeito de plano amostral.

Os dados utilizados nos cálculos foram obtidos no Censo 2000, tanto os referentes aos números de pessoas, total e por cor ou raça, quanto os referentes aos efeitos de plano amostral, e são apresentados na Tabela 1 a seguir.

**Tabela 1 - Pessoas 15 anos ou mais de idade e efeito do plano amostral, segundo as Unidades da Federação selecionadas e a cor ou raça - 2000**

Unidades da Federação selecionadas e cor ou raça	Pessoas de 15 anos ou mais de idade		Efeito do plano amostral
	Total	Percentual (%)	
<b>Total</b>			
Branca	39 265 552	67,7	2,2
Preta	2 618 226	4,5	1,5
Amarela	495 838	0,9	1,5
Parda	15 023 658	25,9	3,8
Indígena	262 336	0,5	5,9
<b>Amazonas</b>			
Branca	681 717	24,2	2,6
Preta	87 471	3,1	3,7
Amarela	9 343	0,3	3,2
Parda	1 884 507	66,9	3,4
Indígena	113 391	4,0	6,9
<b>Paraíba</b>			
Branca	1 467 260	42,6	2,2
Preta	136 577	4,0	2,1
Amarela	2 439	0,1	1,2
Parda	1 801 161	52,3	2,5
Indígena	10 088	0,3	3,7
<b>São Paulo</b>			
Branca	26 185 687	70,7	1,8
Preta	1 627 267	4,4	1,3
Amarela	456 420	1,2	1,4
Parda	8 456 718	22,8	2,7
Indígena	63 789	0,2	1,7
<b>Rio Grande do Sul</b>			
Branca	8 817 727	86,6	1,6
Preta	527 144	5,2	1,6
Amarela	9 656	0,1	1,2
Parda	762 365	7,5	2,1
Indígena	38 718	0,4	4,4
<b>Mato Grosso</b>			
Branca	1 104 962	44,1	2,3
Preta	141 305	5,6	2,0
Amarela	9 984	0,4	1,3
Parda	1 200 602	47,9	2,4
Indígena	29 196	1,2	6,9
<b>Distrito Federal</b>			
Branca	1 008 199	49,4	2,0
Preta	98 462	4,8	1,3
Amarela	7 996	0,4	1,2
Parda	918 305	45,0	2,4
Indígena	7 154	0,4	1,1

Fonte: IBGE, Censo Demográfico 2000.

O tamanho total da amostra foi calculado avaliando níveis de precisão para as estimativas das proporções de pessoas por categoria de cor ou raça no conjunto das seis Unidades da Federação selecionadas e, para cada uma destas, o tamanho da amostra foi obtido após a alocação do tamanho total nos estratos definidos na amostra mestra.

A cor ou raça com menor proporção de pessoas no conjunto das seis Unidades da Federação era a indígena, uma vez que somente 0,5% das pessoas se declararam indígenas no Censo 2000, por isso o tamanho escolhido foi o necessário para estimar esta proporção com um coeficiente de variação de 30%. O tamanho obtido foi de 14 237 pessoas.

Para calcular o tamanho da amostra de primeiro estágio, foi preciso definir o número de domicílios/pessoas a ser selecionado por setor. O número escolhido foi 10 domicílios; desta forma o tamanho da amostra de setores ficou em 1 424 setores.

Este tamanho inicial sofreu um pequeno acréscimo decorrente da alocação nos estratos definidos para a amostra mestra da qual foram selecionados efetivamente os setores. A alocação nos estratos, diretamente no conjunto das seis Unidades da Federação, foi proporcional ao produto do número de setores pelo máximo entre a proporção de indígenas e de amarelos. Esta combinação do tamanho do estrato com as menores proporções de cor ou raça foi realizada para aumentar o tamanho da amostra em estratos com um maior número de indígenas ou amarelos, permitindo uma possível melhora nas estimativas das características destas populações no conjunto das seis Unidades da Federação. Os tamanhos da amostra por Unidades da Federação são mostrados na Tabela 2 a seguir.

**Tabela 2 - Tamanho da amostra de setores, domicílios e pessoas, segundo as Unidades da Federação selecionadas - 2007**

Unidades da Federação selecionadas	Tamanho da amostra	
	Setores	Domicílios/Pessoas
<b>Total</b>	<b>1 511</b>	<b>15 110</b>
Amazonas	227	2 270
Paraíba	68	680
São Paulo	637	6 370
Rio Grande do Sul	307	3 070
Mato Grosso	227	2 270
Distrito Federal	45	450

Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Métodos e Qualidade.

Após a determinação do tamanho da amostra para cada uma das Unidades da Federação selecionadas, foi feita uma avaliação da precisão das estimativas das proporções de pessoas por cor ou raça. Na Tabela 3, são apresentados os coeficientes de variação esperados das estimativas por cor ou raça segundo as Unidades da Federação.

Observa-se, pelos coeficientes de variação apresentados na Tabela 3, que se deve ter cuidado ao analisar as estimativas das proporções de pessoas em algumas categorias de cor ou raça e, conseqüentemente, em categorias de outras características investigadas na pesquisa, pois o nível de precisão pode não ser aceitável.

**Tabela 3 - Coeficiente de variação esperado da estimativa da proporção de pessoas de 15 anos ou mais de idade, segundo as Unidades da Federação selecionadas e a cor ou raça - 2007**

Unidades da Federação selecionadas e a cor ou raça	Coeficiente de variação esperado da estimativa da proporção das pessoas de 15 anos ou mais de idade (%)
<b>Total</b>	
Branca	0,8
Preta	4,6
Amarela	10,5
Parda	2,7
Indígena	29,1
<b>Amazonas</b>	
Branca	5,9
Preta	22,5
Amarela	62,4
Parda	2,7
Indígena	26,9
<b>Paraíba</b>	
Branca	6,7
Preta	27,0
Amarela	138,0
Parda	5,8
Indígena	135,0
<b>São Paulo</b>	
Branca	1,1
Preta	6,7
Amarela	13,2
Parda	3,8
Indígena	38,3
<b>Rio Grande do Sul</b>	
Branca	0,9
Preta	9,6
Amarela	56,0
Parda	9,2
Indígena	61,0
<b>Mato Grosso</b>	
Branca	3,6
Preta	12,1
Amarela	34,5
Parda	3,4
Indígena	50,5
<b>Distrito Federal</b>	
Branca	6,8
Preta	23,6
Amarela	80,0
Parda	8,2
Indígena	81,0

Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Métodos e Qualidade.

## Expansão da amostra e precisão das estimativas

Os pesos amostrais ou fatores de expansão, utilizados para a obtenção de estimativas dos parâmetros relacionados com as características investigadas na PCERP 2008, foram calculados considerando as probabilidades de seleção, os ajustes pela não resposta, e os ajustes para calibração dos totais estimados.

Para cada domicílio da amostra, foram calculados dois pesos: um para obtenção das estimativas das características gerais de todos os moradores e do domicílio, associado ao domicílio e a cada um de seus moradores; e outro para as características específicas sobre o tema da pesquisa das pessoas de 15 anos ou mais de idade, associado à pessoa selecionada no domicílio.

O cálculo dos pesos amostrais seguiu os seguintes passos, dentro de cada Unidade da Federação:

- a) Peso amostral básico, definido como o inverso da probabilidade de seleção
- probabilidade de seleção do setor para a amostra mestra

$$p_{hi}^a = m_h^a \cdot \frac{N_{hi}}{N_h}$$

onde:

$N_{hi}$  é o número de domicílios no universo do setor  $i$ , do estrato  $h$ , com base nos dados do Censo 2000;

$N_h$  é o número de domicílios no universo do estrato  $h$ , com base nos dados do Censo 2000; e

$m_h^a$  é o número de setores selecionados para a amostra mestra no estrato  $h$ .

- probabilidade de seleção do setor para a PCERP 2008, dado que o setor foi selecionado para a amostra mestra

$$p_{hi} = \frac{m_h}{m_h^a}$$

onde:

$m_h$  é o número de setores selecionados para a PCERP 2008 no estrato  $h$ .

- probabilidade de seleção do domicílio para a PCERP 2008, dado que o seu setor foi selecionado para a pesquisa

$$p_{hij}^d = \frac{10}{N_{hi}^*}$$

onde:

10 é o número de domicílios selecionados por setor da amostra da PCERP 2008; e

$N_{hi}^*$  é o número de domicílios no universo do setor  $i$ , do estrato  $h$ , com base nos dados da listagem realizada nos setores da amostra da PCERP 2008.

- probabilidade de seleção do domicílio

$$p_{hij}^{d*} = p_{hi}^a \cdot p_{hi} \cdot p_{hij}^d = m_h^a \cdot \frac{N_{hi}}{N_h} \cdot \frac{m_h}{m_h^a} \cdot \frac{10}{N_{hi}^*}$$

- peso amostral básico do domicílio e de todos os seus moradores

$$w_{hij}^d = \frac{1}{p_{hij}^{d*}} = \frac{1}{m_h^a} \cdot \frac{N_h}{N_{hi}} \cdot \frac{m_h^a}{m_h} \cdot \frac{N_{hi}^*}{10}$$

- probabilidade de seleção da pessoa, dado que seu domicílio foi selecionado para a PCERP 2008

$$p_{hijk}^p = \frac{1}{O_{hij}}$$

onde:

$O_{hij}$  é o número de pessoas de 15 anos ou mais de idade no domicílio  $j$ , do setor  $i$ , do estrato  $h$ .

- probabilidade de seleção da pessoa

$$p_{hijk}^{p*} = p_{hij}^{d*} \cdot p_{hijk}^p = m_h^a \cdot \frac{N_{hi}}{N_h} \cdot \frac{m_h}{m_h^a} \cdot \frac{10}{N_{hi}^*} \cdot \frac{1}{O_{hij}}$$

- peso amostral básico da pessoa selecionada para a PCERP 2008

$$w_{hijk}^p = \frac{1}{p_{hijk}^{p*}} = \frac{1}{m_h^a} \cdot \frac{N_h}{N_{hi}} \cdot \frac{m_h^a}{m_h} \cdot \frac{N_{hi}^*}{10} \cdot O_{hij}$$

b) Peso amostral com ajuste de não resposta, calculado para compensar a ocorrência de perda de entrevista por não resposta na PCERP 2008 (por domicílio fechado, por recusa dos moradores em atender o entrevistador, ou por recusa da pessoa selecionada em responder o questionário)

- peso amostral do domicílio e todos os moradores, com ajuste por não resposta

$$w_{hij}^{d*} = \frac{1}{m_h^a} \cdot \frac{N_h}{N_{hi}} \cdot \frac{m_h^a}{m_h} \cdot \frac{N_{hi}^*}{10} \cdot \frac{r_{hi}^d + nr_{hi}^d}{r_{hi}^d}$$

onde:

$r_{hi}^d$  é o número de domicílios selecionados para a amostra da PCERP 2008 no setor  $i$ , do estrato  $h$ , com entrevista realizada sobre dados gerais de todos os moradores e do domicílio; e

$nr_{hi}^d$  é o número de domicílios selecionados para a amostra da PCERP 2008 no setor  $i$ , do estrato  $h$ , sem entrevista realizada por não resposta (domicílio fechado, recusa dos moradores em atender o entrevistador, ou outro motivo em domicílios ocupados).

- peso amostral da pessoa selecionada para a PCERP 2008, com ajuste por não resposta

$$w_{hijk}^{p*} = \frac{1}{m_h^a} \cdot \frac{N_h}{N_{hi}} \cdot \frac{m_h^a}{m_h} \cdot \frac{N_{hi}^*}{10} \cdot \frac{r_{hi}^p + nr_{hi}^p}{r_{hi}^p}$$

onde:

$r_{hi}^p$  é o número de pessoas selecionadas para a amostra da PCERP 2008 no setor  $i$ , do estrato  $h$ , com entrevista realizada sobre dados específicos do tema da pesquisa; e

$nr_{hi}^p$  é o número de pessoas selecionadas para a amostra da PCERP 2008 no setor  $i$ , do estrato  $h$ , sem entrevista realizada por não resposta (domicílio fechado, recusa dos moradores em atender o entrevistador, outro motivo em domicílios ocupados, e recusa da pessoa selecionada).

- c) Peso amostral final, em que optou-se por calibrar a estimativa do total de pessoas de 15 anos ou mais de idade, obtida com os dados de todos os moradores da PCERP 2008, pelo total estimado com dados da POF 2008-2009

Além disso, observou-se que a não resposta da pessoa selecionada foi diferenciada por sexo e por faixa etária, levando à calibração das estimativas das pessoas de 15 anos ou mais de idade nestes domínios, obtidas com os dados da pessoa selecionada da PCERP 2008 pelos totais estimados pela POF 2008-2009.

- peso amostral final do domicílio e de todos os seus moradores

$$W_{hij}^d = W_{hij}^{d*} \cdot \frac{\hat{Y}^{pof}}{\hat{Y}} = \frac{I}{m_h^a} \cdot \frac{N_h}{N_{hi}} \cdot \frac{m_h^a}{m_h} \cdot \frac{N_{hi}^*}{10} \cdot \frac{r_{hi}^d + n_{hi}^d}{r_{hi}^d} \cdot \frac{\hat{Y}^{pof}}{\hat{Y}}$$

onde:

$\frac{\hat{Y}^{pof}}{\hat{Y}}$  é o fator de ajuste do peso dos domicílios selecionados e de todos os seus moradores na Unidade da Federação;

$\hat{Y}^{pof}$  é a estimativa do total de pessoas de 15 anos ou mais de idade na Unidade da Federação, proveniente da POF 2008-2009; e

$\hat{Y}$  é a estimativa do total de pessoas de 15 anos ou mais de idade na Unidade da Federação, proveniente da PCERP 2008, considerando todos os moradores.

- peso amostral final da pessoa selecionada para a PCERP 2008

$$W_{hijk}^{psg} = W_{hijk}^{p*} \cdot \frac{\hat{Y}_s^{pof}}{\hat{Y}_s} = \frac{1}{m_h^a} \cdot \frac{N_h}{N_{hi}} \cdot \frac{m_h^a}{m_h} \cdot \frac{N_{hi}^*}{10} \cdot \frac{r_{hi}^p + n_{hi}^p}{r_{hi}^p} \cdot \frac{\hat{Y}_{sg}^{pof}}{\hat{Y}_{sg}}$$

onde:

$\frac{\hat{Y}_{sg}^{pof}}{\hat{Y}_{sg}}$  é o fator de ajuste do peso das pessoas selecionadas do sexo  $s$  e da faixa etária  $g$  na Unidade da Federação;

$\hat{Y}_{sg}^{pof}$  é a estimativa do total de pessoas de 15 anos ou mais de idade do sexo  $s$  e da faixa etária  $g$  na Unidade da Federação, proveniente da POF 2008-2009; e

$\hat{Y}_{sg}$  é a estimativa do total de pessoas de 15 anos ou mais de idade do sexo  $s$  e da faixa etária  $g$  na Unidade da Federação, proveniente da PCERP 2008, considerando a pessoa selecionada.

As faixas etárias consideradas foram as seguintes:

- para o Distrito Federal e o Estado da Paraíba: 15 a 24 anos; 25 a 34 anos; 35 a 44 anos; 45 a 54 anos; e 55 anos ou mais de idade; e
- para os Estados do Amazonas, São Paulo, Rio Grande do Sul e Mato Grosso: 15 a 24 anos; 25 a 34 anos; 35 a 44 anos; 45 a 54 anos; 55 a 64 anos; e 65 anos ou mais de idade.

## Precisão das estimativas

A precisão das estimativas produzidas com os dados da PCERP 2008 é expressa pelo coeficiente de variação (CV), estimado, utilizando-se o Método do Conglomerado Primário (*Ultimate Cluster*)<sup>4</sup>, por meio do *software* SUDAAN (*Survey Data Analysis*). Apresenta-se, a seguir, o estimador da variância do estimador de total de uma variável  $d$  associada ao domicílio e todos os seus moradores e de uma variável  $x$  associada à pessoa selecionada.

$$\hat{V}(\hat{D}) = \sum_h \frac{m_h}{m_h - 1} \sum_i \left( \hat{Z}_{hi} - \frac{\hat{Z}_h}{m_h} \right)^2$$

onde

$$\hat{Z}_{hi} = \sum_j w_{hij}^d \cdot \left( d_{hij} - \frac{\hat{D}_h}{m_h} \right);$$

$$\hat{Z}_h = \sum_i \hat{Z}_{hi};$$

$$\hat{D}_h = \sum_{ij} w_{hij}^d \cdot d_{hij} \text{ e}$$

$d_{hij}$  é o valor da variável  $d$  para o domicílio  $j$ , do setor  $i$ , do estrato  $h$

$$\hat{V}(\hat{X}) = \sum_h \frac{m_h}{m_h - 1} \sum_i \left( \hat{Z}_{hi} - \frac{\hat{Z}_h}{m_h} \right)^2$$

onde:

$$\hat{Z}_{hi} = \sum_{jks} w_{hijk}^{psg} \cdot \left( x_{hijk} - \frac{\hat{X}_h^{sg}}{m_h} \right);$$

$$\hat{Z}_h = \sum_i \hat{Z}_{hi};$$

$$\hat{X}_h^{sg} = \sum_{ijk} w_{hijk}^{psg} \cdot x_{hijk} \cdot I_{hijk}^s \cdot I_{hijk}^g;$$

$x_{hijk}$  é o valor da variável  $x$  para a pessoa selecionada no domicílio  $j$ , do setor  $i$ , do estrato  $h$ ; e

$$I_{hijk}^s = \begin{cases} 1 & \text{se a pessoa selecionada no domicílio } j, \text{ do setor } i, \text{ do estrato } h \\ & \text{é do sexo } s \text{ caso contrario} \\ 0 & \end{cases}$$

$$I_{hijk}^g = \begin{cases} 1 & \text{se a pessoa selecionada no domicílio } j, \text{ do setor } i, \text{ do estrato } h \\ & \text{é da faixa etária } g \text{ caso contrario} \\ 0 & \end{cases}$$

<sup>4</sup> Para detalhes sobre o método, consultar a publicação: HANSEN, M. H.; HURWITZ, W. N.; MADOW, W. G. *Sample survey methods and theory*. New York: Wiley, 1953. 2 v. (Wiley publications in statistics). Ver também: PESSOA, D. G. C.; SILVA, P. L. do N. Análise de dados amostrais complexos. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 13., 1998, Caxambu. *Anais...* São Paulo: Associação Brasileira de Estatística, 1998.

E o estimador do coeficiente de variação para as estimativas de total é dado por:

$$cv(\hat{D}) = \frac{\sqrt{\hat{V}(\hat{D})}}{\hat{D}} \text{ e } cv(\hat{X}) = \frac{\sqrt{\hat{V}(\hat{X})}}{\hat{X}}$$

## Instrumento de coleta

O questionário busca, em linhas gerais, reproduzir uma situação de pesquisa de campo na área da construção das identidades raciais, pretendendo “traduzir” para fins estatísticos fenômenos sociais impregnados de subjetividades. Esses fenômenos se referem a processos de construção de identidades, ao mesmo tempo individuais e coletivas. Nesse sentido, o questionário inicia aproximando o entrevistado gradativamente do tema, apresentando os elementos que, segundo os estudos, são constitutivos destes processos de elaboração de identidades.

O questionário é constituído por blocos de identificação do informante, suas características familiares e socioeconômicas e um bloco dedicado às questões específicas de cor ou raça do indivíduo entrevistado e seus pais.

No bloco sobre as questões relativas ao tema central da pesquisa, a primeira pergunta (quesito 3.06 do questionário anexo) propõe-se a dar início ao tratamento da classificação racial com a opinião do entrevistado sobre a influência da cor ou raça na vida das pessoas. Ao mesmo tempo, esta pergunta visa captar o grau de percepção dos fenômenos sociais de discriminação, baseados na cor ou identificação racial das pessoas.

A seguir, estimulando a pessoa entrevistada a refletir sobre sua própria identificação, pergunta-se se ela saberia dizer qual é sua cor ou raça (quesito 3.07) e, caso a resposta fosse afirmativa, pede-se para especificá-la, levantando os termos que o entrevistado escolhe de forma espontânea para definir a sua cor ou raça. A localização da pergunta no início do questionário obedece à necessidade de quem responde não estar influenciado pelo conteúdo e categorias das demais perguntas que seguirão.

As próximas duas perguntas objetivam pesquisar as dimensões que configuram a identificação de cor ou raça das pessoas. De forma geral, duas são as principais dimensões que podem orientar as respostas: por um lado, a origem familiar da pessoa e, por outro, os aspectos físicos, ou aparência das mesmas, aqui melhor especificadas, além de outras opções, como aspectos referentes à cultura (esfera de valores e tradições, e práticas sociais como a religião, festas, etc.), aspectos vinculados às práticas política e ideológica (que faz a pessoa escolher essa ou aquela categoria a partir de sua inserção em movimentos sociais) e, ainda, situações em que a escolha de uma categoria está vinculada a uma inserção de classe ou de status (quando a cor ou raça está determinada pela leitura que se faz de si mesmo ou de uma outra pessoa a partir de sua condição econômica).

As alternativas visam orientar o entrevistado a escolher entre as principais dimensões que, seguindo a teoria em torno da construção da identidade racial, são mais frequentes tanto para a identificação das pessoas em geral (quesito 3.08) quanto para a autoidentificação (quesito 3.09). Ainda que essas dimensões estejam imbricadas na prática, na grande maioria das vezes atuando em conjunto, o que se pretende, justamente, é induzir o entrevistado a não apenas distingui-las como também a hierarquizá-las por graus de influência que exercem no processo de construção ou de imputação de identidades com base nessas características.

O quesito 3.10 - “*Você é brasileiro(a)?*” - foi incluído com a finalidade de melhor estabelecer o objetivo da pergunta seguinte sobre origem familiar no sentido de tentar evitar, na medida do possível, a previsível resposta “brasileira” em relação à origem, a qual, como já foi verificado nos testes do Censo 2000 e da pesquisa suplementar da PME, realizada em julho de 1998, é escolhida por 75% a 90% dos entrevistados. Se o entrevistado possuir mais de uma nacionalidade e entre elas estiver a “brasileira”, deve-se marcar a resposta afirmativa; de outra forma, a resposta é não.

A pergunta sobre origem familiar (quesito 3.11) procura indagar a respeito de elementos suplementares que podem contribuir para definir o perfil de identificação étnico-racial das pessoas. Mesmo reconhecendo que esta adscrição constitui um processo eminentemente simbólico e subjetivo, a referência a um determinado grupo étnico e/ou a uma origem geográfica remete a um passado, real ou mítico, que proporciona a base para essa adscrição. Esta pergunta permite respostas múltiplas e solicita especificar, quando for possível, país, região, comunidade ou etnia a que a resposta sobre a origem familiar está se referindo.

O conjunto de perguntas que segue pretende dar conta da multidimensionalidade do fenômeno de categorização racial na sociedade brasileira contemporânea. Enquanto tradicionalmente nas pesquisas apenas se levanta uma única dimensão composta por cinco categorias ou possibilidades de resposta excludentes, neste questionário se abrem diversas oportunidades de resposta, permitindo, assim, expressar a reconhecida multiétnicidade que caracteriza a população do País. Ao mesmo tempo em que as opções de resposta possibilitam uma superposição com as categorias clássicas de identificação étnico-racial utilizadas, novos termos são introduzidos, ampliando o escopo de identificação: no primeiro quesito deste sub-bloco (quesito 3.12), pergunta-se se a pessoa se reconhece ou identifica como afrodescendente, sem proporcionar nenhuma definição do termo; no segundo (quesito 3.13), indaga-se sobre a autoidentificação como indígena e, em caso afirmativo, se quer especificar etnia e língua indígena falada; no terceiro (quesito 3.14), se o entrevistado se reconhece como amarelo e, quando a resposta for afirmativa, pede-se para especificar origem geográfica familiar; no quarto (quesito 3.15), se o entrevistado se reconhece ou se identifica como negro; no quinto, (quesito 3.16), como branco; no sexto (quesito 3.17), como preto; e no sétimo (quesito 3.18), como pardo. Os cruzamentos entre as respostas deste conjunto de perguntas permitirão reconstruir de maneira mais acurada o quadro de identificação étnico-racial da população.

Todavia, uma outra pergunta de resposta espontânea de identificação racial foi incluída (quesito 3.19), dando à pessoa entrevistada a opção de escolher, eventualmente, algum outro termo, diferente dos anteriores. Procura-se, desta forma, confirmar ou não, se for o caso, a escolha já feita, ou distinguir, entre aqueles entrevistados que podem apresentar ambiguidades na sua identificação, se existe a possibilidade de uma melhor opção não condicionada por hábitos, usos do senso comum ou categorias consideradas oficiais. Através da prática de pesquisas qualitativas, sabe-se que muitas vezes a primeira abordagem da questão não reflete a verdadeira maneira de pensar do entrevistado.

Dentro deste bloco, finalmente, o quesito 3.20 permitirá conhecer a opinião do entrevistado a respeito dos efeitos da categorização racial na vida das pessoas, em algumas áreas de inter-relação social: casamento, trabalho, escola, atendimento à saúde e em repartições públicas, convívio social e relação com a justiça e a polícia.

A pesquisa também objetiva levantar a informação de cor ou raça dos pais da pessoa entrevistada, ainda que obtida de forma indireta. Estas informações permitirão

acrescentar mais elementos, contribuindo para um melhor discernimento da configuração de alguns grupos raciais mais sujeitos a ambiguidades, como o da categoria "parda", mas também o da "branca" e, eventualmente, também de outras categorias. Por outro lado, representam um subsídio para a análise das respostas das perguntas seguintes sobre o nível de instrução e características ocupacionais dos pais da pessoa entrevistada. Estas informações possibilitam o estudo da trajetória social dos entrevistados em relação aos seus pais e foram formuladas com base no suplemento de mobilidade social da PNAD 1996<sup>5</sup>, para permitir a realização de estudos comparativos com aquelas informações.

Cabe sinalizar finalmente que, apoiados na fundamentação de diversos pesquisadores e especialistas na área de relações raciais no Brasil, foi incluído na pesquisa o quesito 3.00, no qual o entrevistador procede à classificação do entrevistado segundo a cor ou raça, o que se conhece como heteroatribuição da pessoa entrevistada, de forma aberta ou sem pré-codificação. Enquanto alguns estudos apontam para discrepâncias entre ambas as formas de classificação, o que levaria a resultados divergentes na análise das desigualdades raciais associadas à renda, à escolaridade ou à habitação, outras análises apontam para um razoável grau de concordância entre elas. Em todo caso, ainda não se dispõe de informações suficientes para determinar o sentido nem o tamanho desta discrepância. Há de se sublinhar, todavia, que os entrevistadores não receberam nenhum tipo de instrução nem foram orientados sobre como preencher este quesito.

<sup>5</sup> Para informações complementares, consultar a publicação: MOBILIDADE social 1996. Rio de Janeiro: IBGE, 1997. 2 v. Acima do título: Pesquisa Nacional por Amostra de Domicílios.