

TEXTOS PARA DISCUSSÃO, ISSN 0103-6661

APURAÇÃO DOS DADOS INVESTIGADOS  
NO QUESTIONÁRIO BÁSICO (CD 1.01)  
DO CENSO DEMOGRÁFICO DE 1991

NÚMERO 71

OUTUBRO DE 1994

**APURAÇÃO DOS DADOS INVESTIGADOS NO QUESTIONÁRIO  
BÁSICO (CD 1.01) DO CENSO DEMOGRÁFICO DE 1991**

**LUIS CARLOS DE SOUZA OLIVEIRA**

**Bacharel em Estatística**

**LAURA BARIDÓ INDÁ**

**Bacharel em Estatística**

**RITA LUZIA AGUIAR LIMA**

**Bacharel em Estatística**

**ZÉLIA MAGALHÃES BIANCHINI**

**Matemática e Mestre em Estatística**

**RIO DE JANEIRO**

**1994**

FUNDAÇÃO INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE

Av. Franklin Roosevelt, 166 - Centro  
CEP 20021-120 - Rio de Janeiro, RJ - Brasil

**DIRETORA DE PESQUISAS**

TEREZA CRISTINA NASCIMENTO ARAÚJO

**DIRETORA-ADJUNTA DE PESQUISAS**

MARIA MARTHA MALARD MAYER

**CHEFE DA COORDENAÇÃO TÉCNICA DO CENSO DEMOGRÁFICO**

VALÉRIA DA MOTTA LEITE

**CHEFE DA DIVISÃO DE METODOLOGIA**

ZÉLIA MAGALHÃES BIANCHINI

© IBGE

**TEXTOS PARA DISCUSSÃO**

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

Edição: Divisão de Documentação e Disseminação da Diretoria de Pesquisas.  
(DDI/DPE)

Apuração dos dados investigados no questionário básico  
(CD1.01) do censo demográfico de 1991 / Luis Carlos de Souza  
Oliveira ... (et.al.) . - Rio de Janeiro : IBGE, Diretoria de  
Pesquisas, 1994.  
91 p. - (Textos para discussão / IBGE, Diretoria de Pesquisas.  
ISSN 0103-6661 ; n. 71)

ISBN 85-240-0504-1

1. Crítica de dados - Metodologia. 2. Brasil - Censo demográfico,  
1991 - Questionários. I. Oliveira, Luis Carlos de Souza. II. IBGE.  
Diretoria de Pesquisas. III. Série.

IBGE.CDDI. Dep. de Documentação e Biblioteca  
RJ/IBGE-94/21

CDU 311.214  
EST

**Informações:** Biblioteca Setorial da Diretoria de Pesquisas -  
Rua Visconde de Niterói, 1.246. Bloco B, sala 1.211-B, Mangueira  
Telefone: (021) 284-8938 / 284-3322 - ramal 303

## APRESENTAÇÃO

Este documento tem por objetivo apresentar as especificações adotadas para o tratamento dos dados investigados no questionário básico - CD 1.01 aplicado no Censo Demográfico de 1991, desde a etapa de transcrição dos dados em papel para o meio magnético até o término da apuração e o armazenamento dos dados em base de dados.

Este texto surge como registro e documentação para que se possa ter uma visão global, abrangente e ao mesmo tempo minuciosa do tratamento que os dados coletados sofreram antes de serem tabulados e divulgados.

É resultado da reunião de especificações elaboradas por vários técnicos da Coordenação Técnica do Censo Demográfico - CTD e da Divisão de Metodologia - DME da Diretoria de Pesquisas e da Coordenação de Informática do Censo Demográfico - COI e da Divisão de Sistemas para Censo e Pesquisas Demográficas - DIDEM do Departamento de Atendimento - DEATE da Diretoria de Informática.

## **AGRADECIMENTOS**

São tantas as pessoas envolvidas no processo de apuração do Censo Demográfico que não seríamos capazes de citar todas nominalmente. Foram anos de trabalho e alguns dos técnicos que participaram do planejamento hoje já não trabalham mais no IBGE. Queremos registrar nosso reconhecimento pela atuação e contribuição dos técnicos envolvidos no decorrer de todo o processo de apuração.

Agradecimentos especiais a Heleno Ferreira Mansoldo - chefe da COI pelas informações, comentários e sugestões que contribuíram para a preparação deste documento; a Edvaldo Monteiro de Oliveira pela colaboração na elaboração das figuras e gráficos; e a Sonia Albieri e Mauro dos Santos Mendonça pela atenciosa leitura, comentários e revisão final do presente texto. Os erros e omissões são, no entanto, de nossa exclusiva responsabilidade.

OS AUTORES

## SUMÁRIO

1. INTRODUÇÃO . . . . .	9
2. VISÃO GERAL DO PROCESSO DE APURAÇÃO DOS DADOS . . . . .	11
3. A RECEPÇÃO E O EMPASTAMENTO DO MATERIAL COLETADO . . . . .	14
4. A TRANSCRIÇÃO DOS DADOS . . . . .	18
4.1 Planejamento da Transcrição dos Dados . . . . .	18
4.2 Execução da Transcrição dos Dados . . . . .	21
5. A PRÉ-CRÍTICA DOS DADOS . . . . .	26
5.1 Variáveis Tratadas na Pré-Crítica . . . . .	27
5.2 Criação de Questionários Faltosos . . . . .	29
5.3 Execução da Pré-Crítica . . . . .	30
6. A DETECÇÃO E A CORREÇÃO AUTOMÁTICA DOS ERROS . . . . .	33
6.1 A Motivação para a Busca de uma Nova Metodologia . . . . .	33
6.2 A Metodologia Proposta por Fellegi e Holt . . . . .	37
6.3 O Sistema DIA . . . . .	39
6.3.1 Subsistema de especificação . . . . .	40
6.3.2 Subsistema de tratamento dos dados . . . . .	43
6.3.3 Informações para análise do processo de detecção e correção dos erros . . . . .	46
6.4 A Aplicação do Sistema DIA aos Dados do CD 1.01 . . . . .	49
6.4.1 A definição das regras de crítica . . . . .	49
6.4.2 A estratégia para o tratamento dos dados . . . . .	51
6.4.3 A formação e seleção dos lotes de apuração . . . . .	58
6.4.4 A execução da detecção e correção automática dos erros pelo sistema DIA . . . . .	62
6.5 A Imputação dos Dados em Questionários Faltosos . . . . .	66
6.6 Procedimentos Adicionais . . . . .	69

6.7 A Análise do Processo Automático de Correção . . . . .	70
6.7.1 A análise para a liberação da correção automática . . . . .	71
6.7.2 A análise consolidada dos resultados da correção automática . . . . .	73
 7. O ARMAZENAMENTO E A LIBERAÇÃO DOS DADOS PARA ANÁLISE, TABULAÇÃO E DISSEMINAÇÃO . . . . .	 81
 REFERÊNCIAS . . . . .	 87
 ANEXOS . . . . .	 91
Anexo 1 - Cópia do Questionário Básico - CD 1.01	
Anexo 2 - Gráfico das Percentagens de Divergências na Digitação por pasta dos dados do Censo Experimental de Limeira de 1988	
Anexo 3 - Plano de Transcrição de Dados do CD 1.01	
Anexo 4 - Plano de Pré-Crítica do CD 1.01	
Anexo 5 - Planos de Crítica para a Detecção e Correção Automática dos erros pelo Sistema DIA	
Anexo 5.1 - Aplicação de Domicílio	
Anexo 5.2 - Aplicação de Chefe	
Anexo 5.3 - Aplicação de Não-Chefe	
Anexo 6 - Exemplo das Tabelas Especiais para a Análise da Correção Automática	
Anexo 7 - Gráficos sobre a Correção Automática do CD 1.01 por Unidades da Federação	

## 1. INTRODUÇÃO

O Censo Demográfico é das mais importantes pesquisas (se não a mais importante) que o IBGE realiza. Seus resultados servem aos mais variados tipos de usuários, para os mais diversos propósitos analíticos, de planejamento e tomada de decisão. Além disso, os dados do Censo servem de base para todo um conjunto de outras pesquisas a serem desenvolvidas durante a década, e em muitos casos a dependência dos dados provenientes do Censo Demográfico é crucial para essas pesquisas. Veja-se, por exemplo, o caso das pesquisas domiciliares por amostragem do IBGE, que baseiam toda a seleção e estimação de suas amostras nas informações provenientes do Censo.

Por essa razão, o conhecimento preciso dos dados e informações resultantes do Censo Demográfico de 1991 é indispensável não só para aqueles que dão uso imediato a essas informações, mas também para aqueles que, ao longo dos anos vindouros, terão o Censo como base para o planejamento de suas pesquisas. Para que esse conhecimento seja não só preciso mas completo, é fundamental ter registrado e devidamente documentado como os dados coletados foram tratados nas diversas etapas de processamento, antes de serem armazenados em base de dados, a partir da qual são recuperados para a produção das tabelas e dos demais produtos de divulgação (arquivos de pronta entrega, sistemas de recuperação *on-line*, etc), bem como para as futuras análises a que servirão de base.

Na coleta das informações do Censo Demográfico de 1991, foram usados dois modelos de questionário: um questionário básico (denominado CD 1.01) aplicado nas unidades não selecionadas para a amostra contendo os quesitos referentes às características que foram investigadas para 100% das unidades domiciliares, e um questionário de amostra (denominado CD 1.02) contendo, além dos quesitos que também constam do questionário básico, outros quesitos mais abrangentes sobre características do domicílio e das pessoas, tais como: raça/cor, religião, nupcialidade, migração, escolaridade, fecundidade, mão-de-obra e rendimento.

Nos Censos Demográficos anteriores, as informações contidas nos dois questionários eram digitadas através de sistemas distintos, dando origem a dois arquivos magnéticos, um contendo os dados da amostra e o outro contendo os dados da não-amostra. Posteriormente, os dois arquivos eram agrupados para a formação de um arquivo único, sendo então acrescentadas aos registros as informações de controle necessárias, bem como o peso associado a cada registro pertencente à amostra para possibilitar o cálculo das estimativas.

Durante o planejamento do Censo Demográfico de 1991, com a finalidade de agilizar a divulgação dos dados referentes às características investigadas para 100% da população, decidiu-se priorizar a apuração das características comuns aos dois questionários.

A fim de garantir o processamento simultâneo dessas características foram "criados" questionários CD 1.01, para as unidades da amostra, a partir das informações investigadas nos questionários CD 1.02. Essa criação foi feita através de transcrição manual dos dados pelos recenseadores, ainda durante a fase de coleta dos dados.

Desse modo, as atividades subseqüentes de apuração passaram a ser planejadas de forma isolada para cada um dos questionários, o que implica em considerar que o Censo Demográfico de 1991 é composto por duas pesquisas independentes<sup>1</sup>. Este documento enfoca as especificações adotadas para o tratamento dos dados do questionário básico - CD 1.01.

Grande parte do que aqui vai se procurar reunir numa documentação mais unificada já foi objeto de documentos específicos, mas sua inclusão se justifica, como já mencionado, a fim de propiciar uma fonte de referência mais abrangente sobre o tratamento dos dados do CD 1.01 do Censo Demográfico de 1991.

---

<sup>1</sup> Ver Albieri e Bianchini (1993).

## 2. VISÃO GERAL DO PROCESSO DE APURAÇÃO DOS DADOS

A apuração dos Censos Demográficos, realizados anteriormente no Brasil, era executada de forma totalmente centralizada, desde a transcrição até a tabulação final dos dados.

A centralização dos processos de apuração<sup>2</sup> gerava a necessidade de contratação (temporária) de equipes numerosas, manuseando um grande volume de documentos. Como decorrência, diversos níveis de supervisores eram necessários para gerenciar as várias equipes de preparo, transcrição, conferência, codificação, etc.

Esse gigantismo nas equipes centralizadas de apuração acarretava problemas sérios na administração da produção e gerava diversos transtornos na época de liberação dos recursos humanos, ao término do Censo.

Outra característica típica da produção centralizada é a de que os processos voltados a grandes volumes quantitativos exigem controles operacionais sofisticados e quase não permitem paralelismo entre eles. Uma consequência da implantação de controles rígidos (necessários neste contexto) é a inflexibilidade da linha de produção e o resultado da linearidade adotada na execução dos processos é a dilatação dos prazos de apresentação dos resultados.

Com os objetivos de encurtar os prazos de produção das informações, evitando concentração excessiva de material e pessoal, e de melhorar a qualidade dos resultados finais, o Censo Demográfico de 1991 teve parte de sua apuração descentralizada nas diversas Unidades Regionais do IBGE.

O modelo de descentralização adotado foi testado no Censo Demográfico Experimental de Limeira, realizado em 1988. Os resultados obtidos foram positivos e, se, por um lado, confirmaram a viabilidade da proposta, por outro lado subsidiaram o aperfeiçoamento do modelo na implementação do sistema definitivo.

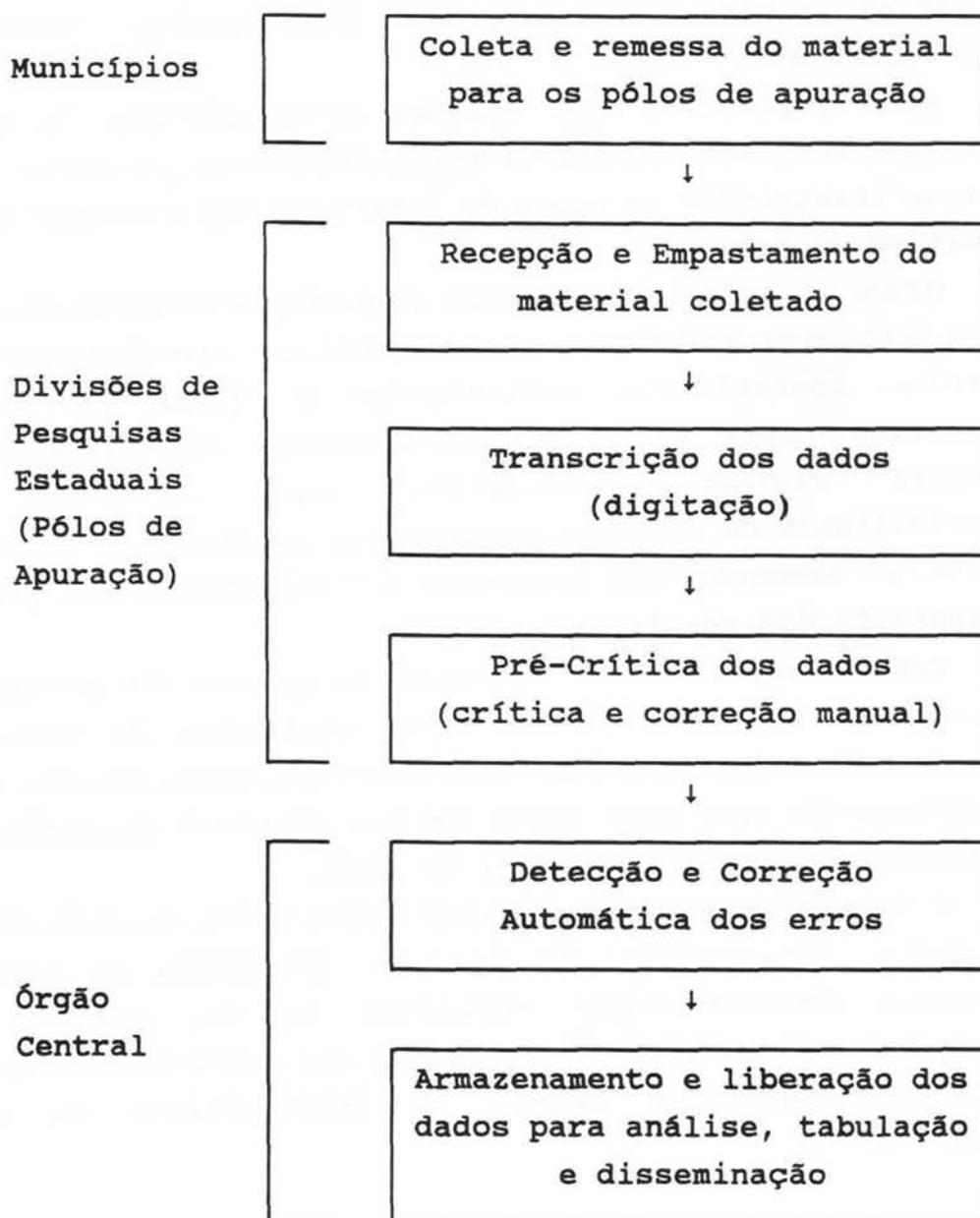
---

<sup>2</sup> Ver Mansoldo(1990).

O sistema de descentralização considerou a complexidade e o volume de trabalho de apuração do questionário básico - CD 1.01 e do questionário da amostra - CD 1.02.

As grandes etapas pelas quais passaram os dados coletados através do CD 1.01 podem ser visualizadas na figura 1.

**Figura 1** - Visão geral do processo de apuração do CD 1.01



O processo de apuração compreende as fases de recepção e empastamento do material, transcrição dos dados para o meio magnético (digitação), pré-crítica (crítica e correção manual de possibilidades, quantidade e parte das de incompatibilidade), detecção e correção automática dos erros e armazenamento e liberação dos dados para análise, tabulação e disseminação.

O CD 1.01 teve as fases de recepção e empastamento do material, transcrição dos dados e pré-crítica realizadas em 21 pólos de apuração localizados em todas as Unidades Regionais, exceto Rondônia que foi apurado no Mato Grosso do Sul; Acre, Roraima e Amapá que foram apurados em Mato Grosso; Amazonas no Espírito Santo; e Tocantins em Goiás.<sup>3</sup> Em seguida, os dados foram transmitidos para o computador central no Rio de Janeiro.

Na etapa centralizada foi realizada a fase de detecção e correção automática dos erros, antecedendo à liberação dos dados para análise, tabulação e disseminação.

Este documento descreve as especificações elaboradas para o tratamento dos dados investigados no CD 1.01 em cada uma dessas etapas.

É fato que a figura 1 fornece uma visão geral muito agregada do processo de apuração do CD 1.01. Por trás desse processo existe todo um sistema de controle e de execução das várias fases do qual derivam, em grande parte, as diversas informações quantitativas, bem como os períodos de execução apresentados nos capítulos posteriores.<sup>4</sup>

---

<sup>3</sup> A apuração do CD 1.01 das Unidades da Federação de Rondônia, Acre, Roraima, Amapá e Amazonas estava inicialmente prevista para ser realizada no Pará e foi deslocada para outras unidades face a disponibilidade de recursos em outros locais.

<sup>4</sup> O detalhamento sobre o sistema computacional montado para viabilizar, apoiar e executar o Censo Demográfico de 1991 encontra-se em Mansoldo(1990).

### 3. A RECEPÇÃO E O EMPASTAMENTO DO MATERIAL COLETADO

Durante a coleta do Censo Demográfico de 1991 foram utilizados os seguintes formulários:

- CD 1.01 - Questionário Básico<sup>5</sup>

- CD 1.02 - Questionário da Amostra

- CD 1.03 - Lista de Domicílio Coletivo

Utilizada para o arrolamento das famílias e das pessoas sós moradoras em domicílios coletivos e a seleção de algumas dessas unidades para a aplicação do CD 1.02.

- CD 1.06 - Caderneta do Setor

Utilizada para acondicionar as folhas de coleta e as listas de domicílio coletivo, além de conter o mapa e a descrição do setor censitário e o resumo das informações coletadas do setor.

- CD 1.07 - Folha de Coleta

Utilizada para a enumeração dos endereços, registro da utilização das unidades visitadas, numeração dos domicílios particulares e respectiva seleção para a aplicação do CD 1.02, numeração dos domicílios coletivos e nome do responsável pelos domicílios ocupados. Além disso, contém o registro do número de questionários preenchidos e de pessoas recenseadas.

- CD 1.10 - Boletim-Resumo

Utilizado para resumir os trabalhos de coleta de cada supervisão. Cada linha preenchida corresponde aos dados transcritos da capa da caderneta de cada setor.

O material coletado de cada supervisão era enviado ao agente de coleta municipal, ou para local estabelecido por esse, acompanhado do respectivo CD 1.10.

---

<sup>5</sup> Ver cópia do questionário no Anexo 1.

A Coordenação Técnica do Censo Demográfico - CTD deslocou dois grupos de técnicos, em meados de novembro de 1991, para algumas Unidades da Federação consideradas pólos de treinamento, a fim de que fossem explicitados todos os passos que deveriam ser seguidos durante a recepção, a conferência e o empastamento do material de coleta, cujas instruções encontram-se descritas em IBGE(1991a).

Desde o momento em que o material era recebido dos supervisores de coleta até o empastamento físico dos questionários existiram alguns passos importantes, pelos quais passaram os formulários (questionários, folhas de coleta, boletins-resumo) utilizados na coleta das informações. Um maior detalhamento pode ser alcançado recorrendo-se ao documento supra citado.

Primeiramente, os setores coletados e liberados pelos supervisores eram confrontados com a relação descrita no Boletim-Resumo - CD 1.10. A orientação foi para somente serem aceitos os setores acompanhados pelo CD 1.10; e no caso de falta de algum setor relacionado, a tarefa era paralisada a fim de que o técnico da rede de coleta responsável pela área resolvesse o problema.

Visando a divulgação dos totais registrados no CD 1.10, que originou a publicação **Resultados Preliminares**, foi realizado um trabalho de digitação preliminar chamado de "digitação em L", no qual digitava-se somente os números dos setores registrados no CD 1.10 e as linhas de total. Deste modo, foi possível exercer um controle na recepção dos setores coletados.

Concomitantemente à digitação preliminar do CD 1.10, os formulários utilizados na coleta de cada setor eram separados, por modelo, para facilitar o trabalho de conferência da quantidade de questionários (CD 1.01 e CD 1.02) com os registros correspondentes no resumo da capa da Caderneta do Setor - CD 1.06. Após a conferência das quantidades dos questionários, também eram conferidos os registros dos totais do CD 1.10 com o resumo do CD 1.06.

Uma vez realizado o processo de conferência era iniciada a digitação integral do CD 1.10. Esses arquivos com as informações digitadas serviram de base para permitir o processamento lógico

do empastamento e, posteriormente, utilizando as listagens produzidas, os questionários eram carimbados e empastados.

Os questionários CD 1.01 e CD 1.02 foram empastados em pastas distintas, já prevendo a independência na apuração dos dois questionários. Os demais formulários (CD 1.06, CD 1.07 e CD 1.03) foram acondicionados em outra série de pastas.

Durante o empastamento físico os questionários eram colocados nas pastas e simultaneamente eram colocadas etiquetas nas capas das pastas, com a indicação do material nelas contido. Cada pasta era identificada pelo código do modelo do questionário e numerada com 6 algarismos, sendo os dois primeiros correspondentes à unidade de produção<sup>6</sup> e os quatro últimos ao seqüencial da pasta na unidade de produção. Todos os questionários constantes de uma pasta foram carimbados com o número da pasta e com a numeração seqüencial dentro de cada pasta.

As pastas do CD 1.01 foram compostas por aproximadamente 600 questionários e as do CD 1.02 por 200 questionários, existindo uma margem de 10% de folga. A tabela 1 apresenta o número de pastas e de questionários por tipo de questionário para as Unidades da Federação.

Durante o processamento lógico do empastamento dos questionários foram emitidos os SLIPs<sup>7</sup>. Somente após todos os questionários de uma mesma pasta terem sido devidamente empastados eram preparadas as unidades de trabalho da digitação, denominadas tarefas de digitação, contendo cerca de 30 questionários CD 1.01 e de 10 questionários no caso do CD 1.02.

---

<sup>6</sup> A unidade de produção correspondeu a cada Unidade da Federação, com exceção de São Paulo que constituiu duas unidades de produção: uma constituída pelos municípios de Diadema, Guarulhos, Mogi das Cruzes, Osasco, Santo André, São Bernardo do Campo, São Caetano do Sul e São Paulo e a outra unidade de produção pelos demais municípios de São Paulo.

<sup>7</sup> SLIP é o nome dado ao documento utilizado para definir, separar e controlar os questionários de cada tarefa de digitação. O SLIP era impresso com vários campos de controle da digitação: modelo do questionário, município, número da pasta, distrito, subdistrito, setor e números dos questionários inicial e final que compunham a tarefa.

**Tabela 1 - Número de pastas e de questionários por tipo de questionário segundo as Unidades da Federação**

Unidades da Federação	CD 1.01		CD 1.02	
	Pastas	Questionários	Pastas	Questionários
TOTAL	65.067	35.452.415	27.819	4.031.788
Rondônia	458	258.808	188	26.875
Acre	163	89.774	73	9.834
Amazonas	760	395.228	373	45.768
Roraima	83	45.438	39	5.514
Pará	1.871	1.001.783	825	107.073
Amapá	113	54.233	52	6.079
Tocantins	387	194.961	244	29.808
Maranhão	1.851	990.970	831	106.459
Piauí	1.000	522.589	541	66.490
Ceará	2.511	1.356.758	1.122	151.238
Rio G. Norte	1.008	525.656	569	72.066
Paraíba	1.346	699.410	714	89.699
Pernambuco	3.000	1.603.580	1.249	172.833
Alagoas	1.019	532.110	496	61.572
Sergipe	634	332.384	328	42.139
Bahia	4.831	2.542.694	2.405	306.858
Minas Gerais	7.094	3.777.298	3.319	462.429
Espírito Santo	1.138	626.346	478	70.539
Rio de Janeiro	6.070	3.521.704	2.093	357.179
São Paulo	15.113	8.268.360	5.528	881.415
Paraná	3.810	2.123.294	1.673	249.258
Santa Catarina	2.095	1.141.353	966	141.046
Rio G. do Sul	4.465	2.537.288	1.832	292.745
Mato G. do Sul	841	441.853	358	52.972
Mato Grosso	870	474.167	431	60.918
Goiás	1.863	1.007.980	855	124.532
Dist. Federal	673	386.396	237	38.450

Fonte: Nº de pastas fornecido pela DI/DIDEM e nº de questionários obtidos do arquivo da Sinopse.

#### 4. A TRANSCRIÇÃO DOS DADOS

A transcrição dos dados dos questionários em papel para o meio magnético compreendeu a primeira fase propriamente dita do processo de apuração do CD 1.01.

##### 4.1 Planejamento da Transcrição dos Dados

Uma atividade importante da fase de planejamento da apuração dos dados do Censo Demográfico de 1991 foi a pesquisa de métodos e sistemas alternativos para o processo de transcrição dos dados. Tal pesquisa que envolveu a Divisão de Metodologia - DME, no âmbito do projeto de apoio metodológico ao Censo Demográfico de 1991<sup>8</sup>, iniciou-se em março de 1988, quando o chefe do então Núcleo de Metodologia participou de uma viagem de estudos ao *U.S. Bureau of the Census* para avaliação da possibilidade de aquisição de equipamentos e sistemas para a transcrição dos dados dos questionários em papel para meio magnético - sistema FOSDIC - *Film Optical Sensing Device for Input to Computers*, quando se decidiu pela não aquisição de tais equipamentos para a transcrição dos dados, por razões de custo, complexidade do sistema e por tratar-se de sistema já obsoleto, em vias de ser descontinuado pelo próprio *U.S. Bureau of the Census*. Tal viagem foi realizada em conjunto com técnicos da Diretoria de Informática, e seus resultados estão relatados em Silva(1988).

Paralelamente, a Diretoria de Informática vinha estudando o emprego de leitoras óticas de marcas, mas acabou desistindo da idéia, principalmente pelo fato de seu uso implicar em executar essa operação de forma centralizada, também em razão do custo e capacidade dos equipamentos necessários, e da necessidade de especialização do pessoal para sua operação.

A decisão a que se chegou foi que os dados seriam transcritos mediante digitação, em operação descentralizada, utilizando programas de entrada de dados em LTD (Linguagem de

---

<sup>8</sup> Ver Silva e Bianchini(1992).

Transcrição de Dados), com dois ciclos completos de digitação, com o segundo ciclo sendo denominado de verificação.

Foi apresentada uma proposta de emprego de controle de qualidade (com verificação por amostragem) da digitação do questionário básico - CD 1.01<sup>9</sup>. Tal proposta foi motivada pela avaliação prévia de que o esquema de transcrição proposto, com dois ciclos completos de digitação baseado na verificação a 100%, não incorporava nenhum mecanismo formal ou medida de qualidade que permitisse agir sobre o processo em busca da garantia da qualidade. Além disso, por implicar na digitação do material em dois ciclos completos, significaria gastar o dobro do tempo para digitar o material todo.

A proposta de adoção de um controle de qualidade por amostragem tinha dois objetivos básicos: adotar mecanismos formais de controle e medição da qualidade, que permitissem o gerenciamento do processo com uma ênfase na qualidade; e reduzir significativamente a carga de trabalho, o custo e o tempo necessário para a transcrição dos dados. Acreditava-se que com a digitação de uma massa de questionários equivalente a 1,2 vezes o número total de questionários (número indicado com grande sobra, para incluir a digitação das amostras de verificação) seria possível obter as medidas de qualidade desejadas, e ainda assim completar a operação em muito menos tempo, pois a redução do volume de trabalho seria da ordem de 40%.

Ainda dentro desse esforço de planejamento do controle de qualidade, foi executado um experimento preparatório de avaliação da qualidade da transcrição dos questionários CD 1.01 coletados durante o Censo Experimental de Limeira de 1988, com a redigitação desse material em 6 pólos regionais de apuração (Espírito Santo, Pernambuco, Rio de Janeiro, Rio Grande do Sul, Rio Grande do Norte e Pará<sup>10</sup>).

---

<sup>9</sup> Tal proposta, elaborada pela DME em conjunto com as equipes técnicas da Coordenação Técnica do Censo Demográfico e da Diretoria de Informática, acha-se descrita em Silva, Indá e Lima (1991).

<sup>10</sup> Os dados redigitados no estado do Pará não puderam ser aproveitados devido a problemas ocorridos com o sistema de digitação usado.

Nesse experimento foram considerados 37.407 questionários digitados 2 vezes por equipes independentes, adotando o mesmo sistema de digitação em dois ciclos completos: digitação e verificação. A taxa global de divergências entre campos<sup>11</sup> mostrou-se muito baixa, da ordem de 0,43%, porém o experimento serviu para revelar uma situação preocupante: foram detectadas divergências nos campos de identificação dos questionários (município.distrito.setor) com uma taxa de 0,284%, quando esses campos não poderiam conter erros devido à existência de mecanismos de controle na digitação (batimento contra os chamados SLIPs de digitação).

A taxa de divergência em campos abertos (como por exemplo idade, rendimento do chefe, etc.) foi de 0,82%, bem maior que nos campos pré-codificados, para os quais foi baixa (0,22%).

Mas se os números globais parecem suportar a idéia de uma boa qualidade da digitação, o exame dessas taxas separadamente para cada uma das pastas digitadas revelou o principal problema do processo, que é a falta de uma homogeneidade, de um padrão de qualidade. As taxas variaram bastante de uma pasta para outra, sem que o processo tivesse qualquer mecanismo de atuação sobre esse aspecto do processo. O gráfico apresentado no Anexo 2 ilustra essa situação.

Essa proposta de aplicação de controle de qualidade com verificação por amostragem, depois de bastante discutida em reuniões e num seminário realizado na Diretoria de Informática, foi apresentada à Comissão de Coordenação do Censo Demográfico, que em 17.04.91, a aprovou e decidiu pela sua adoção. Posteriormente, já iniciada a coleta de dados do Censo, foi anunciado pela Diretoria de Informática o abandono da idéia, por falta de tempo e de recursos humanos para operacionalizar a substituição do sistema anterior pelo novo, e o retorno do esquema de dupla digitação, que passou a ser empregado na transcrição dos dados.

---

<sup>11</sup> A taxa de divergências foi calculada como o número total de valores diferentes nos dois arquivos dividido pelo número total de valores comparados (iguais + diferentes).

## 4.2 Execução da Transcrição dos Dados

Para a etapa de digitação, os questionários CD 1.01 empastados foram subdivididos em tarefas de cerca de 30 questionários, as quais formaram as unidades de trabalho da digitação. Cada tarefa foi digitada por um digitador e verificada por um outro digitador distinto do primeiro.

Os dados foram digitados em dois ciclos completos, isto é, com verificação em 100% da digitação, sendo que os SLIPs foram digitados apenas uma vez, no primeiro ciclo. O primeiro ciclo de digitação de uma tarefa compreendia a digitação do SLIP, sendo transcritas as informações de controle que descreviam o conteúdo da tarefa, e em seguida, eram digitados os questionários pertencentes à tarefa.

O fluxograma da transcrição dos dados do CD 1.01 é apresentado na figura 2.

Uma das alterações mais significativas na estratégia global de apuração dos dados do Censo Demográfico de 1991<sup>12</sup> foi a aplicação de algumas regras de crítica fundamentais<sup>13</sup> durante a operação de entrada de dados, o que permite dizer que a entrada de dados do questionário básico do Censo Demográfico é do tipo "semi-inteligente". Foram introduzidas algumas verificações de identificação e de validade de códigos para os quesitos pré-codificados, evitando assim a possibilidade de introdução de alguma identificação ou códigos inválidos, como resultado de erros de digitação. O plano de crítica especificado para a etapa de transcrição dos dados encontra-se no Anexo 3.

A introdução dessas verificações não afetou significativamente a velocidade da operação de entrada de dados, porém auxiliou na redução do número de erros apontados nas etapas subsequentes de crítica dos dados.

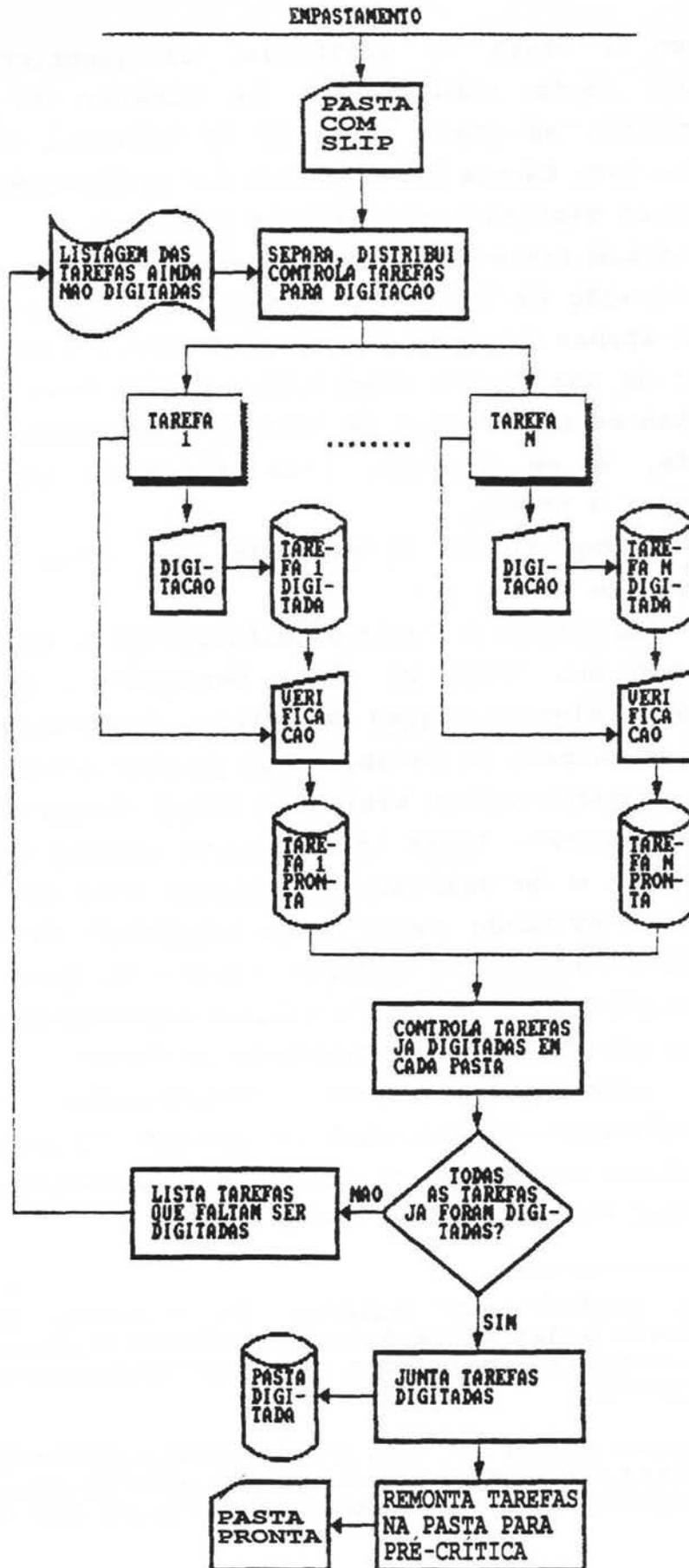
---

<sup>12</sup> Uma revisão dos métodos de crítica, dos sistemas computacionais e das estratégias de crítica e processamento dos dados das pesquisas realizadas pelo IBGE encontra-se em Silva e Bianchini (1994).

<sup>13</sup> Entende-se por crítica fundamental o processo através do qual o sistema só aceita a entrada dos valores digitados quando forem aceitáveis segundo as regras de crítica definidas.

FIGURA 2

FLUXOGRAMA DA TRANSCRIÇÃO DOS DADOS DO CD 1.01



Na digitação dos questionários propriamente ditos, quando eram digitados os quesitos do quadro 1 (identificação geográfica), era feita uma verificação dos campos município (V101), número da pasta (V102), distrito (V104), subdistrito (V105), e número do setor (V106), contra os valores digitados a partir do SLIP. Essa verificação foi prevista para assegurar a qualidade dos valores digitados nos campos de identificação dos questionários e também evitar a perda de dados durante a digitação (críticas de identificabilidade e cobertura). Os erros eventualmente detectados durante essa verificação causavam a interrupção da digitação da tarefa, que só prosseguia quando a inconsistência detectada fosse resolvida (acertos nos dados do SLIP ou do questionário).

Durante o segundo ciclo de digitação, o digitador não tinha acesso às informações que tinham sido digitadas durante o primeiro ciclo, e o processo era virtualmente idêntico ao da digitação inicial, a menos do fato de que não eram digitados novamente os dados do SLIP. Se durante essa segunda digitação fosse detectada alguma divergência em relação aos dados digitados no primeiro ciclo, o operador era alertado e o sistema não aceitava o dado.

Para que o programa de entrada de dados aceitasse o valor do campo que estava sendo redigitado era necessário que houvesse uma coincidência com o primeiro valor digitado. Caso não houvesse essa coincidência, o valor para aquele campo era novamente digitado e, para que esse último valor fosse definitivamente incorporado ao arquivo de dados, era necessário que coincidisse com qualquer um dos outros dois valores digitados anteriormente.

No caso dos quesitos do quadro 1 (identificação geográfica), o sistema somente aceitava os dados caso coincidissem exatamente com aqueles digitados durante a primeira rodada de digitação, ocasião em que foram comparados com os dados do SLIP e a partir da qual não podiam mais ser modificados durante a etapa de transcrição. Com relação aos demais quesitos do questionário, os dados aceitos e armazenados para as etapas seguintes da apuração eram aqueles resultantes da digitação durante o segundo ciclo (confirmando ou não a primeira digitação).

A transcrição dos dados do CD 1.01 foi realizada no período de outubro de 1991 a março de 1993. O período da execução dessa etapa para cada Unidade da Federação é apresentado no quadro 1 a seguir.

**Quadro 1 - Período da digitação do CD 1.01 por Unidade da Federação**

Unidades da Federação	Período	
	Início	Final
Rondônia	19.05.92	05.08.92
Acre	01.07.92	24.07.92
Amazonas	15.05.92	07.08.92
Roraima	08.06.92	01.07.92
Pará	17.02.92	02.10.92
Amapá	21.05.92	08.06.92
Tocantins	10.07.92	12.08.92
Maranhão	12.02.92	23.10.92
Piauí	09.01.92	08.06.92
Ceará	20.11.91	02.07.92
Rio Grande do Norte	01.12.91	20.07.92
Paraíba	09.01.92	04.08.92
Pernambuco	16.01.92	14.08.92
Alagoas	18.11.91	26.06.92
Sergipe	21.11.91	28.05.92
Bahia	05.03.92	16.10.92
Minas Gerais	20.01.92	07.12.92
Espírito Santo	21.11.91	05.06.92
Rio de Janeiro	13.01.92	12.03.93
São Paulo	16.01.92	19.02.93
Paraná	23.03.92	04.12.92
Santa Catarina	21.11.91	10.07.92
Rio Grande do Sul	19.02.92	07.02.93
Mato Grosso do Sul	18.11.91	22.05.92
Mato Grosso	29.10.91	23.06.92
Goiás	18.11.91	10.07.92
Distrito Federal	21.01.92	05.06.92

Cabe esclarecer que, no Pará, Bahia, Minas Gerais, Rio de Janeiro, São Paulo, Paraná e Rio Grande do Sul, a digitação foi iniciada com número de teclados bem inferior ao planejado para a operação censitária devido a problemas ocorridos na entrega dos equipamentos definitivos. E ainda que, Goiás até o final trabalhou com apenas um equipamento quando o número previsto era dois.

## 5. A PRÉ-CRÍTICA DOS DADOS

A operação de crítica e correção dos erros detectados nos dados investigados no CD 1.01 envolveu duas fases de trabalho, uma manual e outra automática, tal como na apuração do Censo Demográfico de 1980.

A pré-crítica correspondeu à primeira fase da crítica e correção dos dados do CD 1.01, durante a qual a correção dos erros detectados era feita manualmente.

É importante registrar que, apesar da semelhança na divisão dos trabalhos de apuração no Censo 91 em relação ao Censo 80, os procedimentos adotados nas fases de crítica e correção dos dados foram diferentes. Isto significa dizer que enquanto no Censo 80 a fase de pré-crítica foi realizada de modo centralizado, em 1991 introduziu-se o processo de descentralização da atividade de pré-crítica em 21 Unidades da Federação, seguindo as orientações fornecidas pela Coordenação Técnica do Censo Demográfico - CTD; quanto à fase de correção automática realizada de forma centralizada, houve um grande avanço metodológico em relação a 1980, com a utilização da metodologia de Fellegi e Holt<sup>14</sup> para detecção e correção dos erros, cuja aplicação foi feita através do SISTEMA DIA<sup>15</sup> - *Detección e Imputación Automática de errores para datos cualitativos* desenvolvido pelo INE - Instituto Nacional de Estadística da Espanha.

Um ponto importante a destacar diz respeito à decisão de quais variáveis deveriam ser corrigidas de forma manual ou automática. Um princípio que norteou a equipe responsável pela elaboração dos planos da pré-crítica e das críticas para a detecção dos erros para a submissão à correção automática foi o de definir o mínimo possível de críticas de incompatibilidade ou de inconsistência a serem analisadas na etapa descentralizada.

As críticas definidas na fase de pré-crítica compreenderam as de possibilidades onde eram verificados se os valores digitados correspondiam a códigos válidos para cada quesito

---

<sup>14</sup> Maiores detalhes no item 6.2 deste texto.

<sup>15</sup> Ver Rubio e Criado(1988) e item 6.3 deste texto.

investigado, incluindo o preenchimento de quesitos obrigatórios; as críticas de quantidade onde era feita a conferência da seqüência da identificação, do número dos questionários dentro de cada pasta, do número da pessoa dentro de cada questionário, dos totais de pessoas; algumas críticas de incompatibilidade que dizem respeito basicamente à verificação de inconsistências entre registros; e outras cuja correção dependia do acesso aos instrumentos de coleta (questionários, folhas de coleta, etc).<sup>16</sup>

O tratamento dos erros detectados pelas demais críticas de incompatibilidade foi realizado na etapa centralizada, já que no processo automático de correção, os erros seriam acertados mediante critérios homogêneos, sem a interferência de pessoas com decisões distintas. No capítulo 6 encontra-se a descrição do processo de detecção e correção automática dos erros nos dados investigados no CD 1.01.

#### **5.1 Variáveis Tratadas na Pré-Crítica**

As inconsistências entre registros foram tratadas na fase de pré-crítica, uma vez que não podiam ser tratadas pelo sistema DIA, cuja metodologia só leva em conta as relações existentes entre variáveis de um mesmo registro.

##### **V111 (total de homens), V112 (total de mulheres) e V301 (sexo)**

Os valores atribuídos a essas variáveis foram confrontados e corrigidos manualmente de forma a ficarem consistentes entre si, tendo sido relevante a consulta aos instrumentos de coleta para analisar o nome das pessoas registradas como subsídio para a correção.

##### **V201 (espécie do domicílio), V109 (nº no CD 1.07), V110 (nº no CD 1.03) e V302 (parentesco ou relação com o chefe do domicílio)**

Optou-se por corrigir essas variáveis conjuntamente na pré-crítica pois, além de tratar de inconsistências entre registros,

---

<sup>16</sup> Ver Plano de pré-crítica do CD 1.01 no Anexo 4.

o acesso aos questionários e a outros instrumentos de coleta era necessário para dirimir as dúvidas existentes.

Outras variáveis tratadas na pré-crítica cujo acesso aos instrumentos de coleta era fundamental:

**V303/304 (mês e ano de nascimento/idade presumida)**

Levando-se em conta a relevância dessas variáveis para a obtenção da idade calculada, optou-se pelo tratamento na pré-crítica pelos seguintes fatos:

- caso houvesse erro de digitação ou de preenchimento, a correção seria imediata, já que naquele momento o operador tinha acesso aos questionários;

- caso fosse necessário excluir registro de pessoas incluídas indevidamente (nascidas após a data de referência), somente seria possível durante a pré-crítica.

**Preenchimento da V305 (sabe ler e escrever) e da V3083 (rendimento mensal bruto)**

O preenchimento da V3083 dependia do preenchimento da variável V305. A crítica que envolve essas variáveis visava detectar possíveis erros do tipo "não digitação" da V3083 por ser a última variável da frente do questionário.

Nessa fase a variável V3083 foi criticada através da detecção de possíveis inconsistências entre o valor do rendimento e o nº de dígitos. A correção manual da variável rendimento obedeceu os seguintes critérios:

- caso o campo da variável estivesse erroneamente em branco, o operador deveria verificar no questionário se havia alguma observação indicando que o valor do rendimento era superior ao tamanho do campo; se assim fosse, o critério adotado foi preencher o campo nº de dígitos com código 7 e o campo valor do rendimento com 9999997, considerado o maior valor permitido.

- caso não existisse nenhuma observação e, de acordo com a V305, devesse estar preenchido, o critério adotado foi registrar no campo nº de dígitos o código 7 e 9999998 (rendimento ignorado) no campo valor do rendimento mensal bruto.

Vale dizer que o código 9999998 foi considerado código ignorado apenas nessa fase de correção. Após a fase de correção automática o código de ignorado foi alterado para 9999999.

Essa mudança teve origem na retirada de uma advertência do plano de pré-crítica, que servia para o operador verificar se aquele valor (9999999) era realmente um valor de rendimento ou um erro de digitação. Durante a elaboração do plano, o grupo responsável pela definição das críticas julgou desnecessária a permanência de advertências nessa fase de correção, e mais apropriado na fase de transcrição de dados, já que esse tipo de crítica funciona apenas como uma proteção contra possíveis erros de digitação.

Ao retirar a advertência do plano de pré-crítica, o valor 9999999 foi considerado um valor inválido para que fosse possível ao operador acertar a variável, e o valor 9999998 foi considerado, naquela fase, como rendimento ignorado.

## **5.2 Criação de Questionários Faltosos**

Considerando a possibilidade de ocorrência de faltas de questionário por ocasião do empastamento, durante os trabalhos de elaboração do plano de pré-crítica foi discutida a possibilidade de recuperação de algumas informações como total de pessoas por sexo, nº no CD 1.03 e espécie do domicílio, deixando os dados das outras variáveis para serem imputados posteriormente.

Assim sendo, foi decidido colocar código 9999 na V109 (nº no CD 1.07) nos questionários desaparecidos no empastamento e o seguinte tratamento para ser executado durante a pré-crítica:

- caso o questionário fosse encontrado, seria alterado o código 9999 para o valor correspondente ao lançamento da Folha de Coleta (CD 1.07) e digitados os dados relativos aos quesitos do questionário;

- caso o questionário não fosse encontrado, o código 9999 da V109 deveria ser alterado para 9996 e os campos das variáveis V110 (nº no CD 1.03), V111 (total de homens), V112 (total de mulheres) e V201 (espécie do domicílio), seriam preenchidos com

base nos registros das Folhas de Coleta. Além disso, foi decidido que iria digitar um campo para cada pessoa existente no domicílio, preenchendo-se apenas a V301 (sexo), de modo que ficasse coerente com o número de pessoas, por sexo, registrado nas V111 e V112. Na verdade, o código 9996 na V109 funcionou como uma "marca" desses questionários faltosos para, posteriormente, serem imputados após a execução da correção automática pelo processo DIA.

### **5.3 Execução da Pré-Crítica**

A execução da pré-crítica foi efetivada após a realização de um período de treinamento descentralizado, ministrado pelos técnicos da CTD nos períodos de 25 a 29 novembro de 1991 e 27 a 30 janeiro de 1992, direcionado aos coordenadores de apuração das Unidades da Federação que, posteriormente, repassaram as instruções para as equipes responsáveis pela execução da operação.

Na fase de pré-crítica foi considerado como lote de trabalho cada pasta de questionários. A operacionalização dos trabalhos se deu através da elaboração de um programa de computador feito sob medida, que emitia listagens de erro contendo o número da pasta, do questionário e da pessoa com erro, para que o operador fizesse as correções manualmente, através do manuseio de alguns instrumentos de coleta, como os questionários (CD 1.01), as folhas de coleta (CD 1.07) e as listas de domicílio coletivo (CD 1.03) e tomando como base as instruções elaboradas pela equipe do Projeto de Acompanhamento e Controle da Apuração Descentralizada da CTD e apresentadas em IBGE(1991b).

Os operadores da pré-crítica recebiam listagens indicando as situações de erro detectadas e os respectivos instrumentos de coleta e preparavam os acertos nas listagens, e posteriormente via *on-line* indicavam as correções que deveriam ser feitas em cada caso para solucionar os problemas detectados em cada pasta.

O início da execução da pré-crítica ocorreu a partir do momento em que já se dispunha de algumas pastas digitadas de cada Unidade da Federação. O período correspondente à execução da pré-

crítica foi de janeiro de 1992 a maio de 1993, sendo que dentre as 27 Unidades da Federação, os trabalhos de Minas Gerais, São Paulo, Rio de Janeiro e Rio Grande do Sul foram concluídos em 1993. A data do final da execução da fase de pré-crítica para cada Unidade da Federação é apresentada no quadro 2.

**Quadro 2** - Data do final da execução da pré-crítica do CD 1.01 por Unidade da Federação

Unidades da Federação	Final da pré-crítica
Rondônia	07.08.92
Acre	31.07.92
Amazonas	02.10.92
Roraima	10.07.92
Pará	23.10.92
Amapá	31.07.92
Tocantins	21.08.92
Maranhão	20.11.92
Piauí	03.07.92
Ceará	17.07.92
Rio Grande do Norte	31.07.92
Paraíba	07.08.92
Pernambuco	28.08.92
Alagoas	03.07.92
Sergipe	12.06.92
Bahia	13.11.92
Minas Gerais	08.01.93
Espírito Santo	12.06.92
Rio de Janeiro	19.03.93
São Paulo	07.05.93
Paraná	04.12.92
Santa Catarina	17.07.92
Rio Grande do Sul	16.04.93
Mato Grosso do Sul	22.05.92
Mato Grosso	10.07.92
Goiás	31.07.92
Distrito Federal	19.06.92

Considerando a estreita ligação existente entre as especificações da pré-crítica e da correção automática, todas as inconsistências verificadas durante a pré-crítica foram solucionadas a fim de que permitisse a execução normal da fase seguinte, a de correção automática. Com isto, algumas variáveis foram corrigidas na primeira fase manualmente e, posteriormente, tornaram-se fixas durante os trabalhos da segunda fase de correção, visando assim, o não comprometimento daquilo que já havia sido corrigido e, conseqüentemente, a manutenção da coerência interna dos questionários sem recorrer a ciclos de acertos.

As variáveis tratadas na pré-crítica V201 (espécie do domicílio), V301 (sexo), V302 (parentesco ou relação com o chefe do domicílio) e V7100 (total de pessoas) foram envolvidas nas regras de crítica para a detecção dos erros para serem tratados na correção automática, porém ficaram fixas. As demais variáveis tratadas na pré-crítica V109 (nº no CD 1.07), V110 (nº no CD 1.03), V111 (total de homens), V112 (total de mulheres) e V303/304 (mês e ano de nascimento/idade presumida) não foram envolvidas nas regras de crítica para a detecção dos erros para serem tratados na correção automática.

## **6. A DETECÇÃO E A CORREÇÃO AUTOMÁTICA DOS ERROS**

Após a liberação da pré-crítica os dados eram transmitidos para o Órgão Central no Rio de Janeiro visando a execução da etapa centralizada da apuração.

Os dados transmitidos passavam primeiramente por um processo de **consolidação**, que tinha por objetivo validar as pastas transmitidas após a realização das atividades descentralizadas. Era então verificado se o total de registros (domicílios e pessoas) transmitidos estava compatível com a informação da Sinopse para o setor e, se o total de pessoas registrado para cada domicílio estava compatível com as pessoas pertencentes ao domicílio em questão. Além disso, era feita uma validação da crítica descentralizada, através da execução de um programa, que rodava no computador central, análogo ao programa executado na fase de pré-crítica.

Após a consolidação dos dados foram iniciados os preparativos para a fase de detecção e correção automática dos erros. Essa fase de correção, como última, traz consigo a necessidade de minimizar os erros que não foram tratados anteriormente tendo como duas funções básicas a agilização de todo o processo de apuração e a uniformização dos critérios de correção.

### **6.1 A Motivação para a Busca de uma Nova Metodologia**

O processo de correção automática de erros de consistência foi implantado no IBGE no Censo Demográfico de 1980, tendo sido objeto de experimentação no Censo Demográfico de 1970, somente nos questionários da não-amostra.

A utilização de um método de correção automática nos dados do Censo Demográfico de 1980 teve os seguintes objetivos:

- economia de tempo no processamento dos dados, eliminando o trabalho de pessoas especialmente treinadas para efetuar as correções necessárias nos questionários rejeitados; e
- obtenção de uma maior homogeneidade nos critérios de correção das inconsistências verificadas.

A realização dessa fase foi baseada num plano de crítica abrangendo todos os quesitos do questionário, passíveis de correção automática, e na definição de um conjunto de procedimentos lógicos, com o objetivo de solucionar cada situação em que alguma regra de incompatibilidade do plano fosse contrariada, obtendo-se assim coerência entre os quesitos pertencentes a um mesmo questionário.

Na definição dos procedimentos de correção automática do Censo Demográfico de 1980, foram utilizados dois métodos: a correção determinística e o método da Matriz Dinâmica, com variáveis de restrição (*hot-deck*)<sup>17</sup>.

O método da matriz dinâmica com variáveis de restrição foi a correção mais empregada. Esse procedimento teve como base a construção de matrizes formadas pelo cruzamento de variáveis pertinentes ao quesito a ser corrigido e que eram continuamente atualizadas com as informações anteriormente consideradas corretas. Sendo assim, procurou-se respeitar a distribuição de freqüências dos dados relativos aos registros "bons" (sem erros).

O método determinístico foi utilizado quando a variável a ser imputada só podia assumir um único código, mantendo assim a consistência com os demais quesitos relacionados.

A preocupação existente no processo de correção automática é a não geração de novos erros após a imputação, mas infelizmente, em alguns casos, isso não foi possível em 1980, já que a metodologia empregada não permitia assegurar, *a priori*, que as correções efetuadas num dado registro deixariam o mesmo livre de novas inconsistências. Como conseqüência, os lotes de apuração tiveram que passar pelo sistema de crítica e correção mais de uma vez, criando assim um ciclo de processamento indesejável numa pesquisa do porte do Censo.

Outro problema decorrente da metodologia adotada no Censo 80 foi a elevada complexidade do trabalho de desenvolvimento do sistema computacional implicado pelo plano de crítica e correção, face à inexistência de ferramentas de automatização do

---

<sup>17</sup> Ver Metodologia do Censo Demográfico de 1980.

desenvolvimento, não disponíveis no IBGE naquela época, e à natureza exploratória do processo de teste dos programas desenvolvidos.

Uma das preocupações fundamentais da Divisão de Metodologia, desde a sua implantação em 1987, com respeito à metodologia de pesquisa tradicionalmente adotada no IBGE se relaciona aos métodos e sistemas de apuração das pesquisas, particularmente devido à ineficiência em termos de prazos de desenvolvimento de sistemas, custo e prazos de execução da crítica e insegurança quanto à qualidade dos dados obtidos.

Com essa preocupação em mente, foram pesquisadas alternativas de metodologia e sistemas (*software*) para apuração de pesquisas, particularmente para correção ou imputação automática de dados, que pudessem ser aplicadas no processamento dos dados do Censo Demográfico de 1991. Além disso, uma avaliação da experiência do Censo anterior revelou que, se o processo pode ter sido considerado à época como um sucesso e tenha representado um avanço no uso de métodos de correção automática no IBGE, nos dias de hoje os métodos então empregados seriam de difícil aplicação, devido tratar-se de métodos extremamente custosos em termos de recursos para a especificação, o desenvolvimento, o teste e validação dos sistemas de correção automática<sup>18</sup>.

Uma nova metodologia deveria ser capaz de solucionar os problemas ocorridos no Censo de 80 preservando, ao mesmo tempo, as vantagens alcançadas. Esse estudo conduziu a um exame mais detalhado das experiências vivenciadas pelos países mais desenvolvidos no tratamento dessa questão e, a um aprofundamento da experiência do Canadá, iniciada a partir do trabalho pioneiro de Fellegi e Holt(1976) e da Espanha, a partir do trabalho de Rubio e Criado(1988), também baseado na mesma metodologia. Essa nova metodologia para detecção e correção automática de erros foi apresentada, em meados de junho de 1989, pelo técnico Pedro Luis do Nascimento Silva da DME, na palestra "Metodologia de Fellegi e Holt para crítica e correção de dados qualitativos ou categóricos".

---

<sup>18</sup> Um relato dessa avaliação e do estudo de alternativas pode ser encontrado em Silva e outros(1990).

O resultado dessa avaliação e da busca de alternativas conduziu ao estudo mais aprofundado de um sistema generalizado para detecção e correção automática de erros em dados qualitativos, desenvolvido por Rubio e Criado(1988) do Instituto Nacional de Estadística - INE - da Espanha, denominado DIA - *Detección e Imputación Automática de errores para datos cualitativos*.

Tal sistema foi cedido ao IBGE em maio de 1989 para estudo e testes, e foi imediatamente aplicado para a realização de um experimento de processamento dos dados do CD 1.01 coletados durante a realização do Censo Experimental de Limeira de 1988. O sucesso desse experimento foi evidente, com a tarefa de processamento executada com o novo sistema sendo concluída um mês e meio após seu início, muito antes do que foi possível conseguir com o sistema tradicional. Tal experimento foi conduzido por uma equipe de apenas 4 técnicos (2 da DME e 2 da equipe da CTD).

Apesar de alcançados resultados tão animadores, que foram descritos e analisados com detalhes em Silva e outros(1990), a aceitação da idéia do uso de um sistema generalizado para processar os dados do Censo só veio mais tarde, depois que técnicos da Diretoria de Informática e da Diretoria de Pesquisas (DME e CTD) visitaram o INE-Espanha<sup>19</sup>. Além disso, foi necessário um grande esforço de convencimento na Instituição, através da realização de seminários e programas de capacitação, bem como a replicação do experimento já realizado com os dados de Limeira pela equipe da Diretoria de Informática. Tanto cuidado se justificava pela importância da decisão que deveria ser tomada, de usar ou não o sistema DIA para processar os dados do Censo Demográfico de 1991.

Finalmente, em julho de 1990, após realizados todos os testes e estudos demandados<sup>20</sup>, e discutidos todos os aspectos considerados relevantes da questão, foi decidida, a nível técnico, a utilização do sistema DIA para o processamento do

---

<sup>19</sup> Ver Hanono, Miranda e Ferreira(1990) e Oliveira e Oliveira(1990).

<sup>20</sup> Ver IBGE(1990).

CD 1.01 do Censo Demográfico de 1991, mas a decisão gerencial só foi tomada em 1991.

Houve uma grande interação entre os técnicos do IBGE que utilizavam o sistema e os do INE responsáveis por seu desenvolvimento, tendo havido três oportunidades em que os técnicos do INE visitaram o Brasil (maio de 1989, fevereiro de 1990 e setembro de 1991<sup>21</sup>), bem como as oportunidades em que técnicos do IBGE visitaram aquele instituto (janeiro de 1990<sup>22</sup>, julho de 1993<sup>23</sup>). Como resultado dessa interação foram introduzidas diversas melhorias e correções, resultando na elaboração da versão 2 do sistema DIA, instalada no IBGE e que foi empregada no processamento dos dados do CD 1.01.

## **6.2 A Metodologia Proposta por Fellegi e Holt**

Nesta seção apresenta-se resumidamente a metodologia proposta por Fellegi e Holt(1976) para a detecção e correção automática de erros, em sua versão para o tratamento de dados categóricos ou qualitativos<sup>24</sup>.

Em seu artigo, Fellegi e Holt(1976) conseguiram dar um tratamento matemático rigoroso à questão da análise de um conjunto de regras de crítica definidas para uma pesquisa, que implicou na possibilidade de integrar perfeitamente o processo de detecção e correção de erros, em apenas um único ciclo de processamento, sem a geração de novos erros, ou seja, não havendo reimputação.

Isso significa dizer que quando um erro é detectado num registro, alguma correção ou alteração é introduzida sobre os

---

<sup>21</sup> Ver Silva e Hanono(1991a,b).

<sup>22</sup> Ver Oliveira e Oliveira(1990) e Hanono, Miranda e Ferreira(1990).

<sup>23</sup> Ver Barbosa(1993).

<sup>24</sup> Uma descrição mais detalhada pode ser encontrada em Silva e outros(1990) e Rubio e Criado(1988).

dados desse registro, de modo a torná-los consistentes, não admitindo saídas do tipo "liberar" o registro sem correção.

Os princípios básicos, no qual se baseia a metodologia, são os seguintes:

- os dados de cada registro devem satisfazer todas as críticas, alterando o menor número possível de variáveis, procurando assim, manter a maior quantidade possível de informação original;

- não se faz necessário especificar critérios de imputação (ao contrário do processo *hot-deck*), os quais são deduzidos automaticamente das regras de crítica, o que assegura que os registros imputados não estarão em condição de erro;

- o processo de imputação procura manter, na medida do possível, as distribuições de frequências marginais e conjuntas das variáveis, baseando-se nas distribuições de dados "bons" (aqueles que passaram pelas críticas sem serem rejeitados).

Para sua implementação é necessário escrever as regras de crítica na **forma normal**<sup>25</sup>, as quais são denominadas críticas explícitas. A partir dessas regras de crítica, pode-se deduzir logicamente outras regras, as quais são denominadas críticas implicadas. Dentre essas regras de crítica só são de interesse as regras de crítica implicadas essencialmente novas que em conjunto com as regras inicialmente especificadas constituem o **conjunto completo das regras de crítica**.

Fellegi e Holt demonstraram, mediante um lema, que sempre se pode obter um conjunto completo de regras de crítica por meio de um algoritmo.

Dispondo do conjunto completo das regras de crítica, para um registro detectado como inconsistente ou errôneo, basta selecionar para ser imputado ao menos um campo ativo<sup>26</sup> em cada crítica do conjunto completo que tenha rejeitado o registro. A

---

<sup>25</sup> Uma regra de crítica é escrita na forma normal se for da seguinte forma:

VAR 1(subconjunto de códigos da VAR 1) e ... e  
VAR K(subconjunto de códigos da VAR K) ⇒ ERRO.

<sup>26</sup> Campo ativo ou explícito é aquele que é incluído na especificação de uma regra de crítica.

seleção se faz empregando o princípio da menor modificação possível dos dados originais, escolhendo o menor número possível de variáveis a modificar em cada registro inconsistente. Para um registro inconsistente, o menor conjunto de variáveis que contemplam as variáveis ativas nas críticas do conjunto completo que rejeitaram esse registro é chamado de conjunto mínimo de campos a imputar.

Uma vez selecionadas as variáveis a serem imputadas em um certo registro inconsistente, é feita a seleção do código a imputar em cada uma dessas variáveis, a partir do dicionário de valores válidos. Fellegi e Holt demonstraram um teorema no qual é sempre possível encontrar ao menos um código para cada variável selecionada de forma que imputando tal código o registro fique "corrigido" e não seja mais rejeitado por nenhuma das regras de crítica do conjunto completo.

### 6.3 O Sistema DIA

O Sistema DIA - *Detección e Imputación Automática de errores para datos cualitativos* trata-se de um pacote computacional desenvolvido por técnicos do Instituto Nacional de Estadística - INE da Espanha, baseado na metodologia de Fellegi e Holt, que tem objetivo facilitar a depuração de Censos e grandes pesquisas estatísticas.

O sistema DIA permite não só a utilização de correção probabilística sugerida na metodologia de Fellegi e Holt, como também a de correção determinística para ser usada no tratamento de erros sistemáticos, e ainda garante uma consistência entre as críticas de incompatibilidade e as críticas para imputação determinística.

O DIA é um sistema para ser operado em ambiente centralizado, tipo *mainframe* IBM e trata somente de arquivos seqüenciais zonados com um único tipo de registro de tamanho fixo com variáveis categóricas ou qualitativas<sup>27</sup>.

---

<sup>27</sup> Ver Hanono(1993).

O sistema DIA é constituído de dois subsistemas, o de especificação e o de tratamento, cada um formado por vários módulos. Além disso, fornece um conjunto amplo de tabelas com informações para subsidiar a análise do processo de detecção e correção automática dos erros.

### **6.3.1 Subsistema de especificação**

O subsistema de especificação consta de 4 etapas:

#### **■ Geração do dicionário interno a ser utilizado pela aplicação**

Para a execução do sistema DIA são necessários os seguintes arquivos de entrada:

- FIVAR: arquivo que contém a relação das variáveis e dos códigos possíveis do registro a ser processado;
- FIPOLLO: arquivo que contém a posição e o tamanho para cada variável listada no arquivo FIVAR;
- EDITS: arquivo que contém o conjunto de regras de crítica, apenas intra-registros e expressas na forma normal, definidas pelos analistas temáticos;
- RIDs: arquivo que contém o conjunto de regras de crítica para a imputação determinística expressas na forma normal. A criação desse arquivo é opcional, sendo realizada em pesquisas em que se verifica a ocorrência de erros sistemáticos nos dados.

#### **■ Análise de consistência do conjunto de regras de crítica**

O analisador de regras é um componente fundamental do sistema DIA. Tem como objetivo garantir que a aplicação do sistema cumpra os requisitos de consistência estabelecidos. As críticas analisadas são as explícitas (descritas pelos analistas temáticos) e aquelas derivadas de RIDs (EDR - Edits Derivadas de RIDs).

Em primeiro lugar o analisador elimina as redundâncias do conjunto de críticas, e em seguida detecta e lista as

inconsistências existentes. O processo de análise só encerra quando não há mais conflito entre as críticas.

#### ■ **Geração do conjunto completo de regras de crítica**

Apesar do algoritmo proposto por Fellegi e Holt para a obtenção do conjunto completo de regras de crítica ser conceitualmente simples, a dificuldade prática está na implementação desse algoritmo, de forma eficiente em computador. A geração do conjunto completo de regras de crítica, em muitos casos, pode consumir muito tempo de computação e implicar em arquivos de difícil manuseio.

O sistema DIA utiliza um teorema de Truncamento para a obtenção do conjunto completo de regras de crítica, que interrompe o processo evitando que se obtenha o verdadeiro conjunto completo, mas que origina uma proxy desse conjunto com a garantia de que toda inconsistência será detectada.

#### ■ **Definição da estratégia para o tratamento dos dados**

Neste módulo é feita a especificação da estratégia a ser considerada para o tratamento dos dados, onde são definidas as variáveis fixas, o peso de acordo com o grau de confiança da variável e o critério de seleção do código a imputar para cada variável. Para tanto são adotadas as seguintes definições e/ou conceituações:

##### **Variável fixa**

Uma variável é marcada como fixa quando aparece nas regras de crítica apenas para indicar conflitos existentes com outras variáveis mas não para ser imputada, exceto se tiver algum código inválido.

As variáveis consideradas fixas são variáveis já depuradas que não devem ser modificadas.

### **Peso de uma variável**

Dada uma variável não fixa, o peso atribuído à variável é uma medida do grau de desconfiança que se tem da informação relativa a ela e representa uma suspeita subjetiva da variável. O peso desempenha papel importante na escolha da variável a ser imputada. A variação do peso é de 1 a 10 sendo que quanto maior o seu valor maior a desconfiança da variável, e portanto, maior a chance da variável ser imputada.

### **Critério de seleção do código a ser imputado em uma variável**

A imputação em uma variável é feita através da seleção de um código. Caso haja vários candidatos, essa seleção é probabilística em função de uma determinada distribuição de freqüências que o DIA procura ajustar. Essa distribuição pode ser marginal ou conjunta de acordo com a especificação definida para cada variável.

Uma variável utiliza distribuição marginal se a imputação for realizada sem levar em conta os códigos das demais variáveis do registro em tratamento.

Uma variável utiliza distribuição conjunta se a imputação for realizada levando em conta a distribuição conjunta com outra(s) variável(is) do registro em tratamento.

O sistema DIA possui duas maneiras de calcular a distribuição de freqüências de uma variável. A primeira, denominada FNS (freqüência de não suspeitos padrão) e a segunda FNS' (freqüência de não suspeitos não padrão).

Para o cálculo da distribuição de freqüências do tipo FNS são contabilizados todos os registros sem erro ("bons") e aqueles registros errôneos, em que a variável é não suspeita<sup>28</sup>.

Para o cálculo da distribuição de freqüências do tipo FNS' são contabilizados todos os registros sem erro ("bons") e aqueles registros errôneos em que a variável é não suspeita após a atuação de RIDS, ou seja, após a correção dos erros sistemáticos.

Além da especificação da distribuição de freqüência marginal ou conjunta e do tipo de freqüência FNS ou FNS' é ainda

---

<sup>28</sup> Variável suspeita é a variável que possui algum valor inválido ou participa em alguma inconsistência detectada.

necessária uma outra especificação que diz respeito ao método de seleção do código. O método recomendado é o proporcional, mas existe uma alternativa no sistema DIA onde a seleção do código pode ser efetuada pela máxima diferença na distribuição de frequência de variáveis com imputação marginal. Este era o método adotado na versão 1, e só foi conservado na versão 2 para possibilitar a compatibilização das duas versões.

A seleção do código pelo método proporcional é feita do seguinte modo: é selecionado aleatoriamente um número dentro do intervalo de variação da função de distribuição adotada e, então, o código selecionado é aquele tal que a função distribuição é maior ou igual ao número aleatório.

### **6.3.2 Subsistema de tratamento dos dados**

O subsistema de tratamento dos dados é realizado pelos módulos de detecção e de imputação.

#### **■ Detecção dos registros com erro**

Para a detecção dos registros com erro o sistema utiliza o dicionário, que contém as variáveis com os respectivos códigos válidos, o conjunto das regras de incompatibilidade e as críticas derivadas de RIDs.

Um registro é considerado com erro quando:

- alguma de suas variáveis tem um código inválido; e/ou
- falha em alguma das regras de crítica ou alguma RID.

O módulo de detecção separa e gera arquivos dos registros sem erro ("bons") e dos registros com erro ("maus"), identificando nesses os campos a corrigir e as regras que falharam.

#### **■ Imputação determinística dos erros**

Caso haja regras de imputação determinística (RIDs) é necessário aplicar o módulo de imputação determinística aos registros "maus".

Para cada registro com erro o sistema verifica a ocorrência de falha em alguma RID e, conforme o caso, efetua as imputações especificadas pelas RIDs que falharam.

A imputação é dita **Imputação Determinística Única (IDU)** se existir um único código a ser atribuído ao registro que falhou. Por outro lado, se existem vários códigos possíveis a variável é pendente de seleção probabilística de um de seus códigos e diz-se que a variável é pendente de **Imputação Determinística Flexível (IDF)**.

#### ■ Imputação probabilística dos erros

Para os registros que falharam em alguma regra de crítica aplica-se o módulo de imputação probabilística que executa duas tarefas:

- seleciona o conjunto de variáveis a modificar para cada registro com erro; e
- seleciona o código a ser imputado para cada variável a modificar.

#### **Seleção das variáveis a imputar**

Para cada registro com erro são verificadas as críticas do conjunto completo que falharam. Dessas críticas o sistema verifica o conjunto de variáveis ativas e seleciona o conjunto de variáveis para serem imputadas.

Na versão 1 do sistema DIA, o conjunto selecionado era o de menor número de variáveis a modificar (conjunto mínimo de campos a imputar). Na versão 2 foi introduzida a estratégia de seleção de variáveis, que toma como base não só o conjunto mínimo como também o peso das variáveis. Dessa forma, na versão 2, o conjunto de variáveis a serem imputadas não necessariamente é o conjunto mínimo.

A seleção das variáveis é feita levando em conta a de maior Índice de Suspeita, cujo cálculo é baseado na atividade da variável, no peso da variável e, em menor medida e em proporção inversa, no grau de atividade do código frente às regras do conjunto completo. A seguir é apresentada a expressão para o

cálculo do Índice de Suspeita de uma dada variável  $i$ , denotado por  $IS(var_i)$ :

$$IS(var_i) = ATIV(var_i) \cdot [PESO(var_i) + [1 - PESO(var_i)] \cdot P_j]$$

onde:

$ATIV(var_i)$  é a atividade da variável  $i$  e representa o número de regras de crítica do conjunto completo em que a variável está ativa; é o fator que mais contribui para o cálculo do  $IS(var_i)$  e varia entre zero e o número de regras de crítica do conjunto completo.

$PESO(var_i)$  é o peso associado à variável  $i$ ;

$[1 - PESO(var_i)] \cdot P_j$  pode ser interpretado como um incremento à suspeita subjetiva e serve para resolver hipotéticos empates;

$GAC_j$  é o grau de atividade de um código  $j$  e representa a quantidade de regras de críticas do conjunto completo em que o código está ativo;

$P_j$  é construído a partir do grau de atividade do código ( $GAC$ ). Quanto maior o grau de atividade de um código menor será o valor associado ao  $P_j$ . A divisão por 10 foi apenas para diminuir o impacto deste fator.

$$P_j = \frac{1 - \frac{GAC_j}{MAX_{GAC} + 1}}{10}$$

$MAX_{GAC}$  representa o  $MAX \{GAC_j \mid j \in \text{conjunto de códigos válidos de todas as variáveis}\}$ .

No caso de ocorrência de mais de uma variável com o valor máximo do Índice de Suspeita é feita uma seleção entre elas para decidir qual será imputada.

#### **Seleção do código a imputar**

Após a seleção das variáveis a imputar de um registro com erro, resta determinar o código a ser imputado para cada uma delas.

Se só existe um código possível que seja diferente do código atribuído ao registro rejeitado, este é eleito e diz-se que ocorreu **Imputação Probabilística Única (IPU)** da variável.

Se existem vários códigos possíveis, é feita a seleção de um deles e diz-se que ocorreu **Imputação Probabilística Flexível (IPF)** da variável. A seleção do código é feita de acordo com a estratégia de tratamento definida no subsistema de especificação, levando em conta a distribuição de frequência marginal ou conjunta, o tipo de frequência FNS ou FNS' e o método de seleção do código (proporcional ou pela máxima diferença na distribuição de frequências).

#### **6.3.3 Informações para análise do processo de detecção e correção dos erros**

Um dos objetivos do sistema DIA é fornecer ampla informação para subsidiar a análise do processo de detecção e correção dos erros. Para cada aplicação do sistema são geradas tabelas (*TABLAS*) que mapeiam o resultado da detecção e da correção automática dos erros.

Os seguintes grupos de informações são impressos nas *TABLAS*:

#### **Informações gerais da aplicação**

##### **a) Sobre os registros**

- número de registros processados;
- número de registros sem erro ("bons");
- número de registros com algum erro ("maus").

b) Sobre a detecção dos erros

- classificação dos registros errôneos segundo o tipo de erro;
- classificação dos registros errôneos segundo o número de variáveis inválidas e suspeitas.

c) Sobre as regras de crítica

- classificação dos registros errôneos segundo o número de RIDs que falharam;
- classificação dos registros errôneos segundo a RID;
- classificação dos registros errôneos segundo o número de críticas (explícitas) que falharam;
- número de registros que falharam em cada regra de crítica.

d) Sobre as imputações

- número de registros segundo o número de variáveis imputadas;
- número de registros por tipo de imputação;
- número total de imputações por tipo de imputação e tipo de código original;
- número de registros segundo o número de variáveis imputadas probabilisticamente.

**Informações de cada variável**

a) Sobre a detecção dos erros

- número de registros em que a variável é inválida;
- indicadores de suspeita da variável
  - número de registros em que a variável é detectada como suspeita;
  - grau de atividade da variável (nº regras de crítica do conjunto completo em que a variável está ativa).

b) Sobre as imputações

- número de registros em que a variável falhou alguma RID;
- número de registros em que a variável foi imputada por RID;
- critério de seleção do código;
- número de registros imputados probabilisticamente;
- número de registros imputados probabilisticamente com seleção entre vários códigos;
- número de registros com imputação da variável.

**Informações dos códigos de cada variável**

- indicadores de suspeita dos códigos;
- número de registros em que o código substituiu um código inválido;
- número de registros em que o código foi imputado por RIDs;
- número de registros em que o código foi imputado probabilisticamente;
- distribuição marginal de cada variável, em que considera:
  - o arquivo de entrada (antes da correção);
  - o arquivo dos registros "bons";
  - o arquivo com todos os registros cuja variável é não suspeita (o arquivo de entrada exceto os registros com valor inválido ou que a variável seja ativa em regras que falharam), que gera a FNS;
  - o arquivo com todos os registros cuja variável é não suspeita após a atuação das RIDs, que gera a FNS';
  - o arquivo corrigido.

Além das informações impressas nas *TABLAS*, é criado um arquivo para proceder outras análises, que contém para cada registro com erro as informações sobre:

- críticas que falharam;
- RIDs que falharam;
- RIDs que foram imputadas;

- variáveis com valor inválido; e
- variáveis suspeitas (ativas em regras de crítica que falharam).

O sistema DIA garante que a correção é realizada com uma única aplicação, ou seja, uma vez tratados os registros com erro nenhum registro será rejeitado em uma nova rodada do módulo de detecção considerando o mesmo conjunto de regras de crítica.

Além da vantagem de evitar os ciclos de processamento, um outro aspecto altamente positivo é o fato de que não há necessidade de desenvolvimento de programas de imputação. Com isso já é possível perceber o ganho significativo desse método em relação ao adotado no Censo de 80.

#### **6.4 A Aplicação do Sistema DIA aos Dados do CD 1.01**

Neste item são abordadas a especificação, a estratégia de tratamento, a formação e seleção dos lotes de apuração e a execução de cada aplicação do sistema DIA para a detecção e correção automática dos erros dos dados do CD 1.01.

##### **6.4.1 A definição das regras de crítica**

A DME foi responsável pela coordenação de um grupo de trabalho integrado por técnicos da DME, da CTD e da DI (DEATE/DIDEM). Esse grupo foi responsável pela especificação das regras de crítica aplicadas para a depuração dos dados do CD 1.01, e por uma revisão profunda das regras de crítica que haviam sido especificadas para aplicação durante as etapas anteriores à apuração centralizada, a saber: a crítica durante a transcrição dos dados e a pré-crítica.

Cabe registrar a necessidade de uma perfeita harmonia entre as fases de apuração dos dados do Censo, ou seja, desde a tarefa de transcrição de dados para o meio magnético até a correção automática dos dados.

Dentre as fases de apuração do CD 1.01, aquelas que se encontram mais intimamente ligadas são a pré-crítica e a detecção

e correção automática dos erros. Quando da definição das variáveis que deviam ser tratadas na pré-crítica, teve-se o cuidado para que as variáveis corrigidas na pré-crítica não sofressem nenhuma alteração durante a correção automática pois, caso contrário, além de se perder o controle sobre o processo de apuração, existiria o risco de transformar casos já tratados em incorretos.

A seguir são apresentadas as variáveis do CD 1.01 que foram submetidas ao processo de detecção e correção automática dos erros:

■ **Características do domicílio:**

- V202 (localização)
- V203 (abastecimento de água)
- V204 (escoadouro)
- V205 (uso de instalação sanitária)
- V206 (condição de ocupação)
- V207 (cômodos)
- V208 (cômodos servindo de dormitório)
- V209 (banheiros)
- V210 (destino do lixo)

■ **Características das pessoas:**

- V3043 (idade calculada com base no mês e ano de nascimento/idade presumida)
- V305 (sabe ler e escrever)
- V306 (última série que concluiu com aprovação)
- V307 (grau da última série concluída com aprovação)
- V3080 (tipo de rendimento).

Uma vez que as críticas entre registros já tinham sido tratadas na fase de pré-crítica, para a especificação das críticas a serem tratadas na fase de detecção e correção automática foi planejado executar o tratamento dos dados em três aplicações do sistema DIA para cada lote<sup>29</sup> de apuração: uma para

---

<sup>29</sup> A formação de lotes está descrita no item 6.4.3.

as características de domicílios e duas referentes às características de pessoas. Quanto às duas últimas aplicações, uma foi desenvolvida para tratar as características dos chefes dos domicílios e individuais em domicílios coletivos, e a outra para as características dos não-chefes de domicílio, considerando a idade já corrigida do chefe para tratar a idade dos demais componentes do domicílio.

Independentemente do método a ser empregado para a correção automática dos dados de um Censo, existe sempre a preocupação em controlar a imputação da variável **idade**, historicamente considerada bastante confiável, tentando assim garantir, na medida do possível, uma certa coerência entre as idades dos membros do domicílio. A principal razão para adotar duas aplicações do sistema DIA para a detecção e correção dos erros em características de pessoas foi possibilitar a construção da variável auxiliar "faixa de idade do chefe do domicílio" (V3044), obtida a partir dos dados "limpos" da variável idade dos chefes. Daí, então, foi possível tratar a variável idade dos não-chefes usando a distribuição conjunta com a variável auxiliar "faixa de idade do chefe" e com a variável "parentesco ou relação com o chefe do domicílio (V302)".

Os planos de crítica foram especificados independentemente para cada aplicação e submetidos ao analisador de regras e à geração do conjunto completo das regras de crítica. O conjunto de regras de imputação determinística (RIDs), o conjunto de regras de crítica de incompatibilidade (especificadas pelos analistas temáticos e com as redundâncias já eliminadas) e o conjunto completo de regras de crítica adotados na versão definitiva da execução da fase de detecção e correção automática dos erros, para cada uma das aplicações do sistema DIA, encontram-se no Anexo 5.

#### **6.4.2 A estratégia para o tratamento dos dados**

No decorrer dos trabalhos de elaboração dos planos de crítica para a execução da detecção e correção automática dos

erros, foram definidas as especificações para subsidiar a execução do tratamento dos dados.

A utilização desses procedimentos envolveu não só a especificação da criação de algumas variáveis auxiliares, mas também a definição da estratégia de imputação dos dados, através de um módulo pertencente ao subsistema de especificações do sistema DIA, que permite ao usuário exercer um controle efetivo sobre a imputação.

Antes da depuração definitiva das informações do CD 1.01 foram feitos testes com alguns lotes de apuração, os quais originaram alguns ajustes na estratégia de tratamento dos dados. Além disso, após o início efetivo do sistema de produção houve duas interrupções<sup>30</sup>, que levaram a novas alterações na estratégia de tratamento dos dados.

Neste item são apresentadas a definição das variáveis auxiliares criadas<sup>31</sup> e a estratégia definida para o tratamento dos dados, para cada uma das três aplicações do sistema DIA, na versão definitiva da execução da detecção e correção automática dos erros do CD 1.01.

#### ■ Aplicação referente às características do domicílio

As seguintes variáveis referentes às características de domicílio foram submetidas à aplicação do sistema DIA para a detecção e correção automática dos erros:

- V202 (localização)
- V203 (abastecimento de água)
- V204 (escoadouro)
- V205 (uso de instalação sanitária)
- V206 (condição de ocupação)
- V207 (cômodos)
- V208 (cômodos servindo de dormitório)
- V209 (banheiros)
- V210 (destino do lixo).

---

<sup>30</sup> Ver item 6.4.4.

<sup>31</sup> Ver IBGE(1992).

**- Variáveis auxiliares criadas**

**V7100 (total de pessoas)**

A variável V7100 foi obtida através da soma das variáveis V111 (total de homens) e V112 (total de mulheres) e criada após a fase da pré-crítica, durante a formação dos lotes de apuração para a detecção e correção automática dos erros.

Códigos possíveis da V7100: 01 a 99.

**V2203 e V2204 (variáveis indicadoras da existência de rede geral de água e esgoto, a nível de distrito/subdistrito)**

Diante da possibilidade de ocorrência de registros indicando a existência de serviço de rede geral de água ou esgoto em domicílios pertencentes a distritos ou subdistritos que não possuíam esses serviços, foi decidido utilizar os resultados de uma pesquisa realizada em 1989, pelo Departamento de Indicadores Sociais (DEISO), que investigou as condições de saneamento básico nos municípios, para apurar as possíveis incompatibilidades envolvendo as variáveis V203 (abastecimento de água) e V204 (escoadouro).

Foram, então, criadas as variáveis V2203 e V2204 durante a formação dos lotes de apuração para detecção e correção automática dos erros, a partir de informações do arquivo do DEISO.

Códigos possíveis da V2203 e V2204: 1 (tem) e 2 (não tem).

**V1062 (tipo de setor)**

A variável V1062 foi criada durante a formação dos lotes de apuração, a partir de informações do arquivo da Sinopse e foi utilizada para a imputação determinística da variável V202 (localização).

Códigos possíveis da V1062: 0 (não especial); 1 (especial - favela, alagado); 2 (especial - quartel); 3 (especial - alojamento); 4 (especial - embarcação); 5 (especial - aldeia); 6 (especial - presídio, cadeia, etc.) e 7 (especial - asilo, orfanato, etc).

## **- Estratégia de imputação**

### **a) Variáveis fixas**

- V7100 (total de pessoas)
- V201 (espécie do domicílio)
- V2203 (indicadora de rede geral de água)
- V2204 (indicadora de rede de esgoto)
- V1062 (tipo de setor).

### **b) Peso das variáveis**

Para todas variáveis envolvidas nessa aplicação foi utilizado o peso médio 5.

### **c) Critério de seleção do código a ser imputado para cada variável**

Todas as variáveis submetidas a essa aplicação tiveram seus códigos imputados probabilisticamente segundo as respectivas distribuições marginais e pelo método proporcional, com exceção da variável V202 (localização) que teve também imputação determinística. Neste caso foi adotada a imputação determinística única (IDU).

A frequência de registros não suspeitos adotada para cada variável foi a padrão (FNS), com exceção da variável V202 (localização), cuja frequência dos registros não suspeitos foi obtida após a atuação das imputações determinísticas (FNS\*).

### **■ Aplicação referente às características dos chefes do domicílio e dos individuais**

As seguintes variáveis referentes às características dos chefes do domicílio e dos individuais foram submetidas à aplicação do sistema DIA para a detecção e correção automática dos erros:

- V3043 (idade calculada)
- V305 (sabe ler e escrever)
- V306 (última série que concluiu com aprovação)

## **- Estratégia de imputação**

### **a) Variáveis fixas**

- V7100 (total de pessoas)
- V201 (espécie do domicílio)
- V2203 (indicadora de rede geral de água)
- V2204 (indicadora de rede de esgoto)
- V1062 (tipo de setor).

### **b) Peso das variáveis**

Para todas variáveis envolvidas nessa aplicação foi utilizado o peso médio 5.

### **c) Critério de seleção do código a ser imputado para cada variável**

Todas as variáveis submetidas a essa aplicação tiveram seus códigos imputados probabilisticamente segundo as respectivas distribuições marginais e pelo método proporcional, com exceção da variável V202 (localização) que teve também imputação determinística. Neste caso foi adotada a imputação determinística única (IDU).

A frequência de registros não suspeitos adotada para cada variável foi a padrão (FNS), com exceção da variável V202 (localização), cuja frequência dos registros não suspeitos foi obtida após a atuação das imputações determinísticas (FNS\*).

### **■ Aplicação referente às características dos chefes do domicílio e dos individuais**

As seguintes variáveis referentes às características dos chefes do domicílio e dos individuais foram submetidas à aplicação do sistema DIA para a detecção e correção automática dos erros:

- V3043 (idade calculada)
- V305 (sabe ler e escrever)
- V306 (última série que concluiu com aprovação)

V307 (grau da última série concluída com aprovação)  
V3080 (tipo de rendimento).

**- Variáveis auxiliares criadas**

**V3043 (idade calculada)**

A variável V3043 foi criada, com base na informação do mês e ano de nascimento ou na idade presumida, após a fase da pré-crítica, durante a formação dos lotes de apuração para a detecção e correção automática dos erros.

Códigos possíveis da V3043: 1 a 130 anos completos e de zero a 11 meses completos, codificados para 900 a 911, respectivamente.

**V3080 (tipo de rendimento)**

A variável V3080 representa a variável V3083 (rendimento mensal bruto do chefe do domicílio) codificada<sup>32</sup> e foi criada após a fase da Pré-Crítica, durante a formação dos lotes de apuração para a detecção e correção automática dos erros.

Códigos possíveis da V3080: b (branco); 0 (sem rendimento); 1 (de 0000001 a 9999997) e 9 (9999998 - ignorado).

Após o tratamento da variável V3080, pelo sistema DIA, foi feita uma decodificação da variável V3083 (rendimento mensal bruto do chefe do domicílio) em função dos valores resultantes da V3080, cuja correspondência encontra-se item 6.6.

**- Estratégia de imputação**

**a) Variável fixa**

V302 (parentesco ou relação com o chefe do domicílio).

**b) Peso das variáveis**

A V3043 (idade calculada) foi imputada utilizando peso 1, por tratar-se de uma variável de maior grau de confiança; para as demais variáveis foi utilizado o peso médio 5.

---

<sup>32</sup> A V3083 é quantitativa e o sistema DIA só trata variáveis qualitativas, com no máximo 4 dígitos.

**c) Critério de seleção do código a ser imputado para cada variável**

As variáveis V305 (sabe ler e escrever), V306 (série concluída) e V307 (grau) tiveram seus códigos imputados de forma probabilística ou determinística. A imputação determinística da V305 foi, em alguns casos, flexível (IDF) e em outros única (IDU), enquanto que para as variáveis V306 e V307 foi a imputação determinística única (IDU).

Para os códigos imputados probabilisticamente da variável V305 (sabe ler e escrever) foi adotado o método proporcional e a distribuição conjunta com a variável V3043 (idade calculada) composta pela freqüência de registros não suspeitos padrão (FNS).

As imputações probabilísticas das variáveis V306 (série concluída) e V307 (grau) foram realizadas segundo o método proporcional e as respectivas distribuições marginais de cada variável, compostas pela freqüência dos registros não suspeitos obtidas após a atuação das imputações determinísticas (FNS\*).

As variáveis V3043 (idade calculada) e V3080 (tipo de rendimento) tiveram seus códigos imputados, probabilisticamente segundo a distribuição marginal da respectiva variável, obtida a partir da freqüência dos registros não suspeitos padrão (FNS) e pelo método proporcional.

**■ Aplicação referente às características dos não-chefes**

As seguintes variáveis referentes às características dos não-chefes do domicílio foram submetidas à aplicação do sistema DIA para a detecção e correção automática dos erros:

V3043 (idade calculada)

V305 (sabe ler e escrever).

**- Variáveis auxiliares criadas**

**V3043 (idade calculada)**

A variável V3043 foi criada, com base na informação do mês e ano de nascimento ou na idade presumida, após a fase da

pré-crítica, durante a formação dos lotes de apuração para a detecção e correção automática dos erros.

Códigos possíveis da V3043: 1 a 130 anos completos e de zero a 11 meses completos, codificados para 900 a 911, respectivamente.

#### **V3044 (faixa de idade do chefe do domicílio)**

A variável V3044 foi criada para os registros de não-chefes com base nas informações dos registros dos respectivos chefes após a execução da aplicação do DIA referente a características de chefes e individuais.

Códigos possíveis da V3044: 1 a 11 (formados por faixas quinquenais a partir de 10 anos até 49 anos; decenais a partir de 50 anos e 70 anos ou mais, respectivamente) e branco para o individual.

#### **- Estratégia de imputação**

##### **a) Variáveis fixas**

V302 (parentesco ou relação com o chefe do domicílio)

V3044 (faixa de idade do chefe do domicílio).

##### **b) Peso das variáveis**

A V3043 (idade calculada) foi imputada utilizando peso 1; para as demais variáveis foi utilizado o peso médio 5.

##### **c) Critério de seleção do código a ser imputado para cada variável**

A variável V3043 (idade calculada) teve seus códigos imputados probabilisticamente segundo a distribuição conjunta com as variáveis V3044 (faixa de idade do chefe) e a V302 (parentesco ou relação com o chefe do domicílio), obtida a partir da frequência dos registros não suspeitos padrão (FNS) e pelo método proporcional.

A variável V305 (sabe ler e escrever) teve seus códigos imputados de forma probabilística ou determinística (IDF ou IDU). Para os códigos imputados probabilisticamente foi adotado o

método proporcional e a distribuição conjunta com a variável V3043 (idade calculada) composta pela frequência de registros não suspeitos padrão (FNS).

#### **6.4.3 A formação e seleção dos lotes de apuração**

Assim que os dados do CD 1.01 foram transmitidos para o Órgão Central, o primeiro procedimento adotado na etapa centralizada foi a consolidação dos dados transmitidos, para em seguida formar os lotes de apuração, com vistas à execução da detecção e correção automática dos erros através do sistema DIA.

Diferentemente da pré-crítica, que considerou como sendo um lote de trabalho cada pasta de questionários, a formação dos lotes com vistas à correção automática foi baseada em critérios um pouco mais exigentes, tendo em vista o uso de uma metodologia que corrige automaticamente as inconsistências.

Quando do início da discussão sobre a necessidade de definir quantos e como seriam os lotes a serem processados automaticamente, o grupo, formado por técnicos da DME e CTD, encarregado deste trabalho, tomou conhecimento da existência de uma regra inicialmente definida para formar os lotes de apuração. De acordo com esta regra os questionários de um mesmo distrito formariam um lote, com exceção dos municípios considerados grandes como Rio de Janeiro, Fortaleza, São Paulo, Recife, Salvador, Belo Horizonte e Distrito Federal, onde para a formação de cada lote seria considerado o subdistrito.

Analisando-se essa regra com maior profundidade, descobriu-se que o total estimado de lotes ficaria em torno de 8.900 e a quantidade de questionários por lote seria bastante variável, podendo obter lotes com 20 questionários e outros com aproximadamente 300.000 questionários.

Não fosse o grande número de lotes a serem processados e devidamente analisados bem como a grande variabilidade no tamanho dos lotes, essa proposta para formação dos lotes de apuração até seria factível, visto que contemplava a divisão regional dentro da Unidade da Federação, procurando preservar as características regionais durante o processo de imputação.

Em vista disso é que o grupo encarregado para a especificação das regras de crítica apresentou uma nova proposta para a formação de lotes de apuração, baseada nos seguintes pontos<sup>33</sup>:

- obtenção do menor número possível de lotes, para minimizar o número de relatórios a serem analisados, após cada aplicação do sistema DIA;

- obtenção de um tamanho mínimo a fim de viabilizar o processo de correção, tendo em vista a utilização das distribuições de registros "bons" (sem erros) como base da imputação;

- geração dos lotes levando em conta a situação do domicílio (urbana e rural) bem como a divisão geográfica do país, contemplando a divisão de cada Unidade da Federação, através da ordenação dos respectivos setores segundo a microrregião, município, distrito, subdistrito, de modo a garantir, na medida do possível, maior homogeneidade quanto às características de cada região durante o processo de correção.

Baseando-se nesses critérios, considerados básicos, foram gerados 528 lotes de apuração, sendo 403 urbanos e 125 rurais, de aproximadamente 70.000 domicílios cada. A tabela 2 apresenta o número de lotes, de setores, de questionários e de pessoas resultantes das informações consolidadas do CD 1.01, segundo as Unidades da Federação.

As variáveis auxiliares especificadas no item anterior (6.4.2), necessárias à execução de cada aplicação do sistema DIA, foram construídas e incluídas nos arquivos de dados durante a preparação dos lotes de apuração. Além disso foram incluídas as seguintes variáveis:

#### **V1061 (situação do setor)**

Criada a partir das informações do arquivo da Sinopse e utilizada para a formação dos lotes.

Códigos possíveis: 1 (área urbanizada); 2 (área não urbanizada); 3 (urbanizada isolada); 4 (rural com extensão

---

<sup>33</sup> Ver Silva, Oliveira e Oliveira(1992).

urbana); 5 (rural povoado); 6 (rural núcleo); 7 (rural outros); e 8 (rural exclusive aglomerados).

Os códigos de 1 a 3 correspondem à área urbana e os demais à área rural.

#### **V100 (número do lote)**

Códigos possíveis: 1 a 528.

Até a criação da variável V100 (número do lote) os dados investigados através do CD 1.01 de cada Unidade da Federação estavam guardados em um ou mais arquivos e, portanto, continham dados de vários lotes. O armazenamento dos dados de cada lote em um único arquivo recebeu a denominação de **seleção dos lotes**. Assim, foram extraídos dos arquivos das Unidades da Federação todos os dados de um mesmo número de lote e gravados no arquivo de lote único.

Após a seleção dos lotes foram preparados os arquivos de dados para serem submetidos ao sistema de produção para a detecção e correção automática dos erros. Para tanto, os dados de cada lote foram divididos em 4 arquivos: arquivo de domicílios, arquivo de pessoas chefes ou individuais, arquivo de pessoas não-chefes e arquivo de questionários faltosos.

**Tabela 2 - Número de lotes, de setores, de questionários e de pessoas do CD 1.01 segundo as Unidades da Federação**

Unidades da Federação	Lotes	Setores	Questionários	Pessoas
TOTAL	528	163.269	35.452.151	147.059.416
Rondônia	5	1.523	258.808	1.132.692
Acre	2	392	89.774	417.718
Amazonas	7	2.148	395.228	2.103.243
Roraima	2	297	45.438	217.583
Pará	15	4.396	1.001.783	5.183.466
Amapá	2	250	54.233	289.397
Tocantins	4	1.087	194.970	919.863
Maranhão	15	5.423	990.947	4.930.374
Piauí	9	3.195	522.589	2.582.137
Ceará	21	6.732	1.356.758	6.366.647
Rio Grande do Norte	9	2.746	525.656	2.415.567
Paraíba	11	3.614	699.410	3.201.114
Pernambuco	24	7.366	1.603.580	7.127.855
Alagoas	8	2.575	532.110	2.514.100
Sergipe	6	1.764	332.384	1.492.122
Bahia	37	12.521	2.542.694	11.868.122
Minas Gerais	54	17.483	3.777.297	15.743.154
Espírito Santo	10	2.709	626.346	2.600.618
Rio de Janeiro	51	14.480	3.521.691	12.807.735
São Paulo	116	33.583	8.268.124	31.588.925
Paraná	31	10.362	2.123.294	8.448.719
Santa Catarina	17	5.210	1.141.353	4.541.994
Rio Grande do Sul	36	13.035	2.537.288	9.138.670
Mato Grosso do Sul	7	2.047	441.853	1.780.373
Mato Grosso	7	2.267	474.167	2.027.231
Goiás	15	4.321	1.007.980	4.018.903
Distrito Federal	7	1.743	386.396	1.601.094

Fonte: Listagem de 21.02.94 contendo as informações dos dados consolidados por lote.

#### 6.4.4 A execução da detecção e correção automática dos erros pelo sistema DIA

O sistema de produção foi desenvolvido de modo a submeter cada lote de questionários, previamente formado, às aplicações de Domicílio, de Chefe e de Não-Chefe, nessa ordem, já que a aplicação de Não-Chefe dependia da realização prévia da correção dos dados dos chefes. Na figura 3 é apresentado um fluxo da operação de detecção e correção automática dos erros dos dados do CD 1.01, pelo sistema DIA.

Antes da depuração definitiva das informações de Domicílio, Chefe e Não-Chefe, foram feitos testes com lotes de trabalho de algumas Unidades da Federação, já liberados da fase de pré-crítica, com o intuito de verificar todo o funcionamento da operação, detectar possíveis problemas que poderiam aparecer posteriormente e, com base nos relatórios de saída da correção, elaborar o Plano de Análise da Correção Automática.

Foi em função desses testes que ficou definida a estratégia de aplicar um peso diferenciado para a variável idade, de modo a controlar a sua imputação.

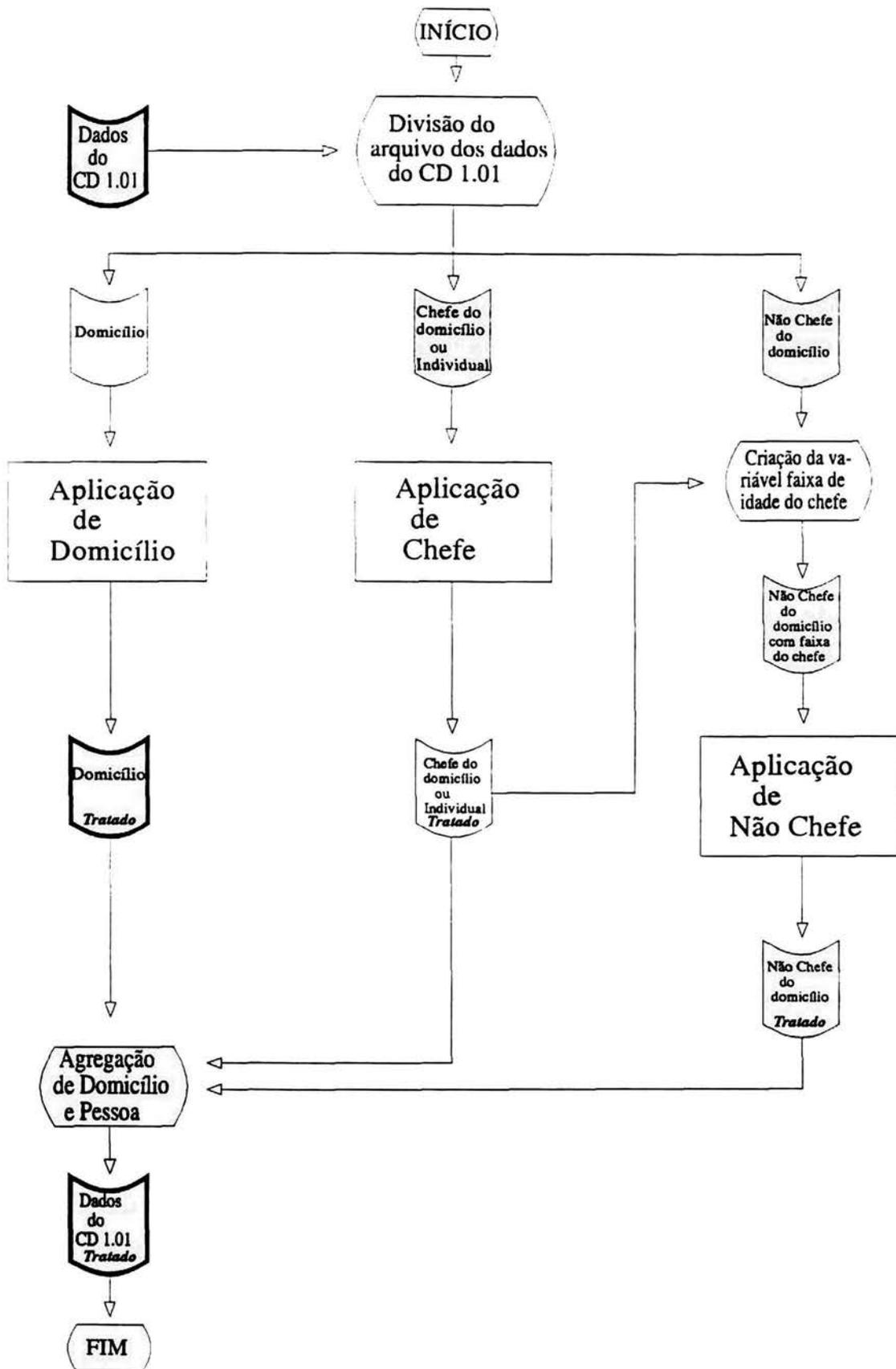
Em dezembro de 1992 deu-se o início do processo de produção da fase de detecção e correção automática dos erros para os lotes de algumas Unidades de Federação que já haviam transmitido os dados após a execução da fase de pré-crítica.

Apesar de ter existido um período de realização de testes com massa real de dados, nem todos os problemas foram detectados durante esse período. Após o início efetivo, o sistema de produção foi interrompido ainda duas vezes, devido à ocorrência de alguns problemas, descritos a seguir.

Em julho de 1993, após o processamento de todos os lotes, durante a análise do processo de correção automática, as análises efetuadas pelo DEPOP (Departamento de População), confirmadas pelas tabelas geradas pela aplicação do sistema DIA (*TABLAS*), indicavam ter havido erro sistemático de coleta para algumas variáveis, cujo tratamento não estava previsto no plano de crítica especificado.

FIGURA 3

Fluxo da Detecção e Correção Automática dos dados do CD 1.01 pelo sistema DIA



Um deles diz respeito ao registro errado da V305 (sabe ler e escrever) para as crianças de 4 e 5 anos que, provavelmente, devido ao erro de cálculo da idade por parte dos recenseadores foi registrado um código na V305 incompatível com a idade real. Para as crianças nascidas nos meses de setembro a dezembro de 1986 (com 4 anos na data de referência do Censo - 01.09.91) foi registrado erroneamente na V305 os códigos 1 (sabe ler e escrever) ou 2 (não sabe); ocorreu também para as crianças com 5 anos um erro na idade fazendo com que fosse registrado código 3 (menos de 5 anos) na V305, ao invés do código 1 ou 2. Apesar da variável "idade calculada" ter recebido um peso que representa um grau de desconfiança bastante reduzido, em alguns registros a correção desses erros foi realizada através da alteração da idade, para mais ou para menos, conforme o tipo de erro, acarretando uma redução do número de crianças, principalmente na idade de 4 anos.

Como consequência, levando em conta o fato da variável "idade calculada" (V3043) ser considerada, historicamente, bastante confiável e que o problema ocorrido estava ligado ao corte do questionário (5 anos ou mais e menos de 5 anos), decidiu-se corrigir o erro sistemático de coleta, através da introdução de regras de imputação determinística (RIDs) no conjunto de regras de crítica para a aplicação de chefe e de não-chefe, o que implicou em nova geração do conjunto completo de regras de crítica pelo sistema DIA.

Dessa forma, foram introduzidas RIDs que abrangiam todas as crianças de 0 a 5 anos, a fim de que esses tipos de erros fossem corrigidos através da imputação na V305, conservando assim o valor atribuído à variável "idade calculada".

O outro erro sistemático que foi detectado diz respeito à declaração dos cursos "primário", "ginasial" e "colegial", incompatível com a idade do informante.

Apesar da reforma do ensino ter sido instituída em 1972, quando esses cursos foram substituídos pelos cursos de 1º grau com 8 séries e de 2º grau, algumas pessoas que freqüentam ou freqüentaram o 1º ou 2º grau tiveram a declaração de série e grau equivalente aos antigos "primário", "ginasial" e

"colegial". Esse fato foi detectado através das críticas que relacionam a frequência a esses tipos de cursos somente para pessoas com idades mais avançadas.

Logo, as pessoas registradas como tendo feito o "primário", "ginasial" ou "colegial" deveriam ter idade compatível com a declaração, ou seja, uma idade mais elevada. Durante a análise percebeu-se que, algumas vezes a correção desses erros estava se dando através da alteração da variável "grau", enquanto a série permanecia constante. Esse fato estava acarretando uma alteração no número de anos de estudo já que, às vezes, para manter a equivalência entre a quantidade de anos, seria necessário alterar a série também, como no exemplo citado a seguir:

- considere uma pessoa com 18 anos de idade que tenha declarado ter freqüentado a 1ª série do ginásial; a informação coerente da idade com esse curso seria a partir de 29 anos. Nesse caso, alterando-se o curso de ginásial para 1º grau, seria necessário alterar, também, a série de 1ª para a 5ª, estabelecendo a equivalência em anos de estudo.

Este problema foi solucionado através da inclusão de algumas RIDs, com a finalidade de estabelecer a equivalência entre os cursos caracterizados erroneamente e os correspondentes atuais.

Em função da introdução das RIDs para o tratamento de erros sistemáticos de coleta, foi alterada também a estratégia de imputação, passando a utilizar a distribuição de registros não suspeitos, obtida após a atuação das RIDs, como base para a imputação probabilística, quando fosse o caso, das variáveis passíveis de imputação determinística única.

A produção começou a reprocessar os lotes novamente, em agosto de 1993, e passou a ser executada no que foi denominada de versão "2" para as três aplicações: de Domicílio, de Chefe e de Não-Chefe. A atualização da versão abrangeu também a aplicação de Domicílio, devido à concepção do sistema de produção montado pela DI.

Após o reprocessamento de 246 lotes, na versão "2", em 04.10.93 houve uma outra interrupção do sistema de produção. A

continuidade da análise dos resultados com as novas especificações indicou a necessidade de uma nova modificação na estratégia de tratamento dos dados a fim de possibilitar a imputação da variável V305 (sabe ler e escrever) condicionada à idade, dado que a distribuição da alfabetização é sabidamente diferenciada ao longo das idades, sendo que a grande proporção de crianças de 5 anos de idade não sabe ler e escrever.

Ao analisar a distribuição dos dados de entrada e de saída da correção observou-se uma discrepância na alfabetização das crianças com 5 anos de idade. O fato ocorreu devido à imputação da V305 estar, até aquele momento, utilizando apenas a distribuição marginal dos registros "bons" da variável V305 dentro de cada lote de operação, quando melhor seria se estivesse condicionada às idades individuais, retratando assim a caracterização predominante em cada uma delas. Este acerto originou a versão "3" da correção automática, mediante a inclusão de uma estratégia de imputação, onde a correção da V305 ficou condicionada às idades individuais.

Cabe ressaltar que apesar do tempo decorrido entre o início efetivo da correção automática e o seu término ter sido bastante extenso, o intervalo gasto para o processamento dos 528 lotes na versão definitiva (versão "3") foi de 18.10.93 a 24.11.93.

Em se tratando de uma metodologia utilizada pela primeira vez em uma pesquisa do IBGE, é natural a ocorrência de alguns ajustes ou calibrações. Destaque deve ser dado para a facilidade com que foram feitas as alterações no conjunto de regras de crítica ou na estratégia de tratamento dos dados com o uso do sistema DIA. O Censo Demográfico de 1991 teve a grande oportunidade de utilizar uma ferramenta poderosíssima e bastante eficiente, obtendo assim, um grande avanço em termos de apuração censitária.

#### **6.5 A Imputação dos Dados em Questionários Faltosos**

Conforme descrito em 5.2, durante a pré-crítica, foram criados os questionários faltosos e recuperadas das Folhas de Coleta as informações referentes às variáveis: V110 (nº no

CD 1.03), V111 (total de homens), V112 (total de mulheres) e V201 (espécie do domicílio) para permitir que esses questionários fossem submetidos ao processo de imputação dos dados.

Foi feita uma avaliação de que no caso de um questionário faltoso, praticamente todo em branco, a imputação pelo processo DIA, como é feita variável a variável, ficaria muito próxima a uma fabricação pura e simples de um questionário fictício. Isto porque o processo é ideal para tratar os casos de não resposta parcial e erros entre campos de um questionário. Para evitar a criação de questionários fictícios, decidiu-se por um processo de duplicação de questionários utilizando as informações disponíveis.

Dessa forma, foi adotado o método *hot-deck* seqüencial para a imputação dos campos sem declaração dos questionários gravados nos arquivos dos faltosos, de acordo com o seguinte procedimento: procurava-se no lote um questionário, para servir de "doador", que tivesse os mesmos valores que o questionário faltoso para as variáveis V201, V111 e V112. Caso fosse encontrado, todas as variáveis de domicílios e de pessoas eram imputadas, segundo o questionário doador, com exceção da identificação e da V109 (nº no CD 1.07) que continuou com o valor 9996, para permitir a sua identificação.

Por outro lado, caso não fosse encontrado um questionário doador com valores idênticos nas variáveis V111 e V112, mas encontrasse algum, cuja soma destas variáveis fosse a mesma da do faltoso, o procedimento era idêntico ao caso anterior. Somente na impossibilidade de encontrar um doador é que o questionário faltoso era eliminado. Foram eliminados 4 questionários faltosos por falta de doadores no lote, sendo 3 da Bahia (2 do lote 26 com 7 e 4 pessoas, respectivamente e 1 do lote 32 com 8 pessoas) e 1 questionário do lote 21 do Rio de Janeiro com 29 pessoas.

A tabela 3 apresenta o número de questionários faltosos e de pessoas com imputações nesses questionários por Unidade da Federação. Em termos relativos o total de questionários faltosos imputados é insignificante, pois representa 0,07% do total de questionários CD 1.01.

**Tabela 3 - Número de questionários e de pessoas com imputações em questionários faltosos segundo as Unidades da Federação**

Unidades da Federação	Questionários faltosos imputados	Pessoas com imputações em faltosos
TOTAL	25.441	97.352
Rondônia	43	216
Acre	0	0
Amazonas	211	1.108
Roraima	1	1
Pará	242	1.071
Amapá	0	0
Tocantins	5	10
Maranhão	6.253	30.633
Piauí	33	192
Ceará	147	590
Rio Grande do Norte	61	272
Paraíba	16	72
Pernambuco	248	1.112
Alagoas	108	437
Sergipe	0	0
Bahia	382	1.823
Minas Gerais	414	1.500
Espírito Santo	81	326
Rio de Janeiro	1.340	4.617
São Paulo	15.192	51.099
Paraná	141	469
Santa Catarina	75	266
Rio Grande do Sul	150	507
Mato Grosso do Sul	22	81
Mato Grosso	84	304
Goiás	105	384
Distrito Federal	83	262

Fonte: Listagem de 21.02.94.

Cabe registrar que foi perdido o material de coleta de 42 setores do Maranhão, restando apenas as informações resumidas referentes ao número de unidades visitadas, de domicílios particulares, de pessoas recenseadas, de questionários CD 1.01 e CD 1.02, que correspondem a 5.921 domicílios e 24.972 pessoas. Durante a pré-crítica foram criados questionários faltosos correspondentes aos CD 1.01 desses setores perdidos. Posteriormente, nos campos sem declaração, foram imputados os valores através do mesmo procedimento de imputação dos outros questionários faltosos.

#### **6.6 Procedimentos Adicionais**

Uma vez terminado o processo de imputação em todos os questionários faltosos, esses dados foram agregados aos arquivos de dados do CD 1.01 tratados pelo sistema DIA.

Ainda durante a fase de detecção e correção automática dos erros foi criada a variável V7200 (tipo de imputação) e feita uma decodificação para a variável V3083 (rendimento mensal bruto do chefe) para cada um dos registros dos arquivos do CD 1.01.

##### **V7200 (tipo de imputação)**

Esta variável foi criada com a finalidade de permitir aos usuários identificar os registros que sofreram imputação, com os seguintes códigos possíveis: branco (se o registro não sofreu imputação durante a correção automática); 1 (se o registro sofreu imputação pelo sistema DIA); e 2 (se o registro sofreu imputação por se referir a um questionário faltoso).

##### **V3083 (Rendimento Mensal Bruto do Chefe)**

A variável V3083 por ser uma variável quantitativa não foi submetida ao processo de detecção e correção automática dos erros pelo sistema DIA. Dessa forma, decidiu-se por manter o rendimento ignorado no caso da não declaração pelo informante. Além disso, para os campos cujo valor do rendimento era "9999998" foi imputado o valor "9999999" (rendimento ignorado) e nesse caso a variável V3080 passou a assumir o valor "8" para permitir a

identificação do rendimento ignorado imputado. Assim, a decodificação da variável V3083 foi feita da seguinte maneira:

Se V3080 = b  $\Rightarrow$  V3083 = branco

Se V3080 = 0  $\Rightarrow$  V3083 = 0000000

Se V3080 = 1 e V3083 = valor  $\Rightarrow$  V3083 = valor

Se V3080 = 1 e V3083 = b  $\Rightarrow$  V3083 = 9999999<sup>34</sup>

Se V3080 = 9  $\Rightarrow$  V3083 = 9999999

## 6.7 A Análise do Processo Automático de Correção

Após a execução da fase de detecção e correção automática dos erros procedeu-se a análise dos resultados para permitir o controle do processo, bem como para avaliar os impactos na qualidade dos dados obtidos ao final do processo.

Algumas recomendações que sobrepassam qualquer discussão sobre o emprego de métodos de correção automática numa pesquisa são destacadas em Silva e outros(1990) e apresentadas a seguir:

- qualquer processo de imputação não pode jamais substituir o cuidado na realização da pesquisa, desde seu planejamento, coleta, transcrição e etc;

- qualquer processo de imputação precisa ser monitorado e controlado cuidadosamente;

- os usuários dos dados devem ser alertados sobre todos os aspectos e detalhes relevantes para que possam compreender e avaliar por si mesmos o processo de imputação realizado.

A análise do efeito da imputação é indispensável para que se possa avaliar se os métodos adotados, embora não compensando plenamente a perda das informações, contribuíram para cobrir a lacuna dos dados incompletos, assegurando consistência entre os dados, preservando a distribuição básica dos dados e permitindo assim que os usuários tenham maior flexibilidade em desenvolver diferentes análises com o arquivo final, livre de omissões.

Diante da ampla disponibilidade de informações geradas para subsidiar a análise e monitorar o processo de imputação,

---

<sup>34</sup> Neste caso, V3080 = 8, para armazenamento no Banco RAPID.

descritas em 6.3.3, foi realizada uma primeira etapa da análise do processo de correção automática dos dados investigados no CD 1.01, visando a liberação dessa fase. Posteriormente, os resultados da correção automática foram consolidados em níveis geográficos de Brasil, Regiões e Unidades da Federação, com o objetivo de revelar, em termos globais, como funcionou a aplicação da metodologia de Fellegi e Holt e do sistema DIA, cuja análise é apresentada no item 6.7.2.

#### **6.7.1 A análise para a liberação da correção automática**

A análise da correção automática dos dados, para a liberação dessa fase de correção, foi realizada em três níveis geográficos: lote, município e subdistrito, sendo este último nível aplicado apenas aos municípios das principais capitais que compreendem: Rio de Janeiro, São Paulo, Brasília, Salvador, Recife, Belo Horizonte e Fortaleza.<sup>35</sup>

##### **- Análise a nível do lote**

Conforme descrito em 6.3.3, para cada aplicação do sistema DIA são geradas tabelas (*TABLAS*) que mapeiam o resultado da detecção e correção automática dos erros.

Para evitar um grande volume de material para análise, tendo em vista que a detecção e correção automática dos erros nos dados do CD 1.01 se daria em 3 aplicações (Domicílio, Chefe e Não-Chefe) em cada um dos 528 lotes de apuração, decidiu-se restringir a impressão e a análise das *TABLAS* apenas para a(s) aplicação(ões) suspeita(s) de cada lote.

Para tanto, foram definidos limites de tolerância para o percentual de registros "maus" e para o percentual de registros que falharam em cada regra de crítica em relação aos registros "maus" de cada aplicação.

Uma aplicação de um lote foi considerada suspeita quando apresentasse pelo menos um de seus indicadores fora dos limites de tolerância ou alguma variável fixa com valor inválido.

---

<sup>35</sup> Uma descrição do plano de análise da correção automática do CD 1.01 encontra-se em IBGE(1993).

Além disso, para cada lote, independentemente dos valores de seus indicadores, foram emitidas 4 tabelas especiais, definidas pela CTD, com o seguinte conteúdo:

- Tabela 1: Número de registros "bons" e de registros "maus" segundo o número de regras que falharam.

- Tabela 2: Número de registros que falharam em cada regra segundo o número da regra/RID que falhou.

- Tabela 3: Número de registros "maus" segundo o número de variáveis imputadas.

- Tabela 4: Número de questionários, de domicílios e de pessoas segundo a ocorrência de imputações.

À guisa de ilustração, encontram-se no Anexo 6 as 4 tabelas com os valores referentes ao lote 9 do Paraná.

#### **- Análise a nível do município e de subdistrito dos principais municípios**

Uma vez que o município é considerado o principal nível geográfico de divulgação dos dados do CD 1.01, tornou-se necessário efetuar, também, uma análise da correção automática a nível de município, imediatamente após a análise dos lotes. Isto porque, de acordo com o critério de formação dos lotes, os dados de cada município faziam parte de pelo menos 2 lotes (um urbano e outro rural) e os lotes eram compostos, na maioria das vezes, por dados de mais de um município.

Para a análise a nível de município ou subdistrito dos principais municípios foram calculados limites de tolerância para os seguintes indicadores:

- percentual de registros com algum erro;

- percentual de registros com valores ignorados nas variáveis V305, V3043 e V3080;

- maior distância, em termos relativos, entre as freqüências marginais dos dados "bons" e dos dados depurados para cada variável;

- medida resumo das distâncias entre as freqüências marginais dos dados "bons" e dos dados depurados para cada variável;

- maior distância, em termos relativos, entre as frequências marginais dos dados de entrada e dos dados depurados para cada variável;

- medida resumo das distâncias entre as frequências marginais dos dados de entrada e dos dados depurados para cada variável.

Para todos os municípios e subdistritos sujeitos à análise foram emitidos relatórios contendo os valores desses indicadores.

Tais indicadores foram utilizados para identificar os municípios e subdistritos suspeitos. O município ou subdistrito que apresentasse pelo menos um desses indicadores fora do limite de tolerância era considerado suspeito. Para os municípios ou subdistritos suspeitos foram emitidas as tabelas especiais 2, 3 e 4, cujos modelos encontram-se no Anexo 6.

Durante os trabalhos de análise da correção automática, observou-se que os resultados da pesquisa realizada em 1989 pelo DEISO, que investigou as condições de saneamento básico nos municípios e que geraram as variáveis V2203 e V2204 (indicadoras da existência de rede geral de água ou esgoto a nível de distrito/subdistrito), estavam desatualizados. Isto fez com que os percentuais de erros correspondentes à existência ou não desses serviços, detectados por meio das regras de crítica, assumissem valores bem superiores aos percentuais de tolerância.

Nesses casos a CTD entrava em contacto com as Divisões de Pesquisas Estaduais (DIPEQs) do IBGE solicitando informações mais atualizadas junto às Prefeituras e órgãos afins. De posse dessas informações era feita a correção nos arquivos preparados para a aplicação de domicílios e rodada novamente a correção automática do lote.

#### **6.7.2 A análise consolidada dos resultados da correção automática**

Os resultados da correção automática foram armazenados em arquivos de lotes, tendo sido efetuada uma consolidação das tabelas especiais a nível de Unidades da Federação, Regiões e Brasil. Algumas estatísticas foram selecionadas, a fim de revelar como funcionou a aplicação da metodologia de Fellegi e Holt e do

sistema DIA para correção automática dos erros detectados nos dados do Censo Demográfico de 1991 referentes ao CD 1.01.

Uma grande parte do tempo de apuração foi dedicada à fase de detecção e correção automática dos erros, principalmente no planejamento dos planos de crítica e das especificações de tratamento para detectar e corrigir o maior número de inconsistências. Porém, cabe ressaltar que a nível Brasil 87,32% dos questionários ficaram isentos de qualquer correção automática, sendo que o processo de imputação foi executado em apenas 12,68% dos questionários, como pode ser constatado através do gráfico 1.

Um questionário foi considerado com alguma imputação quando pelo menos um campo (no registro do domicílio ou de uma pessoa do domicílio - chefe ou não chefe) teve código imputado pelo sistema DIA ou pelo procedimento adotado para tratar de questionário faltoso.

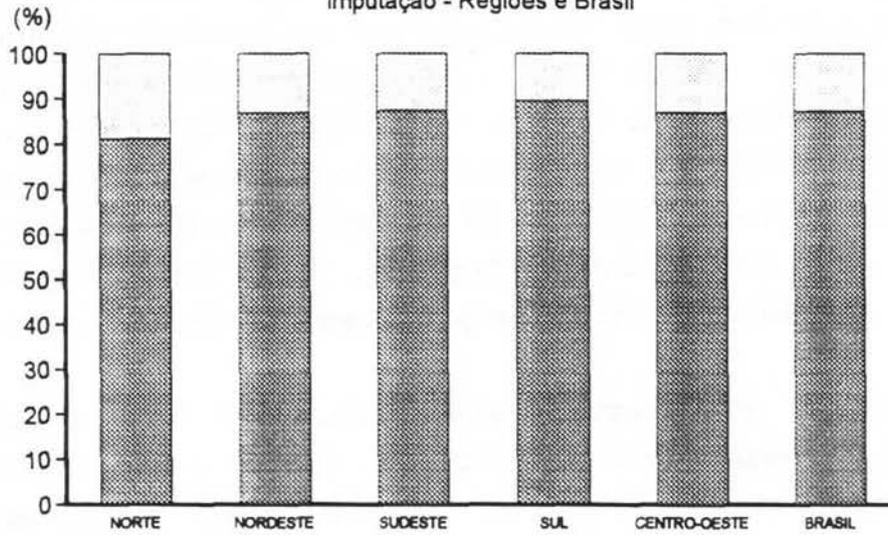
Os questionários que tiveram alguma imputação foram maciçamente tratados pelo sistema DIA, uma vez que apenas 0,07% dos questionários tiveram imputações por se tratarem de questionários faltosos.

No gráfico 1 é apresentada a distribuição de questionários por ocorrência de imputação para cada uma das Regiões e para o Brasil. Essa mesma distribuição para cada Unidade da Federação encontra-se no Anexo 7, gráfico A.7.1. Nota-se uma variabilidade mais acentuada na distribuição de questionários por ocorrência de imputação entre as Unidades da Federação do que entre as Regiões. O declínio do percentual de questionários "sem imputação" na Região Norte é decorrente, em maior proporção, do declínio desse percentual para a Unidade da Federação do Amazonas cujo valor foi de 70,31%.

Uma outra avaliação efetuada diz respeito ao percentual de registros com alguma imputação por tipo de aplicação do sistema DIA aos dados do CD 1.01. O gráfico 2 apresenta esses percentuais por Região e para o Brasil, e os gráficos A.7.2, A.7.3 e A.7.4, no Anexo 7, apresentam os percentuais por Unidade da Federação, para a aplicação de Domicílio, Chefe e Não-Chefe, respectivamente.

Gráfico 1

Distribuição de questionários por ocorrência de imputação - Regiões e Brasil



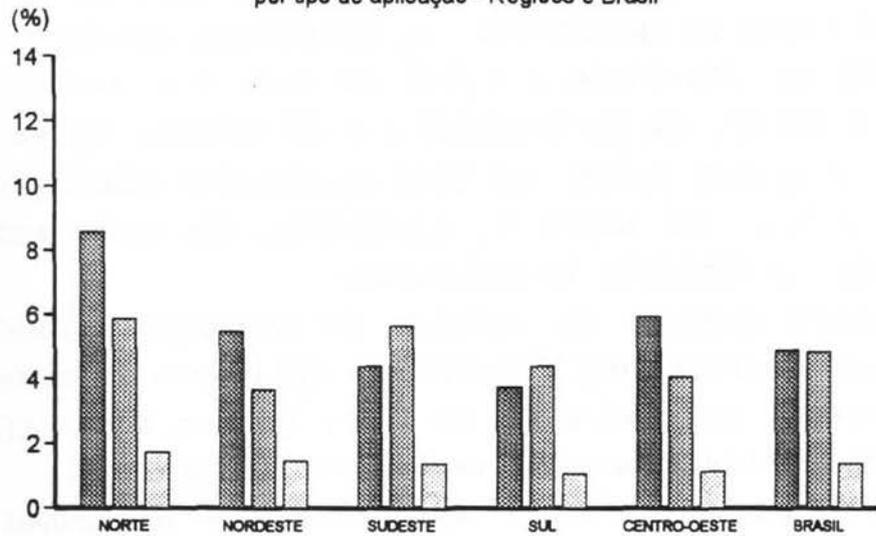
Legenda: Ocorrência

■ Sem imputação □ Com alguma imputação

Fonte: Censo 91 - Resultados da Correção Automática do CD 1.01.

Gráfico 2

Percentual de registros com alguma imputação por tipo de aplicação - Regiões e Brasil



Legenda: Tipo de aplicação

■ Domicílio ■ Chefe □ Não-chefe

Fonte: Censo 91 - Resultados da Correção Automática do CD 1.01, usando o sistema DIA.

Observa-se que a aplicação de Não-Chefe apresenta percentuais de registros com alguma imputação bem inferiores aos percentuais obtidos para as aplicações de Chefe e de Domicílio, o que pode ser justificado pelo menor número de variáveis sujeitas à imputação para essa aplicação. Por outro lado, nas Regiões Sul e Sudeste o percentual de registros com imputação na aplicação de Chefe é superior ao da aplicação de Domicílio, num sentido inverso ao das demais Regiões, o que levou a que para o País esses percentuais ficassem praticamente iguais, em torno de 5%.

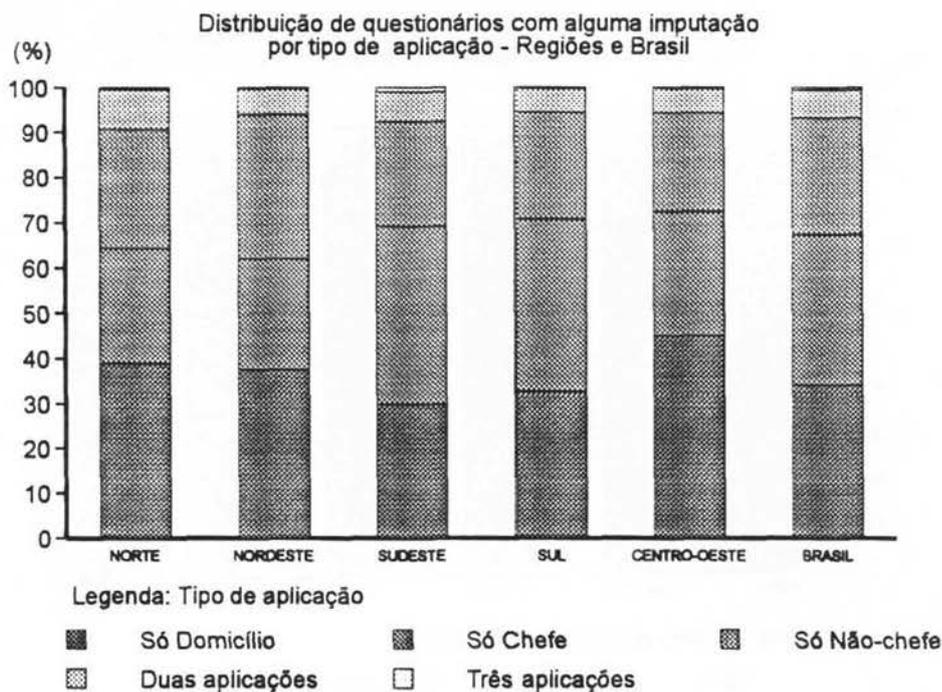
Devido à arquitetura das aplicações que originou a separação dos registros de domicílios, chefes e não-chefes de um mesmo questionário em arquivos distintos para o processamento pelo sistema DIA, é importante analisar os resultados da correção automática a nível de questionário a fim de verificar a sua integridade.

O gráfico 3 apresenta a distribuição de questionários com alguma imputação por tipo de aplicação por Região e para o Brasil, com o intuito de visualizar o quantitativo de questionários submetidos a uma única aplicação ou a mais de uma aplicação. Cabe destacar que, a nível Brasil, dentre os questionários que tiveram alguma imputação através do sistema DIA, 93,12% foram em apenas uma das aplicações (em Domicílio, ou em Chefe ou em Não-Chefe), 6,33% em duas das aplicações (em Domicílio e Chefe, ou em Domicílio e Não-Chefe, ou em Chefe e Não-Chefe) e apenas 0,55% nas três aplicações simultaneamente. O gráfico A.7.5, no Anexo 7, apresenta, de modo análogo, a distribuição por Unidade da Federação.

Um outro aspecto da análise do processo automático de correção visa revelar como funcionou a aplicação do sistema DIA, para a correção dos dados do CD 1.01, no que diz respeito ao número de variáveis imputadas em cada aplicação.

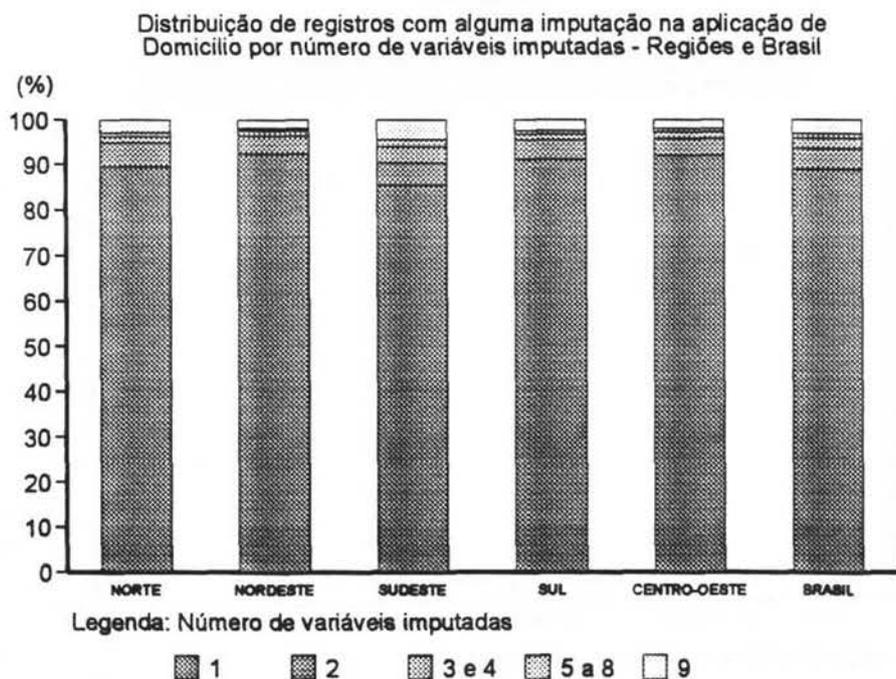
Os gráficos 4, 5 e 6 apresentam as distribuições de registros com alguma imputação na aplicação de Domicílio, Chefe e Não-Chefe, respectivamente, por número de variáveis imputadas para as Regiões e Brasil. Cabe destacar que aproximadamente 90% dos registros com alguma imputação em cada aplicação tiveram

Gráfico 3



Fonte: Censo 91 - Resultados da Correção Automática do CD 1.01, usando o sistema DIA.

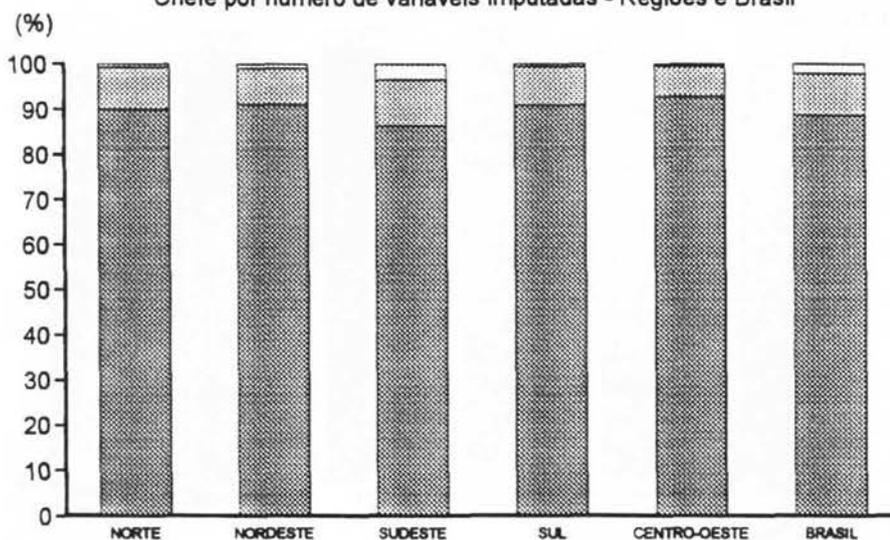
Gráfico 4



Fonte: Censo 91 - Resultados da Correção Automática do CD 1.01, usando o sistema DIA.

Gráfico 5

Distribuição de registros com alguma imputação na aplicação de Chefe por número de variáveis imputadas - Regiões e Brasil



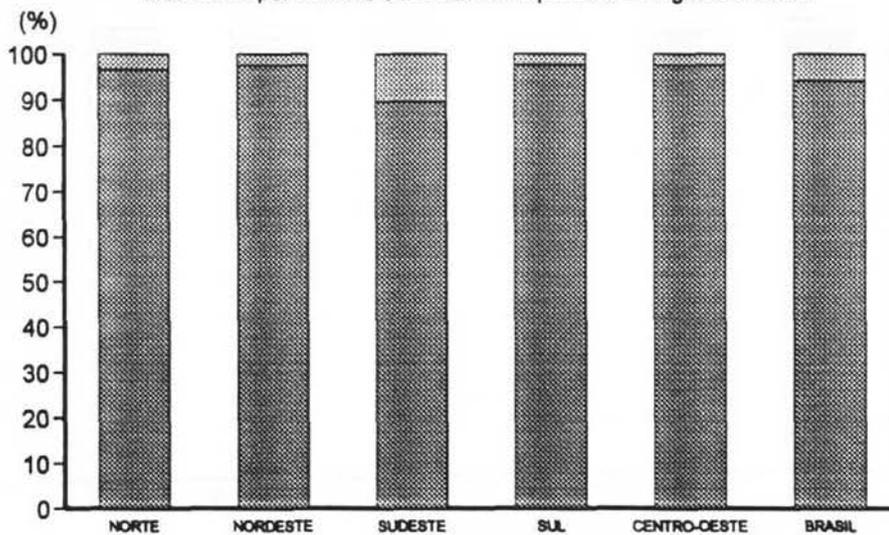
Legenda: Número de variáveis imputadas

■ 1    ■ 2    □ 3 a 5

Fonte: Censo 91 - Resultados da Correção Automática do CD 1.01, usando o sistema DIA.

Gráfico 6

Distribuição de registros com alguma imputação na aplicação de Não-chefe por número de variáveis imputadas - Regiões e Brasil



Legenda: Número de variáveis imputadas

■ 1    ■ 2

Fonte: Censo 91 - Resultados da Correção Automática do CD 1.01, usando o sistema DIA.

imputação em apenas uma variável, sendo 89,09% em Domicílio, 88,53% em Chefe e 94,14% em Não-Chefe. Considerando-se que, pelo gráfico 3, 93,12% dos questionários tiveram imputação em apenas uma das aplicações, pode-se inferir que mais de 80% dos questionários, que tiveram imputação, foram corrigidos com a imputação de apenas uma variável. Tais resultados estão de acordo com o princípio na qual se baseia a metodologia de Fellegi e Holt, mencionado em 6.2: "os dados de cada registro devem satisfazer todas as críticas, alterando o menor número possível de variáveis, procurando assim, manter a maior quantidade possível de informação original".

Os gráficos A.7.6, A.7.7 e A.7.8, no Anexo 7, apresentam, de modo análogo, as distribuições de registros com alguma imputação na aplicação de Domicílio, Chefe e Não-Chefe, respectivamente, por número de variáveis imputadas por Unidade da Federação.

Nos gráficos 4 e A.7.6 pode-se verificar um percentual de imputações em 9 variáveis, na aplicação de Domicílio. Isto se justifica pelo que segue: a variável V201 (espécie do domicílio) foi tratada na pré-crítica e permaneceu fixa durante todo o processo automático de correção; em consequência, quaisquer inconsistências detectadas entre a V201 e as demais variáveis correspondentes às características de domicílios só puderam ser corrigidas através da imputação das outras 9 variáveis envolvidas nessa aplicação.

As análises aqui apresentadas não esgotam outras análises mais aprofundadas que podem ser efetuadas diante das informações armazenadas nos arquivos, pois o sistema fornece todos os recursos necessários para uma análise e monitoramento adequados do processo, como se recomenda.

Das análises efetuadas pode-se concluir que o efeito da imputação sobre os resultados foi pequeno e estava dentro do que se esperava, não introduzindo distorções indesejáveis nos dados.

Por fim, cabe ressaltar o grande sucesso obtido na apuração do questionário básico - CD 1.01 do Censo Demográfico de 1991 com a utilização do sistema DIA que, além de integrar perfeitamente o processo de detecção e correção dos erros em apenas um ciclo

de processamento, permitiu a execução definitiva dessa fase em pouco mais de um mês de processamento e a avaliação dos ganhos significativos de qualidade do processo.

## 7. O ARMAZENAMENTO E A LIBERAÇÃO DOS DADOS PARA ANÁLISE, TABULAÇÃO E DISSEMINAÇÃO

Para o armazenamento dos dados definitivos do CD 1.01 foi utilizado o Banco de Dados RAPID - *Relational Access for Integrated Databases*, que permite o acesso direto aos dados.

Além dos dados investigados, devidamente criticados e "limpos", foram carregadas, no Banco RAPID, algumas variáveis auxiliares criadas durante as etapas de crítica, algumas variáveis de identificação territorial provenientes de outros arquivos, bem como algumas variáveis criadas especificamente para armazenamento.

A seguir são apresentadas as variáveis que foram acrescentadas para serem armazenadas no Banco RAPID<sup>36</sup>:

### **V7001 (mesorregião) e V7002 (microrregião)**

Obtidas do arquivo de Estruturas Territoriais criado a partir do Banco de Estruturas Territoriais (BET), sob a responsabilidade da DGC.

### **V3031 (tipo de idade)**

Essa variável permite a identificação do tipo de informação que foi utilizada para o cálculo da idade.

Códigos possíveis: 1 (se a idade foi calculada com base na idade presumida) e 2 (se a idade foi calculada com base na declaração do mês e ano de nascimento).

### **V3032 (idade em anos)**

A variável V3032 foi construída a partir das informações da variável V3043 (idade calculada) e permite a identificação do número de anos de idade de cada pessoa do arquivo de dados.

Códigos válidos:

V3032 = 000 se  $900 \leq V3043 \leq 911$  (menor que 1 ano)

V3032 = i se  $V3043 = i \quad \forall 001 \leq i \leq 130$ .

---

<sup>36</sup> Ver Formatação para o Banco RAPID em IBGE(1992).

**V3033 (idade em meses)**

A variável V3033 foi construída a partir das informações da variável V3043 (idade calculada) e permite a identificação do número de meses dos menores de 1 ano.

Códigos válidos:

V3033 = j se V3043 = 900 + j  $\forall 0 \leq j \leq 11$

V3033 = branco se  $001 \leq V3043 \leq 130$ .

**V7003 (Área Metropolitana)**

Dada a grande utilização das informações a nível geográfico de Área Metropolitana (AM), decidiu-se incorporar a identificação da cada uma das 9 áreas, a partir das informações do arquivo fornecido pelo CDDI à DI / COI.

Códigos possíveis: 0 (área não metropolitana); 1 (AM de Belém); 2 (AM de Fortaleza); 3 (AM de Recife); 4 (AM de Salvador); 5 (AM de Belo Horizonte); 6 (AM do Rio de Janeiro); 7 (AM de São Paulo); 8 (AM de Curitiba); 9 (AM de Porto Alegre).

**V3050 (anos de estudo)**

A variável V3050 foi construída a partir das informações das variáveis V306 (última série que concluiu com aprovação) e V307 (grau da última série concluída com aprovação) de acordo com critério apresentado no quadro a seguir.

Quadro 3 - Correspondência entre os códigos da V3050 e V306 e V307

Anos de estudo (V3050)	Série (V306)	Grau (V307)
0	0	0
1	0	1
1	1	2
1	1	4
2	2	2
2	2	4
3	3	2
3	3	4
4	4	2
4	5	2
4	4	4
5	5	4
5	1	3
6	6	4
6	2	3
7	7	4
7	3	3
8	8	4
8	4	3
8	5	3
9	1	5
9	1	6
10	2	5
10	2	6
11	3	5
11	4	5
11	3	6
11	4	6
12	1	7
13	2	7
14	3	7
15	4	7
16	5	7
17	6	7
17	0	8

**V3081 (faixa de rendimento mensal bruto do chefe)**

Com a finalidade de facilitar os usuários do Banco RAPID foi criada a variável V3081, cujos códigos representam faixas de rendimento mensal bruto dos chefes, calculadas a partir dos valores correspondentes para a variável V3083. No quadro 4 são apresentados os códigos da V3081 e os respectivos intervalos representados em salários mínimos e em cruzeiros novos (moeda corrente na época da coleta dos dados).

**Quadro 4** - Correspondência entre os códigos da faixa de rendimento e os valores em salários mínimos e em cruzeiros novos de acordo com a V3083

Códigos da V3081	Valores	
	em salários mínimos	em NCz\$ (V3083)
01	até 1/4	de 1 a 9040
02	mais de 1/4 a 1/2	de 9041 a 18081
03	mais de 1/2 a 3/4	de 18082 a 27121
04	mais de 3/4 a 1	de 27122 a 36162
05	mais de 1 a 1 1/4	de 36163 a 45202
06	mais de 1 1/4 a 1 1/2	de 45203 a 54242
07	mais de 1 1/2 a 2	de 54243 a 72323
08	mais de 2 a 3	de 72324 a 108485
09	mais de 3 a 5	de 108486 a 180808
10	mais de 5 a 10	de 180809 a 361616
11	mais de 10 a 15	de 361617 a 542424
12	mais de 15 a 20	de 542425 a 723232
13	mais de 20	de 723233 a 9999997
14	sem rendimento	0
15	sem declaração	9999999
branco	individual com menos de 5 anos ou não-chefe	branco

As variáveis que deixaram de ser armazenadas no Banco RAPID são as seguintes: V110 (nº no CD 1.03); V211 (existência de criança com menos de 2 anos no domicílio); V3082 (nº de dígitos do rendimento); V3044 (faixa de idade do chefe); V303 (mês de nascimento); V304 (ano de nascimento); V3043 (idade calculada); V2203 (indicadora da existência de rede geral de água); V2204 (indicadora da existência de rede geral de esgoto).

Convém destacar que, antes da carga definitiva dos dados no Banco RAPID os dados limpos foram gravados em um banco de dados provisório a fim de permitir uma avaliação estatística e demográfica da qualidade dos resultados definitivos do Censo Demográfico de 1991 relativos à investigação do CD 1.01. Tal avaliação foi realizada através de análises efetuadas pelo Departamento de População - DEPOP e pelo Departamento de Indicadores Sociais - DEISO.

O DEISO ficou responsável pela análise dos dados referentes às características de domicílios e à alfabetização e instrução dos chefes de domicílio,<sup>37</sup> enquanto o DEPOP ficou responsável pela análise dos demais dados referentes às características de pessoas<sup>38</sup>.

Após a liberação dos dados pela CTD, DEPOP e DEISO os dados apurados do CD 1.01 foram carregados no Banco RAPID definitivo e liberados para a tabulação e divulgação.

O quadro 5 apresenta a data, por Unidade da Federação, da carga final dos dados do CD 1.01 no Banco RAPID.

---

<sup>37</sup> Ver *Análise de consistência do CD 1.01. Procedimentos da DISOC/DEISO*.

<sup>38</sup> Uma descrição do plano de análise dos resultados do CD 1.01 encontra-se no documento, elaborado pelo DEPOP: *Censo Demográfico de 1991: análises preliminares (volume I)*.

**Quadro 5 - Data da carga final dos dados do CD 1.01 no Banco RAPID por Unidade da Federação**

Unidades da Federação	Carga final no Banco RAPID
Rondônia	24.01.94
Acre	06.01.94
Amazonas	23.02.94
Roraima	31.01.94
Pará	01.03.94
Amapá	24.01.94
Tocantins	27.01.94
Maranhão	14.12.93
Piauí	20.10.93
Ceará	03.02.94
Rio Grande do Norte	08.12.93
Paraíba	12.01.94
Pernambuco	25.02.94
Alagoas	16.12.93
Sergipe	27.12.93
Bahia	08.02.94
Minas Gerais	18.01.94
Espírito Santo	07.12.93
Rio de Janeiro	01.02.94
São Paulo	31.01.94
Paraná	13.01.94
Santa Catarina	14.01.94
Rio Grande do Sul	01.02.94
Mato Grosso do Sul	09.12.93
Mato Grosso	09.12.93
Goiás	02.02.94
Distrito Federal	08.12.93

## REFERÊNCIAS

- ALBIERI, S. e Bianchini, Z.M. *Censo Demográfico de 1991. Sobre a independência da apuração do CD 1.01 e CD 1.02*. Rio de Janeiro: IBGE, Divisão de Metodologia, 1993. 20p.
- ANÁLISE de consistência do CD-91. Procedimentos da DISOC/DEISO. Rio de Janeiro: IBGE, Departamento de Indicadores Sociais, [1994].
- BARBOSA, D.M.R. Relatório de viagem. Participação na Conferência "Computerising survey support system", do SGCSA (study group on computer and survey analysis). Visita técnica ao INE. Rio de Janeiro: IBGE, Departamento de Tecnologia, Divisão de Sistemas de Apoio, 1993. 9p.
- CENSO Demográfico de 1991: análises preliminares (volume I). Rio de Janeiro: IBGE, 1992.
- CENSO Demográfico 1991: resultados preliminares. Rio de Janeiro: IBGE, 1992. 85p.
- CRIADO, I.V. e Cabria, M.S.B. *Procedimiento de depuración de datos estadísticos*. Vitoria-Gasteiz: EUSTAT - Instituto Vasco de Estadística, 1990. 169p.
- LAS FRECUENCIAS para la imputacion DIA. Madrid: Instituto Nacional de Estadística (INE), [1993]. (mimeo). 15p.
- FELLEGI, I.P. e Holt, D. A systematic approach to automatic edit and imputation. *Journal of the American Statistical Association*, v.71, p.17-35. 1976.
- HANONO, R.M. *DIA - Integração à arquitetura de informática do IBGE*. Rio de Janeiro: IBGE, 1993. (INFOTEC v.2 n. 9).

HANONO, R.M.; Miranda, J.B. e Ferreira, S.B. *Relatório da missão à Espanha e Hungria*. Rio de Janeiro: IBGE, 1990. 22p.

IBGE. *Análise da correção automática no Censo Demográfico de 1991*. Rio de Janeiro: IBGE, Coordenação Técnica do Censo Demográfico, [1993].

IBGE. *Comparação entre o processo de correção automática de dados utilizado no CD-80 e o sistema DIA*. Rio de Janeiro: IBGE, 1990.

IBGE. *Procedimentos e criação de variáveis nas etapas formação de lotes, seleção de lotes, correção automática e formatação para o Banco RAPID*. Rio de Janeiro: IBGE, DI/DEATE/DIDEM, 1992. 9p.

IBGE. *X Recenseamento Geral do Brasil de 1991. Instruções de empastamento. Questionário básico CD 1.01*. Rio de Janeiro: IBGE, Coordenação Técnica do Censo Demográfico, [1991a].

IBGE. *X Recenseamento Geral do Brasil de 1991. Instruções de pré-critica. Questionário básico CD 1.01*. Rio de Janeiro: IBGE, Coordenação Técnica do Censo Demográfico, [1991b].

MANSOLDO, H.F. *X Recenseamento Geral do Brasil. Censo Demográfico*. Rio de Janeiro: IBGE, Diretoria de Informática, 1990.

METODOLOGIA do Censo Demográfico de 1980. Rio de Janeiro: IBGE, 1983. 477p. (Série Relatórios Metodológicos, n. 3).

OLIVEIRA, E.M. e Oliveira, L.C.S. *Relatório de viagem ao INE-Espanha - Instituto Nacional de Estadística da Espanha*. Rio de Janeiro: IBGE, 1990, 20p.

RUBIO, E.G. e Criado, I.V. *Sistema DIA - sistema de Detecção e Imputación Automática de errores para datos cualitativos*.

Volumen I. *DIA: descripción del sistema*. Madrid: Instituto Nacional de Estadística (INE), 1988.

LA SELECCION de variables DIA. Madrid: Instituto Nacional de Estadística (INE), [1993]. (mimeo). 23p.

LA SELECCION de codigos DIA. Madrid: Instituto Nacional de Estadística (INE), [1993]. (mimeo). 9p.

SILVA, P.L.N. *Relatório de viagem aos Estados Unidos no período de 12-25 de março de 1988*. Rio de Janeiro: IBGE, Núcleo de Metodologia, 1988. 40p.

SILVA, P.L.N. e Bianchini, Z.M. *Qualidade de dados e o envolvimento da Divisão de Metodologia no planejamento do Censo Demográfico de 1991. (Versão final)*. Rio de Janeiro: IBGE, Divisão de Metodologia, 1992. 27p.

SILVA, P.L.N. e Bianchini, Z.M. *Apuração de dados no IBGE: Problemas e perspectivas*. Rio de Janeiro: IBGE, 1994. 62p. (Textos para Discussão, n. 68).

SILVA, P.L.N. e Hanono, R.M. *Relatório de visita de consultoria técnica de Emiliano Garcia Rubio ao IBGE para assessoria sobre a utilização do sistema DIA no processamento dos dados do Censo Demográfico de 1991*. Rio de Janeiro: IBGE, 1991a. 11p.

SILVA, P.L.N. e Hanono, R.M. *Relatório de visita de consultoria técnica de Águeda Farina Martinez ao IBGE para instalação da versão 2 do sistema DIA*. Rio de Janeiro: IBGE, 1991b. 3p.

SILVA, P.L.N.; Indá, L.B. e Lima, V.M.U.G. *Projeto para o controle de qualidade da transcrição dos dados do questionário básico CD-1.01 no Censo Demográfico de 1991 (versão preliminar)*. Rio de Janeiro: IBGE, 1991. 28p.

SILVA, P.L.N.; Oliveira, E.M. e Oliveira, L.C.S. *Sobre o critério de formação de lotes de apuração do questionário básico CD 1.01 (versão 2)*. Rio de Janeiro: IBGE, 1992. 10p.

SILVA, P.L.N; Oliveira, E.M.; Oliveira, L.C.S. e Lima, R.L.A. *Uma nova metodologia para correção automática no Censo Demográfico brasileiro: experimentação e primeiros resultados*. Rio de Janeiro: IBGE, 1990. 102p. (Textos para Discussão, n. 28).

**ANEXOS**

**Anexo 1 - Cópia do Questionário Básico - CD 1.01**

4 DIS-TRITO	5 SUB-DISTRITO	8 N.º DO SETOR	7 QUAR-TEIRÃO	8 FACE	9 N.º NO CD 1 (A)	10 N.º NO CD 1 (B)	PESSOAS RESIDENTES	11 Mascu- lino	12 Feme- nino	1 MUNICÍPIO	2 PASTA	3 N.º NA PASTA
-------------	----------------	----------------	---------------	--------	-------------------	--------------------	--------------------	----------------	---------------	-------------	---------	----------------

CENSO DEMOGRÁFICO

CD 1.01 QUESTIONÁRIO BÁSICO

1

Localidade: \_\_\_\_\_ Nome do Informante: \_\_\_\_\_

Logradouro: \_\_\_\_\_ N.º: \_\_\_\_\_ Dependência: \_\_\_\_\_ Assinatura do Informante: \_\_\_\_\_

PARA USO DO ÓRGÃO CENTRAL

QUESTIONÁRIO SUPLEMENTAR

Não tem  Tem  É

2 CARACTERÍSTICAS DO DOMICÍLIO

1 ESPÉCIE	2 LOCALIZAÇÃO	3 ABASTECIMENTO DE ÁGUA	4 INSTALAÇÃO SANITÁRIA	
1 <input type="checkbox"/> Permanente 2 <input type="checkbox"/> Improvisado 3 <input type="checkbox"/> Coletivo (As questões seguintes só serão preenchidas para o domicílio particular permanente)	Casa 1 <input type="checkbox"/> Isolado ou de condomínio 2 <input type="checkbox"/> Em conjunto residencial popular 3 <input type="checkbox"/> Em aglomerado subnormal Apartamento 4 <input type="checkbox"/> Isolado ou de condomínio 5 <input type="checkbox"/> Em conjunto residencial popular 6 <input type="checkbox"/> Em aglomerado subnormal 7 <input type="checkbox"/> Condomínio	Com instalação interna 1 <input type="checkbox"/> Rede geral 2 <input type="checkbox"/> Poço de nascente 3 <input type="checkbox"/> Outra fonte Sem instalação interna 4 <input type="checkbox"/> Rede geral 5 <input type="checkbox"/> Poço ou nascente 6 <input type="checkbox"/> Outra fonte	4 ESCOADOURO Fossa séptica 1 <input type="checkbox"/> Rede geral 2 <input type="checkbox"/> Ligada à rede geral 3 <input type="checkbox"/> Sem escoadouro 4 <input type="checkbox"/> Fossa rudimentar 5 <input type="checkbox"/> Vaso negro 6 <input type="checkbox"/> Outro 7 <input type="checkbox"/> Não sabe 8 <input type="checkbox"/> Não tem	
5 CONDIÇÃO DE OCUPAÇÃO	6 TOTAL DE CÔMODOS	7 CÔMODOS SERVINDO DE DORMITÓRIO	8 BANHEIROS	9 DESTINO DO LIXO
Próprio 1 <input type="checkbox"/> A construção e o terreno 2 <input type="checkbox"/> Só a construção 3 <input type="checkbox"/> Alugado Cedido 4 <input type="checkbox"/> Por empregador 5 <input type="checkbox"/> Por particular 6 <input type="checkbox"/> Outra	Numero de cômodos (Quando o número de cômodos for inferior a 10, registrar 0 no primeiro campo)	1 <input type="checkbox"/> 1 cômodo 2 <input type="checkbox"/> 2 cômodos 3 <input type="checkbox"/> 3 cômodos 4 <input type="checkbox"/> 4 cômodos 5 <input type="checkbox"/> 5 cômodos 6 <input type="checkbox"/> 6 cômodos 7 <input type="checkbox"/> 7 cômodos 8 <input type="checkbox"/> 8 cômodos 9 <input type="checkbox"/> 9 cômodos ou mais	1 <input type="checkbox"/> 1 banheiro 2 <input type="checkbox"/> 2 banheiros 3 <input type="checkbox"/> 3 banheiros 4 <input type="checkbox"/> 4 banheiros 5 <input type="checkbox"/> 5 banheiros ou mais 6 <input type="checkbox"/> Não tem	Coletivo 1 <input type="checkbox"/> Diretamente 2 <input type="checkbox"/> Indirectamente 3 <input type="checkbox"/> Queimado Alugado 4 <input type="checkbox"/> Enterrado 5 <input type="checkbox"/> Terreno baldio 6 <input type="checkbox"/> Rio, lago ou mar 7 <input type="checkbox"/> Outro
10	11	11 NESTE DOMICÍLIO RESIDE CRIANÇA COM MENOS DE 1 ANO, INCLUSIVE ALGUMA RECÉM-NASCIDA?		
		1 <input type="checkbox"/> Sim 2 <input type="checkbox"/> Não		

3 1.ª PESSOA CARACTERÍSTICAS DO CHEFE DO DOMICÍLIO OU INDIVIDUAL NOME

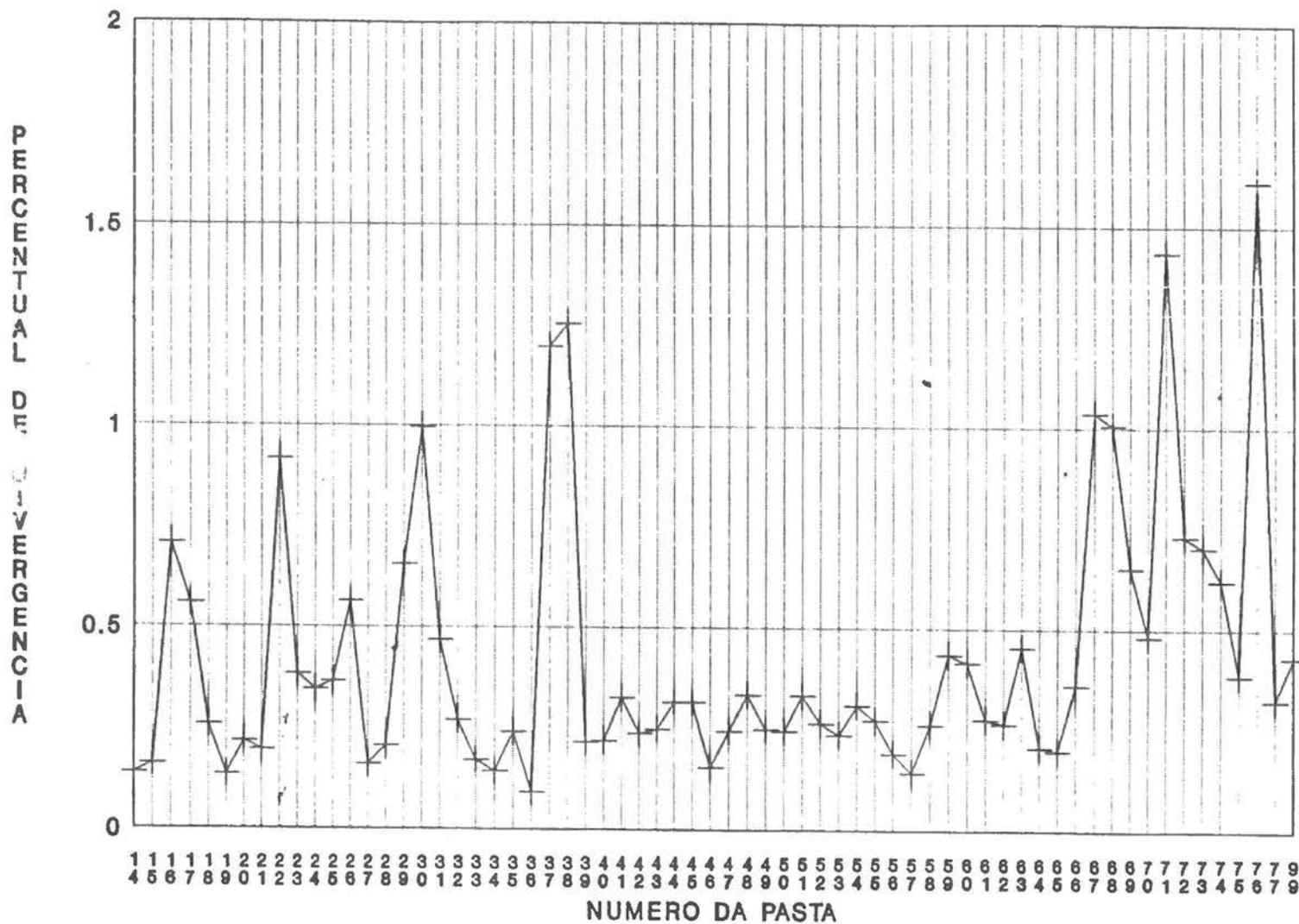
1 Sexo	2 Chefe de domicílio ou indivíduo	3 Mês e ano de nascimento (Não registre o dia)	4 Mês e ano de nascimento (Se não souber e não souber o ano de nascimento, indique quantos anos após ter sido inferior a 1 ano, o número de meses)	5 Sabe ler e escrever ou não sabe	6 Última série que concluiu com aprovação (Para a pessoa que concluiu cursos de alfabetização de adultos, mestrado ou doutorado, assinale a redação 1 - Nenhum)	7 Grau da última série concluída com aprovação	8 Rendimentos mensais brutos (Para os rendimentos fixos, o do mês de agosto de 1990. Para os variáveis, a média dos últimos 12 meses)
1 <input type="checkbox"/> Masculino 2 <input type="checkbox"/> Feminino	01 <input type="checkbox"/> Chefe 20 <input type="checkbox"/> Indivíduo	MM Ano	20 Anos 400 Meses	1 <input type="checkbox"/> Ler e escrever 2 <input type="checkbox"/> Não sabe 3 <input type="checkbox"/> Menos de 5 anos (As questões seguintes só serão preenchidas para as pessoas de 5 anos ou mais)	1 <input type="checkbox"/> 1.ª série 2 <input type="checkbox"/> 2.ª série 3 <input type="checkbox"/> 3.ª série 4 <input type="checkbox"/> 4.ª série 5 <input type="checkbox"/> 5.ª série 6 <input type="checkbox"/> 6.ª série 7 <input type="checkbox"/> 7.ª série 8 <input type="checkbox"/> 8.ª série 9 <input type="checkbox"/> Nenhuma	1 <input type="checkbox"/> Curso de alfabetização de adultos 2 <input type="checkbox"/> Primário ou semestral 3 <input type="checkbox"/> Ginasial ou médio 1.º ciclo 4 <input type="checkbox"/> 1.º grau 5 <input type="checkbox"/> 2.º grau 6 <input type="checkbox"/> Colégio ou médio 2.º ciclo 7 <input type="checkbox"/> Superior 8 <input type="checkbox"/> Mestrado ou doutorado 9 <input type="checkbox"/> Nenhum	0 <input type="checkbox"/> Não tem N.º de dígitos R\$

QUESTÕES	2.ª PESSOA	3.ª PESSOA	4.ª PESSOA	5.ª PESSOA
<b>NOME</b>				
<b>1</b> Sexo	1 <input type="checkbox"/> Masculino 2 <input type="checkbox"/> Feminino	1 <input type="checkbox"/> Masculino 2 <input type="checkbox"/> Feminino	1 <input type="checkbox"/> Masculino 2 <input type="checkbox"/> Feminino	1 <input type="checkbox"/> Masculino 2 <input type="checkbox"/> Feminino
<b>2</b> Parentesco ou relação com o Chefe do domicílio	02 <input type="checkbox"/> Cônjuge 03 <input type="checkbox"/> Filho(a) 04 <input type="checkbox"/> Enteadado(a) 05 <input type="checkbox"/> Pai ou mãe 06 <input type="checkbox"/> Sogro(a) 07 <input type="checkbox"/> Avô(a) ou bisavô(a) 08 <input type="checkbox"/> Neto(a) ou beneto(a) 09 <input type="checkbox"/> Genro ou nora 10 <input type="checkbox"/> Irmão ou irmã 11 <input type="checkbox"/> Cunhado(a) 12 <input type="checkbox"/> Outros parentes 13 <input type="checkbox"/> Agregado(a) 14 <input type="checkbox"/> Parente(a) doméstico(a) 15 <input type="checkbox"/> Empregado(a) 16 <input type="checkbox"/> Parente(a) doméstico(a) empregado(a)	03 <input type="checkbox"/> Filho(a) 04 <input type="checkbox"/> Enteadado(a) 05 <input type="checkbox"/> Pai ou mãe 06 <input type="checkbox"/> Sogro(a) 07 <input type="checkbox"/> Avô(a) ou bisavô(a) 08 <input type="checkbox"/> Neto(a) ou beneto(a) 09 <input type="checkbox"/> Genro ou nora 10 <input type="checkbox"/> Irmão ou irmã 11 <input type="checkbox"/> Cunhado(a) 12 <input type="checkbox"/> Outros parentes 13 <input type="checkbox"/> Agregado(a) 14 <input type="checkbox"/> Parente(a) doméstico(a) 15 <input type="checkbox"/> Empregado(a) 16 <input type="checkbox"/> Parente(a) doméstico(a) empregado(a)	03 <input type="checkbox"/> Filho(a) 04 <input type="checkbox"/> Enteadado(a) 05 <input type="checkbox"/> Pai ou mãe 06 <input type="checkbox"/> Sogro(a) 07 <input type="checkbox"/> Avô(a) ou bisavô(a) 08 <input type="checkbox"/> Neto(a) ou beneto(a) 09 <input type="checkbox"/> Genro ou nora 10 <input type="checkbox"/> Irmão ou irmã 11 <input type="checkbox"/> Cunhado(a) 12 <input type="checkbox"/> Outros parentes 13 <input type="checkbox"/> Agregado(a) 14 <input type="checkbox"/> Parente(a) doméstico(a) 15 <input type="checkbox"/> Empregado(a) 16 <input type="checkbox"/> Parente(a) doméstico(a) empregado(a)	03 <input type="checkbox"/> Filho(a) 04 <input type="checkbox"/> Enteadado(a) 05 <input type="checkbox"/> Pai ou mãe 06 <input type="checkbox"/> Sogro(a) 07 <input type="checkbox"/> Avô(a) ou bisavô(a) 08 <input type="checkbox"/> Neto(a) ou beneto(a) 09 <input type="checkbox"/> Genro ou nora 10 <input type="checkbox"/> Irmão ou irmã 11 <input type="checkbox"/> Cunhado(a) 12 <input type="checkbox"/> Outros parentes 13 <input type="checkbox"/> Agregado(a) 14 <input type="checkbox"/> Parente(a) doméstico(a) 15 <input type="checkbox"/> Empregado(a) 16 <input type="checkbox"/> Parente(a) doméstico(a) empregado(a)
<b>3</b> Mês e ano de nascimento (Não registre o dia)	Mês: <input type="text"/> <input type="text"/> <input type="text"/> Ano: <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	Mês: <input type="text"/> <input type="text"/> <input type="text"/> Ano: <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	Mês: <input type="text"/> <input type="text"/> <input type="text"/> Ano: <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	Mês: <input type="text"/> <input type="text"/> <input type="text"/> Ano: <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
<b>4</b> Idade pretendida (Se não souber o mês e/ou o ano de nascimento, indique quantos anos supõe ter. Se inferior a 1 ano, o número de meses)	1 ano ou mais: <input type="text"/> <input type="text"/> <input type="text"/> anos Menos de 1 ano: <input type="text"/> <input type="text"/> <input type="text"/> meses	1 ano ou mais: <input type="text"/> <input type="text"/> <input type="text"/> anos Menos de 1 ano: <input type="text"/> <input type="text"/> <input type="text"/> meses	1 ano ou mais: <input type="text"/> <input type="text"/> <input type="text"/> anos Menos de 1 ano: <input type="text"/> <input type="text"/> <input type="text"/> meses	1 ano ou mais: <input type="text"/> <input type="text"/> <input type="text"/> anos Menos de 1 ano: <input type="text"/> <input type="text"/> <input type="text"/> meses
<b>5</b> Sabe ler e escrever ou não sabe	1 <input type="checkbox"/> Ler e escrever 2 <input type="checkbox"/> Não sabe 3 <input type="checkbox"/> Menos de 5 anos	1 <input type="checkbox"/> Ler e escrever 2 <input type="checkbox"/> Não sabe 3 <input type="checkbox"/> Menos de 5 anos	1 <input type="checkbox"/> Ler e escrever 2 <input type="checkbox"/> Não sabe 3 <input type="checkbox"/> Menos de 5 anos	1 <input type="checkbox"/> Ler e escrever 2 <input type="checkbox"/> Não sabe 3 <input type="checkbox"/> Menos de 5 anos

QUESTÕES	6.ª PESSOA	7.ª PESSOA	8.ª PESSOA	9.ª PESSOA
<b>NOME</b>				
<b>1</b> Sexo	1 <input type="checkbox"/> Masculino 2 <input type="checkbox"/> Feminino			
<b>2</b> Parentesco ou relação com o Chefe do domicílio	03 <input type="checkbox"/> Filho(a) 04 <input type="checkbox"/> Enteadado(a) 05 <input type="checkbox"/> Pai ou mãe 06 <input type="checkbox"/> Sogro(a) 07 <input type="checkbox"/> Avô(a) ou bisavô(a) 08 <input type="checkbox"/> Neto(a) ou beneto(a) 09 <input type="checkbox"/> Genro ou nora 10 <input type="checkbox"/> Irmão ou irmã 11 <input type="checkbox"/> Cunhado(a) 12 <input type="checkbox"/> Outros parentes 13 <input type="checkbox"/> Agregado(a) 14 <input type="checkbox"/> Parente(a) doméstico(a) 15 <input type="checkbox"/> Empregado(a) 16 <input type="checkbox"/> Parente(a) doméstico(a) empregado(a)	03 <input type="checkbox"/> Filho(a) 04 <input type="checkbox"/> Enteadado(a) 05 <input type="checkbox"/> Pai ou mãe 06 <input type="checkbox"/> Sogro(a) 07 <input type="checkbox"/> Avô(a) ou bisavô(a) 08 <input type="checkbox"/> Neto(a) ou beneto(a) 09 <input type="checkbox"/> Genro ou nora 10 <input type="checkbox"/> Irmão ou irmã 11 <input type="checkbox"/> Cunhado(a) 12 <input type="checkbox"/> Outros parentes 13 <input type="checkbox"/> Agregado(a) 14 <input type="checkbox"/> Parente(a) doméstico(a) 15 <input type="checkbox"/> Empregado(a) 16 <input type="checkbox"/> Parente(a) doméstico(a) empregado(a)	03 <input type="checkbox"/> Filho(a) 04 <input type="checkbox"/> Enteadado(a) 05 <input type="checkbox"/> Pai ou mãe 06 <input type="checkbox"/> Sogro(a) 07 <input type="checkbox"/> Avô(a) ou bisavô(a) 08 <input type="checkbox"/> Neto(a) ou beneto(a) 09 <input type="checkbox"/> Genro ou nora 10 <input type="checkbox"/> Irmão ou irmã 11 <input type="checkbox"/> Cunhado(a) 12 <input type="checkbox"/> Outros parentes 13 <input type="checkbox"/> Agregado(a) 14 <input type="checkbox"/> Parente(a) doméstico(a) 15 <input type="checkbox"/> Empregado(a) 16 <input type="checkbox"/> Parente(a) doméstico(a) empregado(a)	03 <input type="checkbox"/> Filho(a) 04 <input type="checkbox"/> Enteadado(a) 05 <input type="checkbox"/> Pai ou mãe 06 <input type="checkbox"/> Sogro(a) 07 <input type="checkbox"/> Avô(a) ou bisavô(a) 08 <input type="checkbox"/> Neto(a) ou beneto(a) 09 <input type="checkbox"/> Genro ou nora 10 <input type="checkbox"/> Irmão ou irmã 11 <input type="checkbox"/> Cunhado(a) 12 <input type="checkbox"/> Outros parentes 13 <input type="checkbox"/> Agregado(a) 14 <input type="checkbox"/> Parente(a) doméstico(a) 15 <input type="checkbox"/> Empregado(a) 16 <input type="checkbox"/> Parente(a) doméstico(a) empregado(a)
<b>3</b> Mês e ano de nascimento (Não registre o dia)	Mês: <input type="text"/> <input type="text"/> <input type="text"/> Ano: <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	Mês: <input type="text"/> <input type="text"/> <input type="text"/> Ano: <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	Mês: <input type="text"/> <input type="text"/> <input type="text"/> Ano: <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	Mês: <input type="text"/> <input type="text"/> <input type="text"/> Ano: <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
<b>4</b> Idade pretendida (Se não souber o mês e/ou o ano de nascimento, indique quantos anos supõe ter. Se inferior a 1 ano, o número de meses)	1 ano ou mais: <input type="text"/> <input type="text"/> <input type="text"/> anos Menos de 1 ano: <input type="text"/> <input type="text"/> <input type="text"/> meses	1 ano ou mais: <input type="text"/> <input type="text"/> <input type="text"/> anos Menos de 1 ano: <input type="text"/> <input type="text"/> <input type="text"/> meses	1 ano ou mais: <input type="text"/> <input type="text"/> <input type="text"/> anos Menos de 1 ano: <input type="text"/> <input type="text"/> <input type="text"/> meses	1 ano ou mais: <input type="text"/> <input type="text"/> <input type="text"/> anos Menos de 1 ano: <input type="text"/> <input type="text"/> <input type="text"/> meses
<b>5</b> Sabe ler e escrever ou não sabe	1 <input type="checkbox"/> Ler e escrever 2 <input type="checkbox"/> Não sabe 3 <input type="checkbox"/> Menos de 5 anos	1 <input type="checkbox"/> Ler e escrever 2 <input type="checkbox"/> Não sabe 3 <input type="checkbox"/> Menos de 5 anos	1 <input type="checkbox"/> Ler e escrever 2 <input type="checkbox"/> Não sabe 3 <input type="checkbox"/> Menos de 5 anos	1 <input type="checkbox"/> Ler e escrever 2 <input type="checkbox"/> Não sabe 3 <input type="checkbox"/> Menos de 5 anos

**Anexo 2 - Gráfico das Percentagens de Divergências na  
Digitação por pasta dos dados do Censo Experimental de Limeira  
de 1988**

# PERCENTAGENS DE DIVERGENCIAS POR PASTA TOTAL (IDENT. + PRE + ABERTOS)



**Anexo 3 - Plano de Transcrição de Dados do CD 1.01**

PLANO DE TRANSCRIÇÃO DE DADOS

QUESTIONARIO BASICO

CD - 1.01

S L I P   A U T O M A T I C O

MUNICIPIO - se as duas primeiras posições não pertencem a tabela de UF's ===> ERRO

- se a 7a. posição não for DV modulo 10 das 6 primeiras ===> ERRO

PASTA - se as duas primeiras posições não forem iguais as duas primeiras de MUNICIPIO, a exceção quando as duas primeiras posições da PASTA = ( 36 ) ===> ERRO

- se as duas primeiras posições forem 36 e as duas primeiras posições de MUNICIPIO Dif. (35) ===> ERRO

DISTRITO - se ( <= 0 ou = B ) ===> ERRO

SUBDISTRITO - se ( < 0 ou = B ) ===> ERRO

SETOR - se ( <= 0 ou = B ) ===> ERRO

Q\_INICIAL - se ( <= 0 ou = B ) ===> ERRO

Q\_FINAL - se ( < Q\_INICIAL ou > 660 ou = B ) ===> ERRO

TOT\_QUEST - se Dif. ( Q\_FINAL - Q\_INICIAL + 1 ) ===> ERRO

Q U A D R O   1

V101 - se Dif. MUNICIPIO (SLIP) ===> ERRO

V102 - se Dif. PASTA (SLIP) ===> ERRO

V103 - se ( < Q\_INICIAL (SLIP) ou > Q\_FINAL (SLIP) ) ===> ERRO

- se primeiro questionario Dif.Q\_INICIAL (SLIP) ===> ERRO

- se Dif. ((V103 do questionario anterior) + 1) ===> ERRO  
(Fora de Sequencia)

V104 - se Dif. DISTRITO (SLIP) ===> ERRO  
 V105 - se Dif. SUBDISTRITO (SLIP) ===> ERRO  
 V106 - se Dif. SETOR (SLIP) ===> ERRO  
 V107 |  
 V108 | Não tem critica  
 V109 - se ( <= 0 ou = B ) ===> ERRO  
 V110 - se ( < 0 ou = B ) ===> ERRO  
 V111 - se ( < 0 ou = B ) ===> ERRO  
 V112 - se ( < 0 ou = B ) ===> ERRO

Q U A D R O 2

V201 - se Dif. (1-3,B) ===> ERRO  
 V202 - se Dif. (1-7,B) ===> ERRO  
 V203 - se Dif. (1-6,B) ===> ERRO  
 V204 - se Dif. (0-7,B) ===> ERRO  
 V205 - se Dif. (0-2,B) ===> ERRO  
 V206 - se Dif. (1-6,B) ===> ERRO  
 V207 - se Dif. ( > 0,B ) ===> ERRO  
 - se ( = 00 ou > 30 ) ===> ADVERTENCIA ( estes codigos  
 deverão ser aceitos somente apos confirmação )  
 V208 - se Dif. (0-8,B) ===> ERRO  
 V209 - se Dif. (0-5,B) ===> ERRO  
 V210 - se Dif. (1-7,B) ===> ERRO  
 V211 - se Dif. (0,1,B) ===> ERRO

Q U A D R O 3

V300 - se Dif. (1-99) ===> ERRO  
 - se primeira pessoa Dif. (1) ===> ERRO

- se ultima pessoa Dif. ( V111 + V112 ) ===> ERRO
- se Dif. ((V300 da pessoa anterior) + 1)===> ERRO  
(Fora de Sequencia)
- V301 - se Dif. (1,2,B) ===> ERRO
- V302 - se Dif. (1-16,20,B) ===> ERRO
  - se = (2-16) e V300 = (1) ===> ERRO
  - se = (1,20) e V300 Dif. (1) ===> ERRO
  - se = (2) e V300 Dif. (2) ===> ERRO
- V3031 - se Dif. (1-12,B) ===> ERRO
- V3032 - se Dif. (860-991,B) ===> ERRO
- V304 - se Dif. (20001-20130,40000-40011,B) ===> ERRO
- V305 - se Dif. (1-3,B) ===> ERRO
- V306 - se Dif. (0-8,B) ===> ERRO
- V307 - se Dif. (0-8,B) ===> ERRO
- V3081 - se Dif. (0,B) ===> ERRO
- V3062 - se Dif. (1-7,B) ===> ERRO
- V3083 - se Dif. ( > 0,B ) ===> ERRO
  - se ( >= 1000000 ) ===> ADVERTENCIA (estes valores de-  
verão ser aceitos somente apos confirmaçãc )

ESQUEMA DE TRANSFERENCIA DAS VARIÁVEIS 303 / 304 (IDADE)

303	304	TRANSFERENCIA
preenchido	preenchido ou n ã o	303
incompleto	preenchido	304
incompleto	incompleto	303

ESQUEMA DE TRANSFERENCIA DA VARIÁVEL 308 (RENDIMENTO)

3081 (não tem)	3062 (no.digitos)	3083 (rendimento)	TRANSFERENCIA
0	preenchido	preenchido	3082 e 3083
0	preenchido	branco	3081 e 3083
branco	preenchido	preenchido	3082 e 3083
branco	branco	preenchido	3082 e 3083
branco	preenchido	branco	3082 e 3083
branco	branco	branco	3082 e 3083
0	branco	branco	3081 e 3083
0	branco	preenchido	3081 e 3083

**Anexo 4 - Plano de Pré-Crítica do CD 1.01**

PLANO DE PRÉ - CRITICA

QUESTIONARIO BASICO

CD - 1.01

## CRITICA QUANTITATIVA

### Q U A D R O 1

- 11 - Conferir a sequencia : municipio, distrito, subdistrito e numero do setor, indicando as divergencias.
- 12 - Conferir a sequencia do numero dos questionarios dentro de cada pasta, indicando as faltas.
- 13 - Conferir a sequencia do numero dos questionarios dentro de cada pasta, indicando os excessos.

### Q U A D R O 3

- 14 - Conferir a sequencia do numero da pessoa(Q.300) dentro de cada questionario, indicando as faltas.
- 15 - Conferir a sequencia do numero da pessoa(Q.300) dentro de cada questionario, indicando os excessos.
- 16 - Quesito 111  
Diferente do numero de pessoas com codigo 1 no Q.301
- 17 - Quesito 112  
Diferente do numero de pessoas com codigo 2 no Q.301
- 36 - Quesito 111 + Quesito 112 < ou = 0
- 37 - Quesito 111 + Quesito 112  
Diferente do numero de pessoas registradas no questionario

## CRITICAS DE INCOMPATIBILIDADE

- 18 - Quesito 300 = 01  
Quesito 302  $\hat{=}$  01 ou 20
- 19 - Quesito 300 = 02  
Quesito 302  $\hat{=}$  02 a 16
- 20 - Quesito 300  $\geq$  03  
Quesito 302  $\hat{=}$  03 a 16
- 21 - Quesito 201 = 1 ou 2  
Quesito 110  $\hat{=}$  0000
- 22 - Quesito 201 = 3  
Quesito 110 = 0000

POSSIBILIDADES

Obs.: - Nos questionarios em que a V109 = 9996, so serão aplicadas as seguintes regras : 1 a 5 , 7 , 11 a 17, 21 , 22 , 36 e 37.

- Nos questionarios em que a V109 = 9997 a 9999 , so sera aplicada a regra 1.

NUMERO	QUESITOS	CODIGO VALIDO
1	109	0001 a 9998
2	110	0000 a 9999
3	111	00 a 69
4	112	00 a 69
5	201	1, 2 e 3
6	3083	0000000 a 9999998 e branco
7	301	1 e 2
8	30341	01 a 12, 20 e 40
9	305	1 a 3 e 9
10	30812	0 a 7 e branco

Onde : - Variavel 30341 = duas primeiras posições da Variavel 303/304.

- Variavel 30342 = tres ultimas posições da Variavel 303/304.

- Variavel 30812 = primeira posição da Variavel 308 ( FAIXA ).

- 23 - Quesito 201 = 3 e (Q.111 + Q.112 = 1) e Quesito 300 = 01  
Quesito 302 ê= 20
- 24 - Quesito 201 = 3 e (Q.111 + Q.112 > 1) e Quesito 300 = 01  
Quesito 302 ê= 01
- 25 - Quesito 201 = 1 ou 2 e Quesito 300 = 01  
Quesito 302 ê= 01
- 26 - Quesito 201 = 1 ou 2  
O numero de pessoas no domicilio com Q.302 = 14 é >= 5
- 27 - Quesito 201 = 3  
Quesito 302 = 14 ou 16
- 28 - Quesito 302 = 16  
Se não existir outra pessoa no questionario com codigo 15 no Q.302
- 29 - Quesito 30341 = 01 a 08  
Quesito 30342 ê= 861 a 991
- 30 - Quesito 30341 = 09 a 12  
Quesito 30342 ê= 860 a 990
- 31 - Quesito 30341 = 20  
Quesito 30342 ê= 001 a 130 ou 999
- 32 - Quesito 30341 = 40  
Quesito 30342 ê= 000 a 011
- 33 - Quesito 300 = 02 e Quesito 302 = 02  
Quesito 301 = Quesito 301 da pessoa 01
- 34 - Quesito 300 = 01 e Quesito 305 = 1 ou 2  
Quesito 30812 = branco e Quesito 3083 = branco
- 35 - Quesitos 30812 e 3083 (rendimento do chefe do domicilio)

FAIXA		RENDIMENTO
branco	ê=	branco
0	ê=	0000000
1	ê=	0000001 a 0000009
2	ê=	0000010 a 0000099
3	ê=	0000100 a 0000999
4	ê=	0001000 a 0009999
5	ê=	0010000 a 0099999
6	ê=	0100000 a 0999999
7	ê=	1000000 a 9999998

**Anexo 5 - Planos de Crítica para a Detecção e Correção Automática dos erros pelo Sistema DIA<sup>39</sup>**

---

<sup>39</sup> Este anexo apresenta uma das saídas do sistema DIA que, para cada aplicação, contém:

1. Dicionário criado pelo DIA
2. RIDs originais
3. Regras de crítica derivadas pelas RIDs
4. RIDs resultantes do analisador
5. Regras de incompatibilidade (especificadas pelos analistas temáticos e com as redundâncias já eliminadas)
6. Conjunto Completo de Regras.

**Anexo 5.1 - Aplicação de Domicílio**

EDIT 29  
V209 (2-5)  
V207 (2)

EDIT 30  
V208 (0,4-8)  
V207 (3)

EDIT 31  
V209 (2-5)  
V208 (2)  
V207 (3)

EDIT 32  
V209 (1-5)  
V208 (3)  
V207 (3)

EDIT 33  
V209 (3-5)  
V207 (3)

EDIT 34  
V208 (0,5-8)  
V207 (4)

EDIT 35  
V209 (3-5)  
V208 (2)  
V207 (4)

EDIT 36  
V209 (2-5)  
V208 (3)  
V207 (4)

EDIT 37  
V209 (1-5)  
V208 (4)  
V207 (4)

EDIT 38  
V209 (4-5)  
V207 (4)

EDIT 39  
V208 (0,6-8)  
V207 (5)

EDIT 40  
V209 (4-5)  
V208 (2)  
V207 (5)

EDIT 41  
V209 (3-5)  
V208 (3)  
V207 (5)

EDIT 42

V209 (2-5)  
V208 (4)  
V207 (5)

EDIT 43  
V209 (1-5)  
V208 (5)  
V207 (5)

EDIT 44  
V209 (5)  
V207 (5)

EDIT 45  
V208 (0,7-8)  
V207 (6)

EDIT 46  
V209 (5)  
V208 (2)  
V207 (6)

EDIT 47  
V209 (4-5)  
V208 (3)  
V207 (6)

EDIT 48  
V209 (3-5)  
V208 (4)  
V207 (6)

EDIT 49  
V209 (2-5)  
V208 (5)  
V207 (6)

EDIT 50  
V209 (1-5)  
V208 (6)  
V207 (6)

EDIT 51  
V208 (0,8)  
V207 (7)

EDIT 52  
V209 (5)  
V208 (3)  
V207 (7)

EDIT 53  
V209 (4-5)  
V208 (4)  
V207 (7)

EDIT 54  
V209 (3-5)  
V208 (5)  
V207 (7)

EDIT 55  
V209 (2-5)  
V208 (6)  
V207 (7)

EDIT 56  
V209 (1 5)  
V208 (7)  
V207 (7)

EDIT 57  
V208 (0)  
V207 (8)

EDIT 58  
V209 (5)  
V208 (4)  
V207 (8)

EDIT 59  
V209 (4-5)  
V208 (5)  
V207 (8)

EDIT 60  
V209 (3-5)  
V208 (6)  
V207 (8)

EDIT 61  
V209 (2-5)  
V208 (7)  
V207 (8)

EDIT 62  
V209 (1-5)  
V208 (8)  
V207 (8)

EDIT 63  
V209 (1-5)  
V208 (0)  
V207 (9)

EDIT 64  
V209 (5)  
V208 (5)  
V207 (9)

EDIT 65  
V209 (4-5)  
V208 (6)  
V207 (9)

EDIT 66  
V209 (3-5)  
V208 (7)  
V207 (9)

EDIT 67

V209 (2-5)  
V208 (8)  
V207 (9)

EDIT 68  
V209 (2-5)  
V208 (0)  
V207 (10)

EDIT 69  
V209 (5)  
V208 (6)  
V207 (10)

EDIT 70  
V209 (4-5)  
V208 (7)  
V207 (10)

EDIT 71  
V209 (3-5)  
V208 (8)  
V207 (10)

EDIT 72  
V209 (3-5)  
V208 (0)  
V207 (11)

EDIT 73  
V209 (5)  
V208 (7)  
V207 (11)

EDIT 74  
V209 (4-5)  
V208 (8)  
V207 (11)

EDIT 75  
V209 (4-5)  
V208 (0)  
V207 (12)

EDIT 76  
V209 (5)  
V208 (8)  
V207 (12)

EDIT 77  
V209 (5)  
V208 (0)  
V207 (13)

EDIT 78  
V208 (0)  
V7100 (1-8)

EDIT 79  
V208 (0,2-8)

V7100 (1)

EDIT 80  
V208 (0,3-8)  
V7100 (1-2)

EDIT 81  
V208 (0,4-8)  
V7100 (1-3)

EDIT 82  
V208 (0,5-8)  
V7100 (1-4)

EDIT 83  
V208 (0,6-8)  
V7100 (1-5)

EDIT 84  
V208 (0,7-8)  
V7100 (1-6)

EDIT 85  
V208 (0,8)  
V7100 (1-7)

EDIT 86  
V205 (1)  
V209 (0)  
V204 (1-3)

EDIT 87  
V201 (1)  
V1062 (1)  
V202 (1-2,4-5)

6.- CONJUNTO COMPLETO DE REGLAS  
=====

EDIT	1		
V201	(1)		
V202	( )	GENERADO POR LOS EDITS	001
EDIT	2		
V201	(2-3)		
V202	(1-7)	GENERADO POR LOS EDITS	002
EDIT	3		
V201	(1)		
V203	( )	GENERADO POR LOS EDITS	003
EDIT	4		
V201	(2-3)		
V203	(1-6)	GENERADO POR LOS EDITS	004
EDIT	5		
V201	(1)		
V204	( )	GENERADO POR LOS EDITS	005
EDIT	6		
V201	(2-3)		
V204	(0-7)	GENERADO POR LOS EDITS	006
EDIT	7		
V201	(1)		
V205	( )	GENERADO POR LOS EDITS	007
EDIT	8		
V201	(2-3)		
V205	(0-2)	GENERADO POR LOS EDITS	008
EDIT	9		
V201	(1)		
V206	( )	GENERADO POR LOS EDITS	009
EDIT	10		
V201	(2-3)		
V206	(1-6)	GENERADO POR LOS EDITS	010
EDIT	11		
V201	(1)		
V207	( )	GENERADO POR LOS EDITS	011
EDIT	12		

V201 (2-3)			
V207 (1-30)		GENERADO POR LOS EDITS	012
EDIT 13			
V201 (1)			
V209 ( )		GENERADO POR LOS EDITS	015
EDIT 14			
V201 (2-3)			
V209 (0-5)		GENERADO POR LOS EDITS	016
EDIT 15			
V201 (1)			
V210 ( )		GENERADO POR LOS EDITS	017
EDIT 16			
V201 (2-3)			
V210 (1-7)		GENERADO POR LOS EDITS	018
EDIT 17			
V2203 (2)			
V203 (1,4)		GENERADO POR LOS EDITS	019
EDIT 18			
V205 (1-2)			
V204 (0)		GENERADO POR LOS EDITS	020
EDIT 19			
V209 (1-5)			
V204 (0)		GENERADO POR LOS EDITS	021
EDIT 20			
V205 (0)			
V204 (1-7)		GENERADO POR LOS EDITS	022
EDIT 21			
V2204 (2)			
V204 (1)		GENERADO POR LOS EDITS	023
EDIT 22			
V205 (0)			
V209 (1-5)		GENERADO POR LOS EDITS	024
EDIT 23			
V209 (1-5)			
V207 (1)		GENERADO POR LOS EDITS	026
EDIT 24			

V209 (2-5)			
V207 (2)		GENERADO POR LOS EDITS	029
EDIT 25			
V209 (3-5)		GENERADO POR LOS EDITS	033
V207 (3)			
EDIT 26			
V209 (4-5)		GENERADO POR LOS EDITS	038
V207 (4)			
EDIT 27			
V209 (5)		GENERADO POR LOS EDITS	044
V207 (5)			
EDIT 28			
V205 (1)		GENERADO POR LOS EDITS	086
V209 (0)			
V204 (1-3)			
EDIT 29			
V209 ( )		GENERADO POR LOS EDITS	015 018
V210 (1-7)			
EDIT 30			
V210 (1-7)		GENERADO POR LOS EDITS	011 018
V207 ( )			
EDIT 31			
V206 ( )		GENERADO POR LOS EDITS	009 018
V210 (1-7)			
EDIT 32			
V205 ( )		GENERADO POR LOS EDITS	007 018
V210 (1-7)			
EDIT 33			
V210 (1-7)		GENERADO POR LOS EDITS	005 018
V204 ( )			
EDIT 34			
V203 ( )		GENERADO POR LOS EDITS	003 018
V210 (1-7)			
EDIT 35			
V202 ( )		GENERADO POR LOS EDITS	001 018
V210 (1-7)			

EDIT 36			
V209 (0-5)			
V210 ( )			
	GENERADO POR LOS EDITS	016	017
EDIT 37			
V209 (0-5)			
V207 ( )			
	GENERADO POR LOS EDITS	011	016
EDIT 38			
V206 ( )			
V209 (0-5)			
	GENERADO POR LOS EDITS	009	016
EDIT 39			
V205 ( )			
V209 (0-5)			
	GENERADO POR LOS EDITS	007	016
EDIT 40			
V209 (0-5)			
V204 ( )			
	GENERADO POR LOS EDITS	005	016
EDIT 41			
V203 ( )			
V209 (0-5)			
	GENERADO POR LOS EDITS	003	016
EDIT 42			
V209 (0-5)			
V202 ( )			
	GENERADO POR LOS EDITS	001	016
EDIT 43			
V210 ( )			
V207 (1-30)			
	GENERADO POR LOS EDITS	012	017
EDIT 44			
V209 ( )			
V207 (1-30)			
	GENERADO POR LOS EDITS	012	015
EDIT 45			
V206 ( )			
V207 (1-30)			
	GENERADO POR LOS EDITS	009	012
EDIT 46			
V205 ( )			
V207 (1-30)			
	GENERADO POR LOS EDITS	007	012
EDIT 47			
V204 ( )			
V207 (1-30)			
	GENERADO POR LOS EDITS	005	012

EDIT 48		
V203 ( )		
V207 (1-30)	GENERADO POR LOS EDITS	003 012
EDIT 49		
V202 ( )		
V207 (1-30)	GENERADO POR LOS EDITS	001 012
EDIT 50		
V206 (1-6)		
V210 ( )	GENERADO POR LOS EDITS	010 017
EDIT 51		
V206 (1-6)		
V209 ( )	GENERADO POR LOS EDITS	010 015
EDIT 52		
V206 (1-6)		
V207 ( )	GENERADO POR LOS EDITS	010 011
EDIT 53		
V205 ( )		
V206 (1-6)	GENERADO POR LOS EDITS	007 010
EDIT 54		
V206 (1-6)		
V204 ( )	GENERADO POR LOS EDITS	005 010
EDIT 55		
V203 ( )		
V206 (1-6)	GENERADO POR LOS EDITS	003 010
EDIT 56		
V206 (1-6)		
V202 ( )	GENERADO POR LOS EDITS	001 010
EDIT 57		
V205 (0-2)		
V210 ( )	GENERADO POR LOS EDITS	008 017
EDIT 58		
V205 (0-2)		
V209 ( )	GENERADO POR LOS EDITS	008 015
EDIT 59		
V205 (0-2)		
V207 ( )	GENERADO POR LOS EDITS	008 011

EDIT	60			
V205	(0-2)			
V206	( )			
		GENERADO POR LOS EDITS	008	009
EDIT	61			
V205	(0-2)			
V204	( )			
		GENERADO POR LOS EDITS	005	008
EDIT	62			
V205	(0-2)			
V203	( )			
		GENERADO POR LOS EDITS	003	008
EDIT	63			
V205	(0-2)			
V202	( )			
		GENERADO POR LOS EDITS	001	008
EDIT	64			
V210	( )			
V204	(0-7)			
		GENERADO POR LOS EDITS	006	017
EDIT	65			
V209	( )			
V204	(0-7)			
		GENERADO POR LOS EDITS	006	015
EDIT	66			
V204	(0-7)			
V207	( )			
		GENERADO POR LOS EDITS	006	011
EDIT	67			
V206	( )			
V204	(0-7)			
		GENERADO POR LOS EDITS	006	009
EDIT	68			
V205	( )			
V204	(0-7)			
		GENERADO POR LOS EDITS	006	007
EDIT	69			
V203	( )			
V204	(0-7)			
		GENERADO POR LOS EDITS	003	006
EDIT	70			
V202	( )			
V204	(0-7)			
		GENERADO POR LOS EDITS	001	006
EDIT	71			
V203	(1-6)			
V210	( )			
		GENERADO POR LOS EDITS	004	017

EDIT 72		
V203 (1-6)		
V209 ( )	GENERADO POR LOS EDITS	004 015
EDIT 73		
V203 (1-6)		
V207 ( )	GENERADO POR LOS EDITS	004 011
EDIT 74		
V203 (1-6)		
V206 ( )	GENERADO POR LOS EDITS	004 009
EDIT 75		
V205 ( )		
V203 (1-6)	GENERADO POR LOS EDITS	004 007
EDIT 76		
V203 (1-6)		
V204 ( )	GENERADO POR LOS EDITS	004 005
EDIT 77		
V203 (1-6)		
V202 ( )	GENERADO POR LOS EDITS	001 004
EDIT 78		
V1062 (1)		
V202 (1-2,4-5)	GENERADO POR LOS EDITS	002 087
EDIT 79		
V202 (1-7)		
V210 ( )	GENERADO POR LOS EDITS	002 017
EDIT 80		
V209 ( )		
V202 (1-7)	GENERADO POR LOS EDITS	002 015
EDIT 81		
V202 (1-7)		
V207 ( )	GENERADO POR LOS EDITS	002 011
EDIT 82		
V206 ( )		
V202 (1-7)	GENERADO POR LOS EDITS	002 009
EDIT 83		
V205 ( )		
V202 (1-7)	GENERADO POR LOS EDITS	002 007

EDIT 84		
V202 (1-7)		
V204 ( )		
	GENERADO POR LOS EDITS	002 005
EDIT 85		
V203 ( )		
V202 (1-7)		
	GENERADO POR LOS EDITS	002 003
EDIT 86		
V205 (1)		
V204 (1-3)		
V207 (1)		
	GENERADO POR LOS EDITS	006 015 026 086
EDIT 87		
V201 (1)		
V208 ( )		
	GENERADO POR LOS EDITS	013
EDIT 88		
V201 (2-3)		
V208 (0-8)		
	GENERADO POR LOS EDITS	014
EDIT 89		
V208 (0,2-8)		
V207 (1)		
	GENERADO POR LOS EDITS	025
EDIT 90		
V208 (0,3-8)		
V207 (2)		
	GENERADO POR LOS EDITS	027
EDIT 91		
V209 (1-5)		
V208 (2)		
V207 (2)		
	GENERADO POR LOS EDITS	028
EDIT 92		
V208 (0,4-8)		
V207 (3)		
	GENERADO POR LOS EDITS	030
EDIT 93		
V209 (2-5)		
V208 (2)		
V207 (3)		
	GENERADO POR LOS EDITS	031
EDIT 94		
V209 (1-5)		
V208 (3)		
V207 (3)		
	GENERADO POR LOS EDITS	032
EDIT 95		

V208 (0,5-8)		
V207 (4)	GENERADO POR LOS EDITS	034
EDIT 96		
V209 (3-5)		
V208 (2)		
V207 (4)	GENERADO POR LOS EDITS	035
EDIT 97		
V209 (2-5)		
V208 (3)		
V207 (4)	GENERADO POR LOS EDITS	036
EDIT 98		
V209 (1-5)		
V208 (4)		
V207 (4)	GENERADO POR LOS EDITS	037
EDIT 99		
V208 (0,6-8)		
V207 (5)	GENERADO POR LOS EDITS	039
EDIT 100		
V209 (4-5)		
V208 (2)		
V207 (5)	GENERADO POR LOS EDITS	040
EDIT 101		
V209 (3-5)		
V208 (3)		
V207 (5)	GENERADO POR LOS EDITS	041
EDIT 102		
V209 (2-5)		
V208 (4)		
V207 (5)	GENERADO POR LOS EDITS	042
EDIT 103		
V209 (1-5)		
V208 (5)		
V207 (5)	GENERADO POR LOS EDITS	043
EDIT 104		
V208 (0,7-8)		
V207 (6)	GENERADO POR LOS EDITS	045
EDIT 105		
V209 (5)		
V208 (2)		
V207 (6)		

GENERADO POR LOS EDITS 046

EDIT 106  
V209 (4-5)  
V208 (3)  
V207 (6)

GENERADO POR LOS EDITS 047

EDIT 107  
V209 (3-5)  
V208 (4)  
V207 (6)

GENERADO POR LOS EDITS 048

EDIT 108  
V209 (2-5)  
V208 (5)  
V207 (6)

GENERADO POR LOS EDITS 049

EDIT 109  
V209 (1-5)  
V208 (6)  
V207 (6)

GENERADO POR LOS EDITS 050

EDIT 110  
V208 (0,8)  
V207 (7)

GENERADO POR LOS EDITS 051

EDIT 111  
V209 (5)  
V208 (3)  
V207 (7)

GENERADO POR LOS EDITS 052

EDIT 112  
V209 (4-5)  
V208 (4)  
V207 (7)

GENERADO POR LOS EDITS 053

EDIT 113  
V209 (3-5)  
V208 (5)  
V207 (7)

GENERADO POR LOS EDITS 054

EDIT 114  
V209 (2-5)  
V208 (6)  
V207 (7)

GENERADO POR LOS EDITS 055

EDIT 115  
V209 (1-5)  
V208 (7)  
V207 (7)

GENERADO POR LOS EDITS 056

EDIT 116  
V208 (0)  
V207 (8)  
GENERADO POR LOS EDITS 057

EDIT 117  
V209 (5)  
V208 (4)  
V207 (8)  
GENERADO POR LOS EDITS 058

EDIT 118  
V209 (4-5)  
V208 (5)  
V207 (8)  
GENERADO POR LOS EDITS 059

EDIT 119  
V209 (3-5)  
V208 (6)  
V207 (8)  
GENERADO POR LOS EDITS 060

EDIT 120  
V209 (2-5)  
V208 (7)  
V207 (8)  
GENERADO POR LOS EDITS 061

EDIT 121  
V209 (1-5)  
V208 (8)  
V207 (8)  
GENERADO POR LOS EDITS 062

EDIT 122  
V209 (1-5)  
V208 (0)  
V207 (9)  
GENERADO POR LOS EDITS 063

EDIT 123  
V209 (5)  
V208 (5)  
V207 (9)  
GENERADO POR LOS EDITS 064

EDIT 124  
V209 (4-5)  
V208 (6)  
V207 (9)  
GENERADO POR LOS EDITS 065

EDIT 125  
V209 (3-5)  
V208 (7)  
V207 (9)  
GENERADO POR LOS EDITS 066

EDIT 126

V209 (2-5)		
V208 (8)		
V207 (9)		
	GENERADO POR LOS EDITS	067
EDIT 127		
V209 (2-5)		
V208 (0)		
V207 (10)		
	GENERADO POR LOS EDITS	068
EDIT 128		
V209 (5)		
V208 (6)		
V207 (10)		
	GENERADO POR LOS EDITS	069
EDIT 129		
V209 (4-5)		
V208 (7)		
V207 (10)		
	GENERADO POR LOS EDITS	070
EDIT 130		
V209 (3-5)		
V208 (8)		
V207 (10)		
	GENERADO POR LOS EDITS	071
EDIT 131		
V209 (3-5)		
V208 (0)		
V207 (11)		
	GENERADO POR LOS EDITS	072
EDIT 132		
V209 (5)		
V208 (7)		
V207 (11)		
	GENERADO POR LOS EDITS	073
EDIT 133		
V209 (4-5)		
V208 (8)		
V207 (11)		
	GENERADO POR LOS EDITS	074
EDIT 134		
V209 (4-5)		
V208 (0)		
V207 (12)		
	GENERADO POR LOS EDITS	075
EDIT 135		
V209 (5)		
V208 (8)		
V207 (12)		
	GENERADO POR LOS EDITS	076
EDIT 136		

V209 (5)		
V208 (0)		
V207 (13)	GENERADO POR LOS EDITS	077
EDIT 137		
V208 (0)		
V7100 (1-8)	GENERADO POR LOS EDITS	078
EDIT 138		
V208 (0,2-8)		
V7100 (1)	GENERADO POR LOS EDITS	079
EDIT 139		
V208 (0,3-8)		
V7100 (1-2)	GENERADO POR LOS EDITS	080
EDIT 140		
V208 (0,4-8)		
V7100 (1-3)	GENERADO POR LOS EDITS	081
EDIT 141		
V208 (0,5-8)		
V7100 (1-4)	GENERADO POR LOS EDITS	082
EDIT 142		
V208 (0,6-8)		
V7100 (1-5)	GENERADO POR LOS EDITS	083
EDIT 143		
V208 (0,7-8)		
V7100 (1-6)	GENERADO POR LOS EDITS	084
EDIT 144		
V208 (0,8)		
V7100 (1-7)	GENERADO POR LOS EDITS	085
EDIT 145		
V210 (1-7)		
V208 ( )	GENERADO POR LOS EDITS	013 018
EDIT 146		
V209 (0-5)		
V208 ( )	GENERADO POR LOS EDITS	013 016
EDIT 147		
V208 ( )		
V207 (1-30)	GENERADO POR LOS EDITS	012 013

EDIT 148			
V206 (1-6)			
V208 ( )	GENERADO POR LOS EDITS	010	013
EDIT 149			
V205 (0-2)			
V208 ( )	GENERADO POR LOS EDITS	008	013
EDIT 150			
V204 (0-7)			
V208 ( )	GENERADO POR LOS EDITS	006	013
EDIT 151			
V203 (1-6)			
V208 ( )	GENERADO POR LOS EDITS	004	013
EDIT 152			
V202 (1-7)			
V208 ( )	GENERADO POR LOS EDITS	002	013
EDIT 153			
V210 ( )			
V208 (0-8)	GENERADO POR LOS EDITS	014	017
EDIT 154			
V209 ( )			
V208 (0-8)	GENERADO POR LOS EDITS	014	015
EDIT 155			
V208 (0-8)			
V207 ( )	GENERADO POR LOS EDITS	011	014
EDIT 156			
V206 ( )			
V208 (0-8)	GENERADO POR LOS EDITS	009	014
EDIT 157			
V205 ( )			
V208 (0-8)	GENERADO POR LOS EDITS	007	014
EDIT 158			
V204 ( )			
V208 (0-8)	GENERADO POR LOS EDITS	005	014
EDIT 159			
V203 ( )			
V208 (0-8)	GENERADO POR LOS EDITS	003	014

EDIT 160  
V202 ( )  
V208 (0-8)  
GENERADO POR LOS EDITS 001 014

EDIT 161  
V205 (1)  
V204 (1-3)  
V208 (0)  
V207 (9)  
GENERADO POR LOS EDITS 014 015 063 086

EDIT 162  
V205 (1)  
V204 (1-3)  
V208 (8)  
V207 (8)  
GENERADO POR LOS EDITS 014 015 062 086

EDIT 163  
V205 (1)  
V204 (1-3)  
V208 (7)  
V207 (7)  
GENERADO POR LOS EDITS 014 015 056 086

EDIT 164  
V205 (1)  
V204 (1-3)  
V208 (6)  
V207 (6)  
GENERADO POR LOS EDITS 014 015 050 086

EDIT 165  
V205 (1)  
V204 (1-3)  
V208 (5)  
V207 (5)  
GENERADO POR LOS EDITS 014 015 043 086

EDIT 166  
V205 (1)  
V204 (1-3)  
V208 (4)  
V207 (4)  
GENERADO POR LOS EDITS 014 015 037 086

EDIT 167  
V205 (1)  
V204 (1-3)  
V208 (3)  
V207 (3)  
GENERADO POR LOS EDITS 014 015 032 086

EDIT 168  
V205 (1)  
V204 (1-3)  
V208 (2)  
V207 (2)  
GENERADO POR LOS EDITS 014 015 028 086

**Anexo 5.2 - Aplicação de Chefe**



DICCIONARIO DIA CREADO

-----

VARIABLE NUMERO	NOMBRE	POSICION	LONGITUD	NUMERO DE CODIGOS	CODIGOS
1	V305	33	1	3	1 3,
2	V3080	46	1	4	8,0,1,9,
3	V306	36	1	10	8,0 8,
4	V307	37	1	10	8,0-8,
5	V300	13	2	1	01,
6	V302	21	2	2	01,20,
7	V3043	28	3	142	001 130,900-911,

-----  
- NUMERO DE VARIABLES A DEPURAR : 7  
NUMERO TOTAL DE CODIGOS PARA TODAS LAS VARIABLES (INCLUIDOS LOS INVALIDOS) : 183  
-----

..... FIN DEL LISTADO DEL DICCIONARIO -DIA-.....

2.- RID-ORIGINALES

=====

R I D 1  
V305 (1-2)  
V302 (20)  
V3043 (1-4,900-911)  
V305 (3)

R I D 2  
V305 (3)  
V302 (20)  
V3043 (5)  
V305 (1-2)

R I D 3  
V306 (1)  
V307 (2)  
V3043 (6-24)  
V307 (4)

R I D 4  
V306 (2)  
V307 (2)  
V3043 (7-25)  
V307 (4)

R I D 5  
V306 (3)  
V307 (2)  
V3043 (8-26)  
V307 (4)

R I D 6  
V306 (4)  
V307 (2)  
V3043 (9-27)  
V307 (4)

R I D 7  
V306 (1)  
V307 (3)  
V3043 (10-28)  
V306 (5)  
V307 (4)

R I D 8  
V306 (2)  
V307 (3)  
V3043 (11-29)  
V306 (6)  
V307 (4)

R I D 9  
V306 (3)  
V307 (3)  
V3043 (12-30)  
V306 (7)  
V307 (4)

3.- EDITS DERIVADOS DE RIDS (-EDR-)

=====

EDIT 1  
V305 (1-2)  
V302 (20)  
V3043 (1-4,900-911)

EDIT 2  
V305 (3)  
V302 (20)  
V3043 (5)

EDIT 3  
V306 (1)  
V307 (2)  
V3043 (6 24)

EDIT 4  
V306 (2)  
V307 (2)  
V3043 (7 25)

EDIT 5  
V306 (3)  
V307 (2)  
V3043 (8-26)

EDIT 6  
V306 (4)  
V307 (2)  
V3043 (9-27)

EDIT 7  
V306 (1)  
V307 (3)  
V3043 (10-28)

EDIT 8  
V306 (2)  
V307 (3)  
V3043 (11-29)

EDIT 9  
V306 (3)  
V307 (3)  
V3043 (12-30)

EDIT 10  
V306 (4)  
V307 (3)  
V3043 (13-31)

EDIT 11  
V306 (1)  
V307 (6)  
V3043 (14-32)

EDIT 12

V306 (2)  
V307 (6)  
V3043 (15-33)

EDIT 13  
V306 (3)  
V307 (6)  
V3043 (16-34)

EDIT 14  
V306 (4)  
V307 (6)  
V3043 (17-35)

4.- RIDS RESULTANTES DEL ANALIZADOR-RID

=====

R I D       1  
V305   (1-2)  
V302   (20)  
V3043   (1-4,900-911)  
          V305   (3)

R I D       2  
V305   (3)  
V302   (20)  
V3043   (5)  
          V305   (1-2)

R I D       3  
V306   (1)  
V307   (2)  
V3043   (6-24)  
          V307   (4)

R I D       4  
V306   (2)  
V307   (2)  
V3043   (7-25)  
          V307   (4)

R I D       5  
V306   (3)  
V307   (2)  
V3043   (8-26)  
          V307   (4)

R I D       6  
V306   (4)  
V307   (2)  
V3043   (9-27)  
          V307   (4)

R I D       7  
V306   (1)  
V307   (3)  
V3043   (10-28)  
          V306   (5)  
          V307   (4)

R I D       8  
V306   (2)  
V307   (3)  
V3043   (11-29)  
          V306   (6)  
          V307   (4)

R I D       9  
V306   (3)  
V307   (3)  
V3043   (12-30)  
          V306   (7)  
          V307   (4)

R I D 10  
V306 (4)  
V307 (3)  
V3043 (13-31)  
V306 (8)  
V307 (4)

R I D 11  
V306 (1)  
V307 (6)  
V3043 (14-32)  
V307 (5)

R I D 12  
V306 (2)  
V307 (6)  
V3043 (15-33)  
V307 (5)

R I D 13  
V306 (3)  
V307 (6)  
V3043 (16-34)  
V307 (5)

R I D 14  
V306 (4)  
V307 (6)  
V3043 (17-35)  
V307 (5)

5.- REGLAS DE INCOMPATIBILIDAD

=====

-SON LOS EDITS ESPECIFICADOS POR EL USUARIO JUNTO CON LOS -EDR-, Y DESPUES DE ELIMINAR REDUNDANCIAS.  
-ESTAS REGLAS SON LA ENTRADA PARA LA GENERACION DEL CONJUNTO COMPLETO, Y TAMBIEN SE USAN EN LA DETECCION DE ERRORES.

EDIT 1  
V305 (1-2)  
V306 ( )

EDIT 2  
V305 (3)  
V306 (0-8)

EDIT 3  
V305 (1-2)  
V307 ( )

EDIT 4  
V305 (3)  
V307 (0-8)

EDIT 5  
V305 (1-2)  
V3080 ( )

EDIT 6  
V305 (3)  
V3080 (0-1,9)

EDIT 7  
V302 (1)  
V3043 (1-9,900-911)

EDIT 8  
V305 (1-2)  
V3043 (1-4,900-911)

EDIT 9  
V305 (3)  
V3043 (5-130)

EDIT 10  
V306 (1)  
V307 (2)  
V3043 (1-24,900-911)

EDIT 11  
V306 (2)  
V307 (2)  
V3043 (1-25,900-911)

EDIT 12  
V306 (3)  
V307 (2)  
V3043 (1-26,900-911)

EDIT 13

V306 (4)  
V307 (2)  
V3043 (1 27,900-911)

EDIT 14  
V306 (5)  
V307 (2)  
V3043 (1-28,900-911)

EDIT 15  
V306 (6)  
V307 (2)  
V3043 (1-29,900-911)

EDIT 16  
V306 (1)  
V307 (3)  
V3043 (1-28,900-911)

EDIT 17  
V306 (2)  
V307 (3)  
V3043 (1-29,900-911)

EDIT 18  
V306 (3)  
V307 (3)  
V3043 (1-30,900-911)

EDIT 19  
V306 (4)  
V307 (3)  
V3043 (1-31,900-911)

EDIT 20  
V306 (5)  
V307 (3)  
V3043 (1-65,900-911)

EDIT 21  
V306 (1)  
V307 (4)  
V3043 (1-5,900-911)

EDIT 22  
V306 (2)  
V307 (4)  
V3043 (1-6,900-911)

EDIT 23  
V306 (3)  
V307 (4)  
V3043 (1-7,900-911)

EDIT 24  
V306 (4)  
V307 (4)  
V3043 (1-8,900-911)

EDIT 25

V306 (5)  
V307 (4)  
V3043 (1-9,900-911)

EDIT 26  
V306 (6)  
V307 (4)  
V3043 (1-10,900-911)

EDIT 27  
V306 (7)  
V307 (4)  
V3043 (1-11,900-911)

EDIT 28  
V306 (8)  
V307 (4)  
V3043 (1-12,900-911)

EDIT 29  
V306 (1)  
V307 (5)  
V3043 (1-13,900-911)

EDIT 30  
V306 (2)  
V307 (5)  
V3043 (1-14,900-911)

EDIT 31  
V306 (3)  
V307 (5)  
V3043 (1-15,900-911)

EDIT 32  
V306 (4)  
V307 (5)  
V3043 (1-16,900-911)

EDIT 33  
V306 (1)  
V307 (6)  
V3043 (1-32,900-911)

EDIT 34  
V306 (2)  
V307 (6)  
V3043 (1-33,900-911)

EDIT 35  
V306 (3)  
V307 (6)  
V3043 (1-34,900-911)

EDIT 36  
V306 (4)  
V307 (6)  
V3043 (1-35,900-911)

EDIT 37

V306 (1)  
V307 (7)  
V3043 (1-16,900-911)

EDIT 38  
V306 (2)  
V307 (7)  
V3043 (1-17,900-911)

EDIT 39  
V306 (3)  
V307 (7)  
V3043 (1-18,900-911)

EDIT 40  
V306 (4)  
V307 (7)  
V3043 (1-19,900-911)

EDIT 41  
V306 (5)  
V307 (7)  
V3043 (1-20,900-911)

EDIT 42  
V306 (6)  
V307 (7)  
V3043 (1-21,900-911)

EDIT 43  
V307 (1)  
V3043 (1-13,900-911)

EDIT 44  
V307 (8)  
V3043 (1-20,900-911)

EDIT 45  
V305 (2)  
V306 (3-8)

EDIT 46  
V305 (2)  
V307 (3.5 8)

EDIT 47  
V306 (1-8)  
V307 (0-1,8)

EDIT 48  
V306 (0,6-8)  
V307 (3)

EDIT 49  
V306 (0)  
V307 (4)

EDIT 50  
V306 (0,5-8)  
V307 (5-6)

EDIT 51  
V306 (0.7-8)  
V307 (2.7)

6.- CONJUNTO COMPLETO DE REGLAS

=====

EDIT	1		
V305	(1-2)		
V306	( )		
		GENERADO POR LOS EDITS	001
EDIT	2		
V305	(3)		
V306	(0-8)		
		GENERADO POR LOS EDITS	002
EDIT	3		
V305	(1-2)		
V307	( )		
		GENERADO POR LOS EDITS	003
EDIT	4		
V305	(3)		
V307	(0-8)		
		GENERADO POR LOS EDITS	004
EDIT	5		
V305	(1-2)		
V3080	( )		
		GENERADO POR LOS EDITS	005
EDIT	6		
V305	(3)		
V3080	(0-1,9)		
		GENERADO POR LOS EDITS	006
EDIT	7		
V305	(2)		
V306	(3-8)		
		GENERADO POR LOS EDITS	045
EDIT	8		
V305	(2)		
V307	(3,5-8)		
		GENERADO POR LOS EDITS	046
EDIT	9		
V306	(1-8)		
V307	(0-1,8)		
		GENERADO POR LOS EDITS	047
EDIT	10		
V306	(0,6-8)		
V307	(3)		
		GENERADO POR LOS EDITS	048
EDIT	11		
V306	(0)		
V307	(4)		
		GENERADO POR LOS EDITS	049
EDIT	12		

V306 (0,5-8)			
V307 (5-6)			
	GENERADO POR LOS EDITS	050	
EDIT 13			
V306 (0,7-8)			
V307 (2,7)			
	GENERADO POR LOS EDITS	051	
EDIT 14			
V308( )			
V307 (0 8)			
	GENERADO POR LOS EDITS	004 005	
EDIT 15			
V308( )			
V306 (0-8)			
	GENERADO POR LOS EDITS	002 005	
EDIT 16			
V308( (0-1,9)			
V307 ( )			
	GENERADO POR LOS EDITS	003 006	
EDIT 17			
V306 (0-8)			
V307 ( )			
	GENERADO POR LOS EDITS	002 003	
EDIT 18			
V308( (0-1,9)			
V306 ( )			
	GENERADO POR LOS EDITS	001 006	
EDIT 19			
V306 ( )			
V307 (0-8)			
	GENERADO POR LOS EDITS	001 004	
EDIT 20			
V302 (1)			
V3043 (1-9,900-911)			
	GENERADO POR LOS EDITS	007	
EDIT 21			
V305 (1-2)			
V3043 (1-4,900-911)			
	GENERADO POR LOS EDITS	008	
EDIT 22			
V305 (3)			
V3043 (5 130)			
	GENERADO POR LOS EDITS	009	
EDIT 23			
V306 (2)			
V307 (2)			
V3043 (1-25,900-911)			
	GENERADO POR LOS EDITS	011	

EDIT 24  
V306 (3)  
V307 (2)  
V3043 (1-26,900-911)  
GENERADO POR LOS EDITS 012

EDIT 25  
V306 (4)  
V307 (2)  
V3043 (1-27,900-911)  
GENERADO POR LOS EDITS 013

EDIT 26  
V306 (5)  
V307 (2)  
V3043 (1-28,900-911)  
GENERADO POR LOS EDITS 014

EDIT 27  
V306 (6)  
V307 (2)  
V3043 (1-29,900-911)  
GENERADO POR LOS EDITS 015

EDIT 28  
V306 (2)  
V307 (3)  
V3043 (1-29,900-911)  
GENERADO POR LOS EDITS 017

EDIT 29  
V306 (3)  
V307 (3)  
V3043 (1-30,900-911)  
GENERADO POR LOS EDITS 018

EDIT 30  
V306 (4)  
V307 (3)  
V3043 (1-31,900-911)  
GENERADO POR LOS EDITS 019

EDIT 31  
V306 (5)  
V307 (3)  
V3043 (1-65,900-911)  
GENERADO POR LOS EDITS 020

EDIT 32  
V306 (2)  
V307 (5)  
V3043 (1-14,900-911)  
GENERADO POR LOS EDITS 030

EDIT 33  
V306 (3)  
V307 (5)  
V3043 (1-15,900-911)  
GENERADO POR LOS EDITS 031

FDIT 34  
V306 (4)  
V307 (5)  
V3043 (1-16,900-911)  
GENERADO POR LOS EDITS 032

EDIT 35  
V306 (2)  
V307 (6)  
V3043 (1-33,900-911)  
GENERADO POR LOS EDITS 034

EDIT 36  
V306 (3)  
V307 (6)  
V3043 (1 34,900-911)  
GENERADO POR LOS EDITS 035

EDIT 37  
V306 (4)  
V307 (6)  
V3043 (1-35,900-911)  
GENERADO POR LOS EDITS 036

EDIT 38  
V306 (2)  
V307 (7)  
V3043 (1-17,900-911)  
GENERADO POR LOS EDITS 038

EDIT 39  
V306 (3)  
V307 (7)  
V3043 (1-18,900-911)  
GENERADO POR LOS EDITS 039

EDIT 40  
V306 (4)  
V307 (7)  
V3043 (1-19,900-911)  
GENERADO POR LOS EDITS 040

EDIT 41  
V306 (5)  
V307 (7)  
V3043 (1-20,900-911)  
GENERADO POR LOS EDITS 041

EDIT 42  
V306 (6)  
V307 (7)  
V3043 (1-21,900-911)  
GENERADO POR LOS EDITS 042

EDIT 43  
V307 (1,5)  
V3043 (1-13,900-911)  
GENERADO POR LOS EDITS 001 004 029 030 031 032 043 050

FDIT 44

V307 (8)  
V3043 (1-20,900-911)  
GENERADO POR LOS EDITS 044

EDIT 45  
V3080 (0-1,9)  
V3043 (1-4,900-911)  
GENERADO POR LOS EDITS 006 008

EDIT 46  
V307 (0-8)  
V3043 (1-4,900-911)  
GENERADO POR LOS EDITS 004 008

EDIT 47  
V306 (0-8)  
V3043 (1-4,900-911)  
GENERADO POR LOS EDITS 002 008

EDIT 48  
V3080 ( )  
V3043 (5-130)  
GENERADO POR LOS EDITS 005 009

EDIT 49  
V307 ( )  
V3043 (5-130)  
GENERADO POR LOS EDITS 003 009

EDIT 50  
V306 ( )  
V3043 (5-130)  
GENERADO POR LOS EDITS 001 009

EDIT 51  
V307 (6)  
V3043 (1-32,900-911)  
GENERADO POR LOS EDITS 001 004 033 034 035 036 050

EDIT 52  
V307 (3)  
V3043 (1-28,900-911)  
GENERADO POR LOS EDITS 001 004 016 017 018 019 020 048

EDIT 53  
V307 (7)  
V3043 (1-16,900-911)  
GENERADO POR LOS EDITS 001 004 037 038 039 040 041 042 051

EDIT 54  
V307 (2)  
V3043 (1-24,900-911)  
GENERADO POR LOS EDITS 001 004 010 011 012 013 014 015 051

EDIT 55  
V307 (2,4-7)  
V3043 (1-5,900-911)  
GENERADO POR LOS EDITS 001 002 003 004 010 011 012 013 014 015 016 017 018 019 020 021 022 023 024 025 026 027  
028 029 030 031 032 033 034 035 036 037 038 039 040 041 042 047 048 049 050 051



EDIT 67  
V305 (3)  
V302 (1)

GENERADO POR LOS EDITS 007 009

**Anexo 5.3 - Aplicação de Não-Chefe**



DICCIONARIO-DIA CREADO  
-----

VARIABLE NUMERO	NOMBRE	POSICION	LONGITUD	NUMERO DE CODIGOS	CODIGOS
1	V306	36	1	1	B,
2	V307	37	1	1	B,
3	V3080	46	1	1	B,
4	V305	33	1	3	1-3,
5	V3044	31	2	11	1-11,
6	V302	21	2	15	02-16,
7	V300	13	2	98	02-99,
8	V3043	28	3	142	001-130,900-911,

-----  
- NUMERO DE VARIABLES A DEPURAR : 8  
- NUMERO TOTAL DE CODIGOS PARA TODAS LAS VARIABLES (INCLUIDOS LOS INVALIDOS) : 285  
-----

..... FIN DEL LISTADO DEL DICCIONARIO -DIA-.....



2.- RID-ORIGINALES  
=====

R I D        1  
V305 (1-2)  
V302 (3-4, 8, 10-14, 16)  
V3043 (1-4, 900-911)  
      V305 (3)

R I D        2  
V305 (3)  
V3043 (5)  
      V305 (1-2)



3.- EDITS DERIVADOS DE RIDS (-EDR-)

=====

EDIT 1  
V305 (1-2)  
V302 (3-4,8,10-14,16)  
V3043 (1-4,900-911)

EDIT 2  
V305 (3)  
V3043 (5)



4.- RIDS RESULTANTES DEL ANALIZADOR-RID  
=====

R I D        1  
V305    (1-2)  
V302    (3-4,8,10-14,16)  
V3043   (1-4,900-911)  
         V305    (3)

R I D        2  
V305    (3)  
V3043   (5)  
         V305    (1-2)



5.- REGLAS DE INCOMPATIBILIDAD  
-----

-SON LOS EDITS ESPECIFICADOS POR EL USUARIO JUNTO CON LOS -EDR-, Y DESPUES DE ELIMINAR REDUNDANCIAS.  
-ESTAS REGLAS SON LA ENTRADA PARA LA GENERACION DEL CONJUNTO COMPLETO, Y TAMBIEN SE USAN EN LA DETECCION DE ERRORES.

EDIT 1  
V302 (2,9,15)  
V3043 (1-9,900-911)

EDIT 2  
V302 (5-6)  
V3043 (1-19,900-911)

EDIT 3  
V302 (7)  
V3043 (1-29,900-911)

EDIT 4  
V305 (1-2)  
V3043 (1-4,900-911)

EDIT 5  
V305 (3)  
V3043 (5-130)



6.- CONJUNTO COMPLETO DE REGLAS

=====

EDIT 1  
V302 (7)  
V3043 (1-29,900-911)  
GENERADO POR LOS EDITS 003

EDIT 2  
V305 (1-2)  
V3043 (1-4,900-911)  
GENERADO POR LOS EDITS 004

EDIT 3  
V305 (3)  
V3043 (5-130)  
GENERADO POR LOS EDITS 005

EDIT 4  
V305 (3)  
V302 (2,5-7,9,15)  
GENERADO POR LOS EDITS 001 002 003 005

EDIT 5  
V302 (2,9,15)  
V3043 (1-9,900-911)  
GENERADO POR LOS EDITS 001

EDIT 6  
V302 (5-6)  
V3043 (1-19,900-911)  
GENERADO POR LOS EDITS 002

**Anexo 6 - Exemplo das Tabelas Especiais para a Análise da  
Correção Automática**



UF :PR

LOTE :009

DATA : 27/10/93

TABELA 1 : REGISTROS "BONS" E REGISTROS "MAUS" SEGUNDO O NUMERO DE REGRAS QUE FALHARAM

DISCRIMINAÇÃO	A P L I C A Ç Ã O					
	D O M I C I L I O S		C H E F E S		N Ã O C H E F E S	
	ABSOLUTO	RELATIVO	ABSOLUTO	RELATIVO	ABSOLUTO	RELATIVO
TOTAL DE REGISTROS	69 678	100,00	69 678	100,00	193 120	100,00
REGISTROS "BONS"	65 031	93,33	66 353	95,23	191 601	99,21
REGISTROS "MAUS"	4 647	6,67	3 325	4,77	1 519	0,79
TOTAL DE "MAUS"	4 647	100,00	3 325	100,00	1 519	100,00
NUMERO DE FALHAS						
ZERO REGRAS	5	0,11	70	2,11	136	8,95
UMA REGRA	4 238	91,20	3 174	95,46	1 375	90,52
DUAS REGRAS	287	6,18	72	2,17	8	0,53
TRES REGRAS	34	0,73	4	0,12	-	-
QUATRO REGRAS	9	0,19	4	0,12	-	-
CINCO REGRAS	19	0,41	1	0,03	-	-
SEIS REGRAS	2	0,04	0	0,00	-	-
SETE REGRAS	1	0,02	0	0,00	-	-
OITO REGRAS	2	0,04	0	0,00	-	-
NOVE REGRAS	48	1,03	0	0,00	-	-
DEZ REGRAS	2	0,04	0	0,00	-	-
11 A 15 REGRAS	0	0,00	0	0,00	-	-
16 A 20 REGRAS	0	0,00	0	0,00	-	-
21 A 25 REGRAS	0	0,00	0	0,00	-	-
26 A 30 REGRAS	0	0,00	0	0,00	-	-
31 A 35 REGRAS	0	0,00	0	0,00	-	-
36 A 40 REGRAS	0	0,00	0	0,00	-	-
41 A 45 REGRAS	0	0,00	0	0,00	-	-
46 A 50 REGRAS	0	0,00	0	0,00	-	-
51 A 60 REGRAS	0	0,00	0	0,00	-	-
61 A 70 REGRAS	0	0,00	-	-	-	-
MAIS DE 70 REGRAS	0	0,00	-	-	-	-

TABELA 2 : NUMERO DE REGISTROS QUE FALHARAM EM CADA REGRA SEGUNDO O NUMERO DA REGRA/RID QUE FALHOU

NUMERO DA REGRA/RID QUE FALHOU	NUMERO DE REGISTROS "MAUS"					
	DOMICILIOS		CHEFES		NAO CHEFES	
	ABSOLUTO	RELATIVO	ABSOLUTO	RELATIVO	ABSOLUTO	RELATIVO
TOTAL DE REGISTROS MAUS	4 647	-	3 325	-	1 519	-
01	68	1,46	111	3,34	20	1,32
02	26	0,56	5	0,15	13	0,86
03	116	2,50	96	2,89	8	0,53
04	26	0,56	5	0,15	945	62,21
05	122	2,63	0	0,00	405	26,66
06	27	0,58	5	0,15	-	-
07	161	3,46	9	0,27	-	-
08	27	0,58	10	0,30	-	-
09	108	2,32	5	0,15	-	-
10	27	0,58	46	1,38	-	-
11	69	1,48	173	5,20	-	-
12	27	0,58	285	8,57	-	-
13	112	2,41	897	26,98	-	-
14	27	0,58	35	1,05	-	-
15	110	2,37	23	0,69	-	-
16	27	0,58	63	1,89	-	-
17	144	3,10	53	1,59	-	-
18	27	0,58	67	2,02	-	-
19	409	8,80	206	6,20	-	-
20	239	5,14	157	4,72	-	-
21	128	2,75	0	0,00	-	-
22	149	3,21	0	0,00	-	-
23	2 427	52,23	0	0,00	-	-
24	122	2,63	0	0,00	-	-
25	2	0,04	0	0,00	-	-
26	98	2,11	0	0,00	-	-
27	3	0,06	0	0,00	-	-
28	26	0,56	2	0,06	-	-
29	3	0,06	0	0,00	-	-
30	0	0,00	0	0,00	-	-
31	1	0,02	1	0,03	-	-
32	3	0,06	0	0,00	-	-
33	0	0,00	12	0,36	-	-



SEPLAN/IBGE

CENSO DEMOGRAFICO DE 1991

ICCA0084

UF : PR                    LOTE : 009

DATA : 27/10/93

TABELA 2 : NUMERO DE REGISTROS QUE FALHARAM EM CADA REGRA SEGUNDO O NUMERO DA REGRA RID QUE FALHOU

NUMERO DA REGRA/RID QUE FALHOU	NUMERO DE REGISTROS "MAUS"					
	DOMICILIOS		CHEFES		NÃO CHEFES	
	ABSOLUTO	RELATIVO	ABSOLUTO	RELATIVO	ABSOLUTO	RELATIVO
34	0	0,00	20	0,60	-	-
35	0	0,00	158	4,75	-	-
36	1	0,02	8	0,24	-	-
37	2	0,04	0	0,00	-	-
38	0	0,00	0	0,00	-	-
39	0	0,00	0	0,00	-	-
40	0	0,00	0	0,00	-	-
41	4	0,03	0	0,00	-	-
42	0	0,00	0	0,00	-	-
43	2	0,04	0	0,00	-	-
44	0	0,00	0	0,00	-	-
45	0	0,00	135	4,06	-	-
46	0	0,00	10	0,30	-	-
47	0	0,00	240	7,22	-	-
48	1	0,02	369	11,10	-	-
49	0	0,00	3	0,09	-	-
50	0	0,00	36	1,08	-	-
51	0	0,00	106	3,19	-	-
52	0	0,00	-	-	-	-
53	0	0,00	-	-	-	-
54	0	0,00	-	-	-	-
55	0	0,00	-	-	-	-
56	0	0,00	-	-	-	-
57	0	0,00	-	-	-	-
58	0	0,00	-	-	-	-
59	0	0,00	-	-	-	-
60	0	0,00	-	-	-	-
61	0	0,00	-	-	-	-
62	0	0,00	-	-	-	-
63	0	0,00	-	-	-	-
64	0	0,00	-	-	-	-
65	0	0,00	-	-	-	-
66	0	0,00	-	-	-	-
67	0	0,00	-	-	-	-
68	0	0,00	-	-	-	-
69	0	0,00	-	-	-	-
70	0	0,00	-	-	-	-



TABELA 2 : NUMERO DE REGISTROS QUE FALHARAM EM CADA REGRA SEGUNDO O NUMERO DA REGRA/RID QUE FALHOU

NUMERO DA REGRA/RID QUE FALHOU	NUMERO DE REGISTROS "MAUS"					
	DOMICILIOS		CHEFES		NÃO CHEFES	
	ABSOLUTO	RELATIVO	ABSOLUTO	RELATIVO	ABSOLUTO	RELATIVO
71	0	0,00	-	-	-	-
72	1	0,02	-	-	-	-
73	0	0,00	-	-	-	-
74	0	0,00	-	-	-	-
75	0	0,00	-	-	-	-
76	0	0,00	-	-	-	-
77	0	0,00	-	-	-	-
78	2	0,04	-	-	-	-
79	166	3,57	-	-	-	-
80	221	4,76	-	-	-	-
81	59	1,27	-	-	-	-
82	15	0,32	-	-	-	-
83	4	0,09	-	-	-	-
84	3	0,06	-	-	-	-
85	2	0,04	-	-	-	-
86	188	4,05	-	-	-	-
R I D S						
01	4	0,09	4	0,12	938	61,75
02	0	0,00	0	0,00	233	15,34
03	-	-	46	1,38	-	-
04	-	-	172	5,17	-	-
05	-	-	284	8,54	-	-
06	-	-	895	26,92	-	-
07	-	-	63	1,89	-	-
08	-	-	53	1,59	-	-
09	-	-	67	2,02	-	-
10	-	-	206	6,20	-	-
11	-	-	12	0,36	-	-
12	-	-	20	0,60	-	-
13	-	-	158	4,75	-	-
14	-	-	8	0,24	-	-



SEPLAN-IBGE

CENSO DEMOGRAFICO DE 1991

ICCA0086

UF : PR            LOTE : 009

DATA : 27/10/93

TABELA 3: NUMERO DE REGISTROS "MAUS" SEGUNDO O NUMERO DE VARIAVEIS IMPUTADAS

NUMERO DE VARIAVEIS IMPUTADAS	NUMERO DE REGISTROS "MAUS"					
	DOMICILIOS		CHEFES		NAO CHEFES	
	ABSOLUTO	RELATIVO	ABSOLUTO	RELATIVO	ABSOLUTO	RELATIVO
TOTAL DE REGISTROS MAUS	4 647	100,00	3 325	100,00	1 519	100,00
UMA	4 459	95,95	2 858	85,95	1 479	97,37
DUAS	88	1,89	453	13,62	40	2,63
TRES	20	0,43	5	0,15	-	-
QUATRO	8	0,17	9	0,27	-	-
CINCO	17	0,37	0	0,00	-	-
SEIS	2	0,04	-	-	-	-
SETE	2	0,04	-	-	-	-
OITO	1	0,02	-	-	-	-
NOVE	50	1,08	-	-	-	-



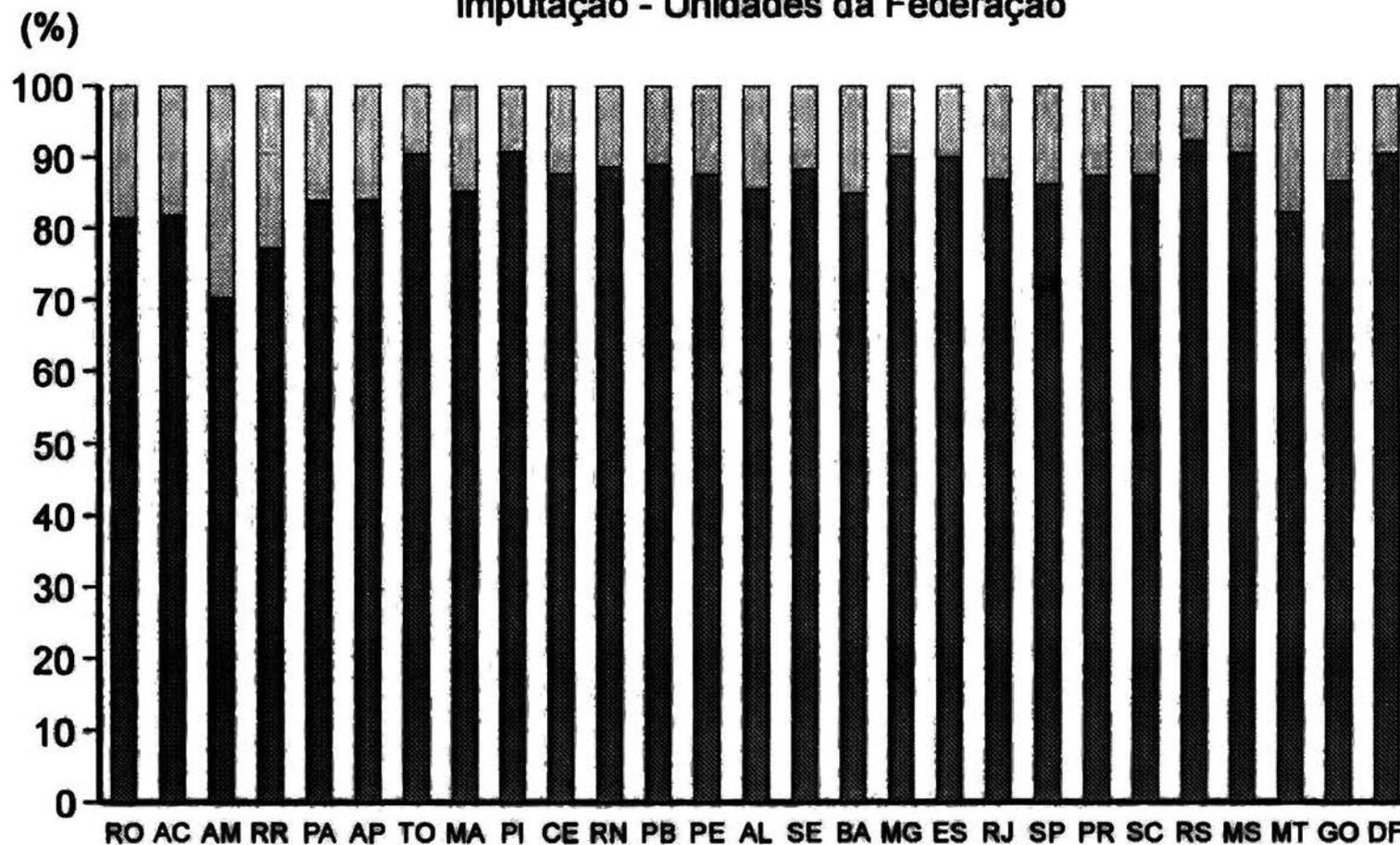
TABELA 4 : NUMERO DE QUESTIONARIOS , DOMICILIOS E PESSOAS SEGUNDO A OCORRENCIA DE IMPUTAÇÕES

DISCRIMINAÇÃO	NUMERO DE QUESTIONARIOS		NUMERO DE DOMICILIOS		NUMERO DE PESSOAS			
					CHEFES		NÃO CHEFES	
	ABSOLUTO	RELATIVO	ABSOLUTO	RELATIVO	ABSOLUTO	RELATIVO	ABSOLUTO	RELATIVO
TOTAL	69 682	100,00	69 682	100,00	69 682	100,00	193 140	100,00
CORRETOS ( SEM IMPUTAÇÃO )	60 737	87,16	65 031	93,33	66 353	95,22	191 601	99,20
COM ALGUMA IMPUTAÇÃO	8 945	12,84	4 651	6,67	3 329	4,78	1 539	0,80
FALTOSOS	4	0,01	4	0,01	4	0,01	20	0,01
IMPUTAÇÃO PELO "D.I.A"	8 941	12,83	4 647	6,67	3 325	4,77	1 519	0,79
TOTAL PELO "D.I.A"	8 941	100,00	-	-	-	-	-	-
SO EM DOMICILIOS	4 288	47,96	-	-	-	-	-	-
SO EM CHEFES	2 947	32,96	-	-	-	-	-	-
SO EM NÃO CHEFES	1 253	14,01	-	-	-	-	-	-
DOMICILIOS E CHEFES	272	3,04	-	-	-	-	-	-
DOMICILIOS E NÃO CHEFES	75	0,84	-	-	-	-	-	-
CHEFES E NÃO CHEFES	94	1,05	-	-	-	-	-	-
DOMIC.,CHEFES E NÃO CHEFES	12	0,13	-	-	-	-	-	-

**Anexo 7 - Gráficos sobre a Correção Automática do CD 1.01 por  
Unidades da Federação**

Gráfico A.7.1

Distribuição de questionários por ocorrência de imputação - Unidades da Federação

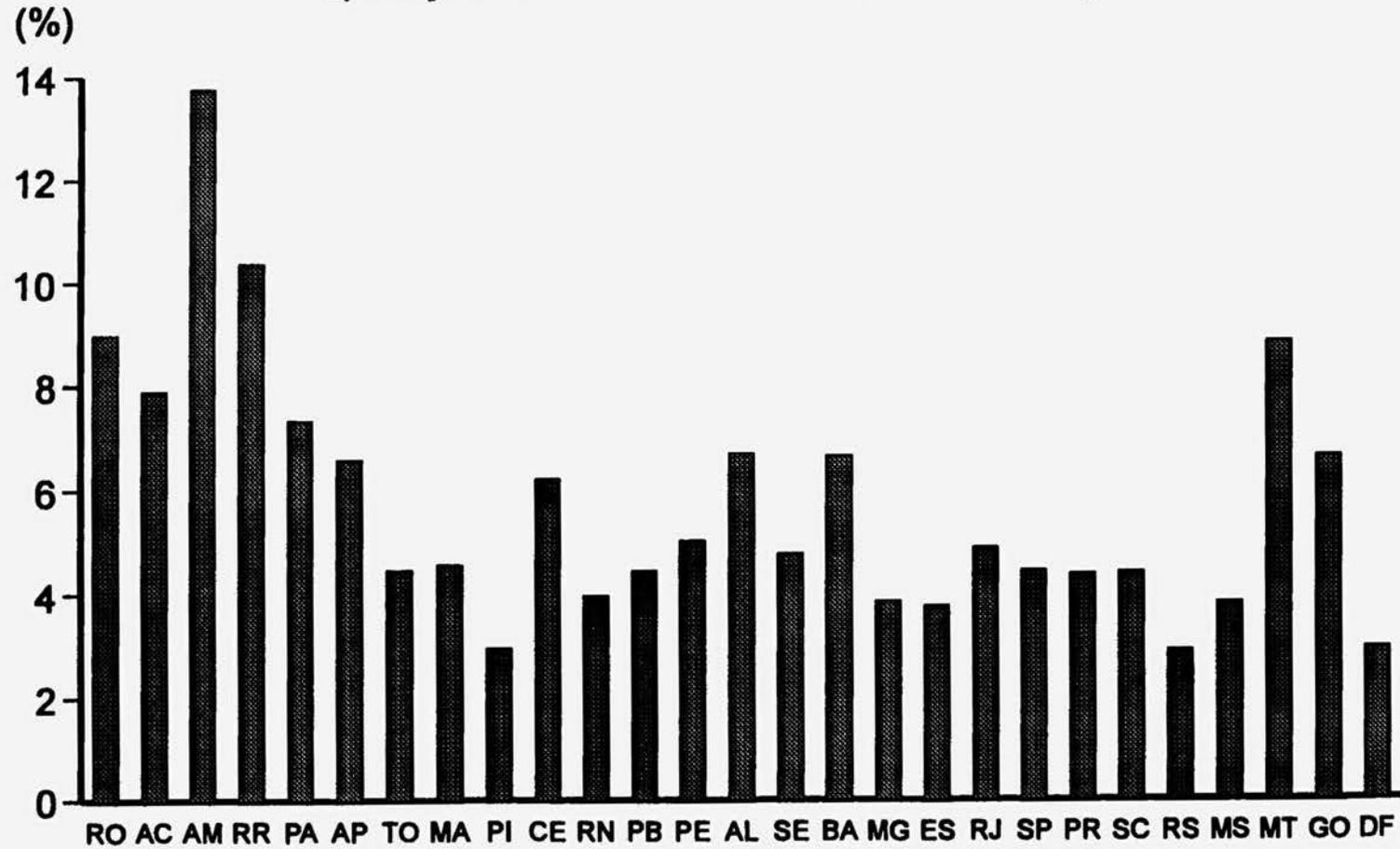


Legenda: Ocorrência

■ Sem imputação ■ Com alguma imputação

Gráfico A.7.2

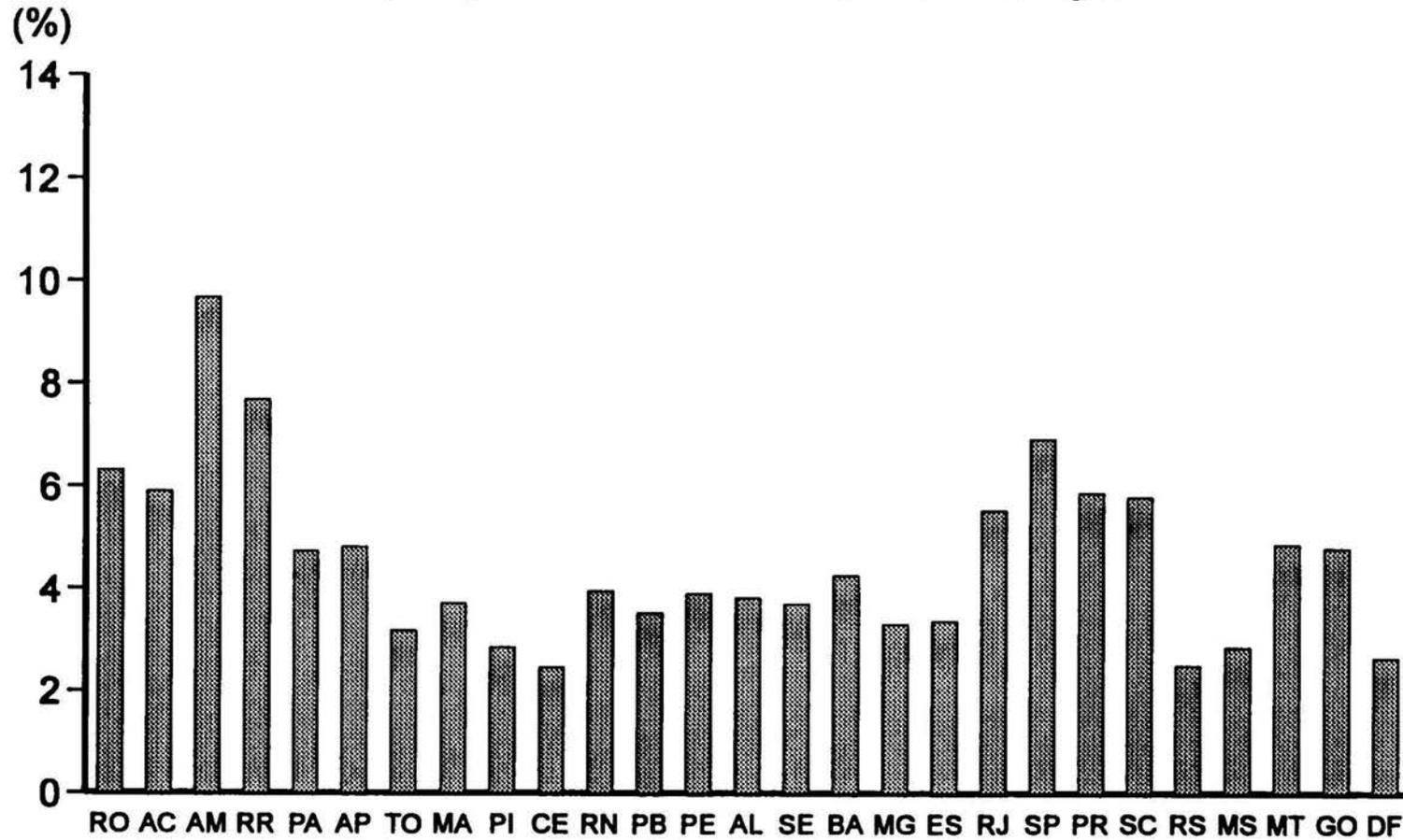
Percentual de registros com alguma imputação na aplicação de Domicílio - Unidades da Federação



Fonte: Censo 91 - Resultados da Correção Automática do CD 1.01, usando o sistema DIA.

Gráfico A.7.3

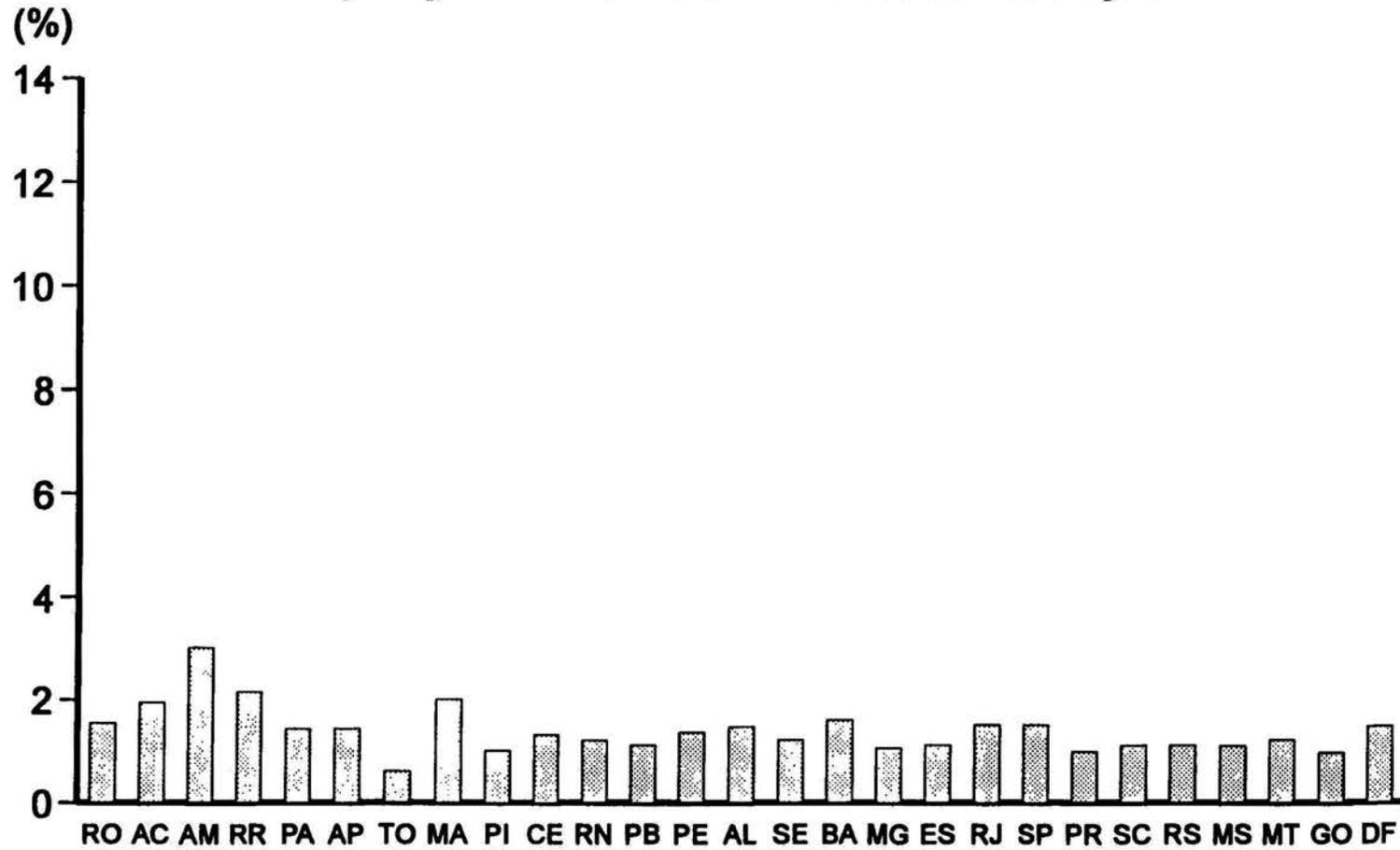
Percentual de registros com alguma imputação na aplicação de Chefe - Unidades da Federação



Fonte: Censo 91 - Resultados da Correção Automática do CD 1.01, usando o sistema DIA.

**Gráfico A.7.4**

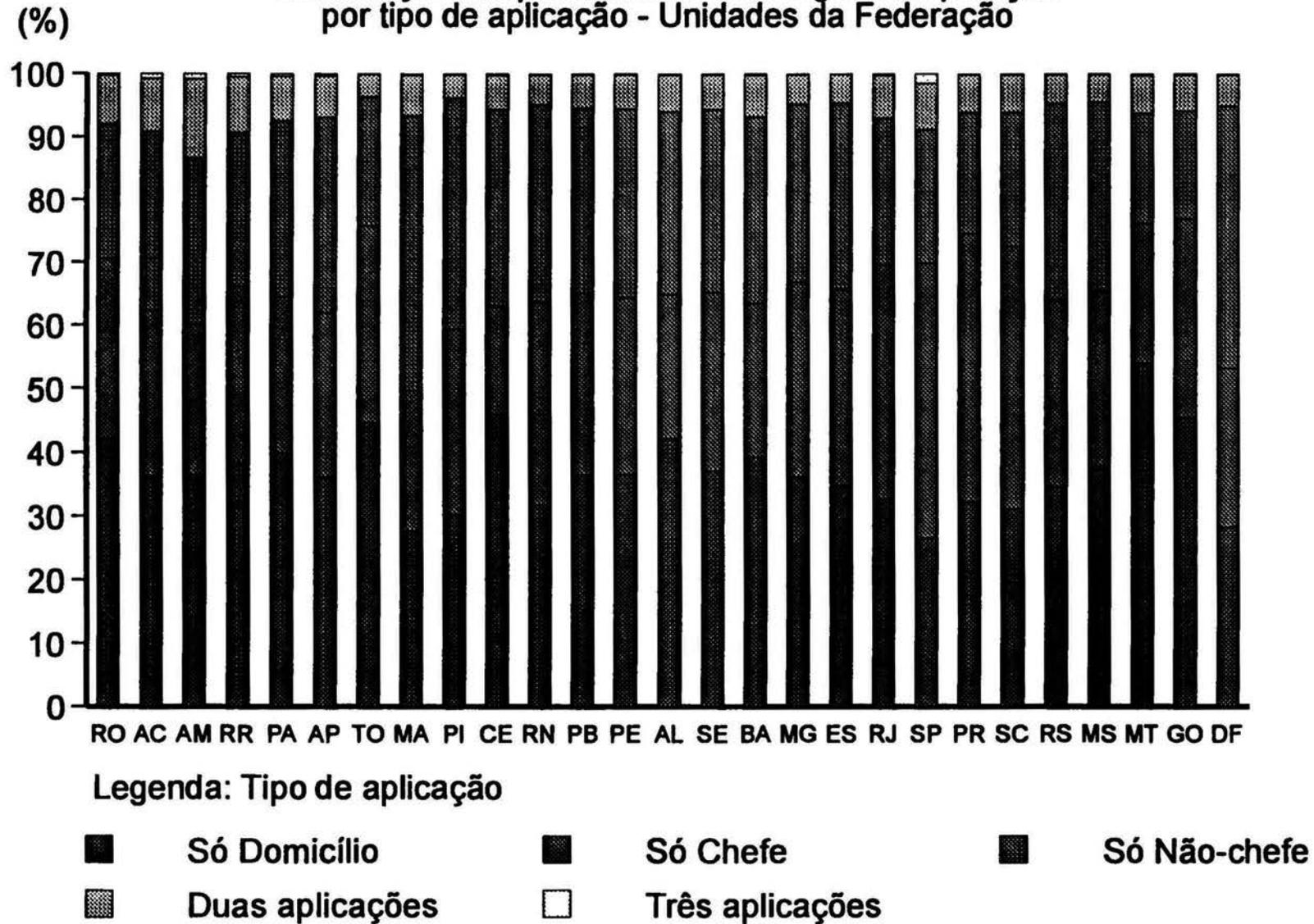
**Percentual de registros com alguma imputação na aplicação de Não-chefe - Unidades da Federação**



Fonte: Censo 91 - Resultados da Correção Automática do CD 1.01, usando o sistema DIA.

Gráfico A.7.5

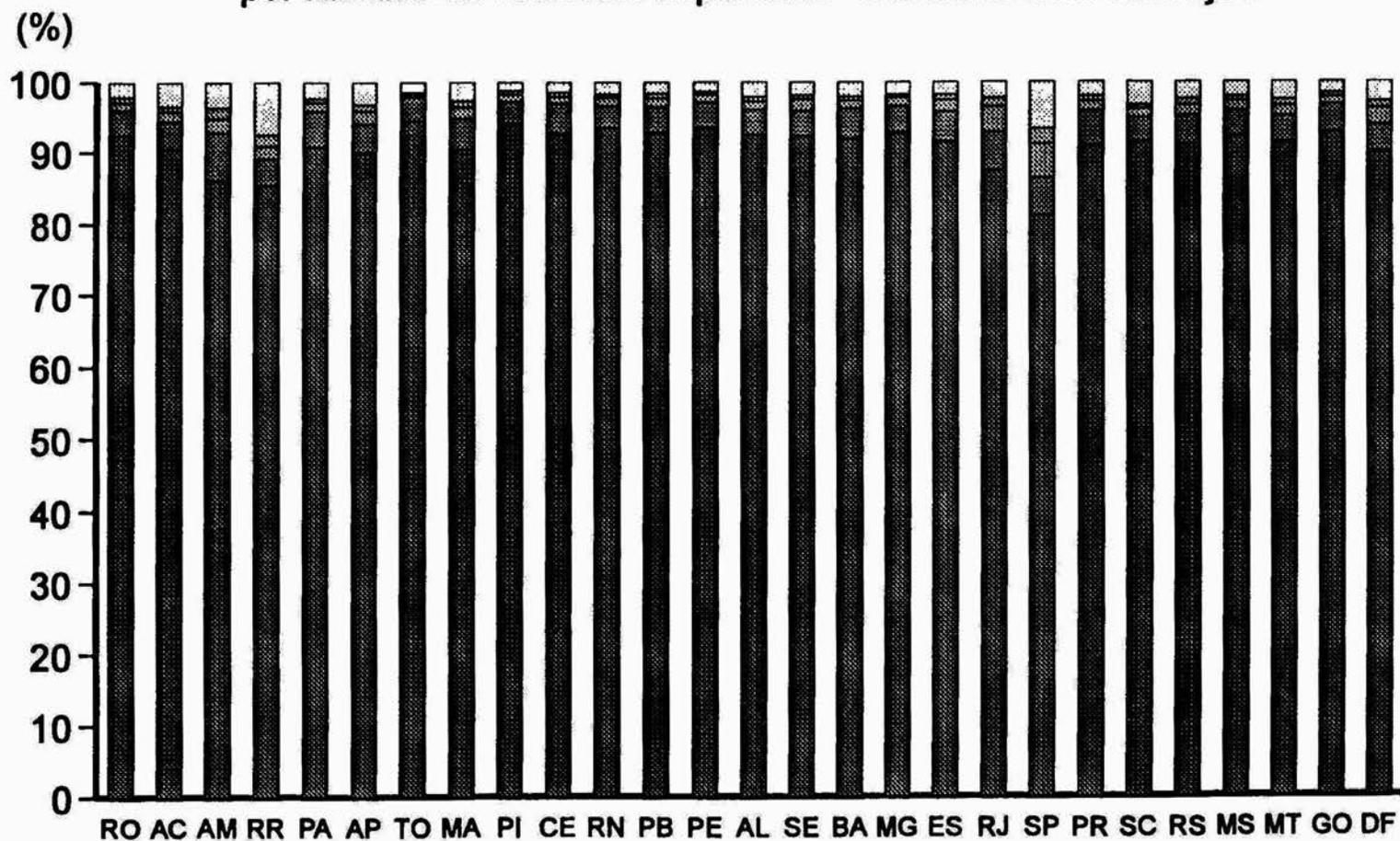
Distribuição de questionários com alguma imputação por tipo de aplicação - Unidades da Federação



Fonte: Censo 91 - Resultados da Correção Automática do CD 1.01, usando o sistema DIA.

Gráfico A.7.6

Distribuição de registros com alguma imputação na aplicação de Domicílio por número de variáveis imputadas - Unidades da Federação



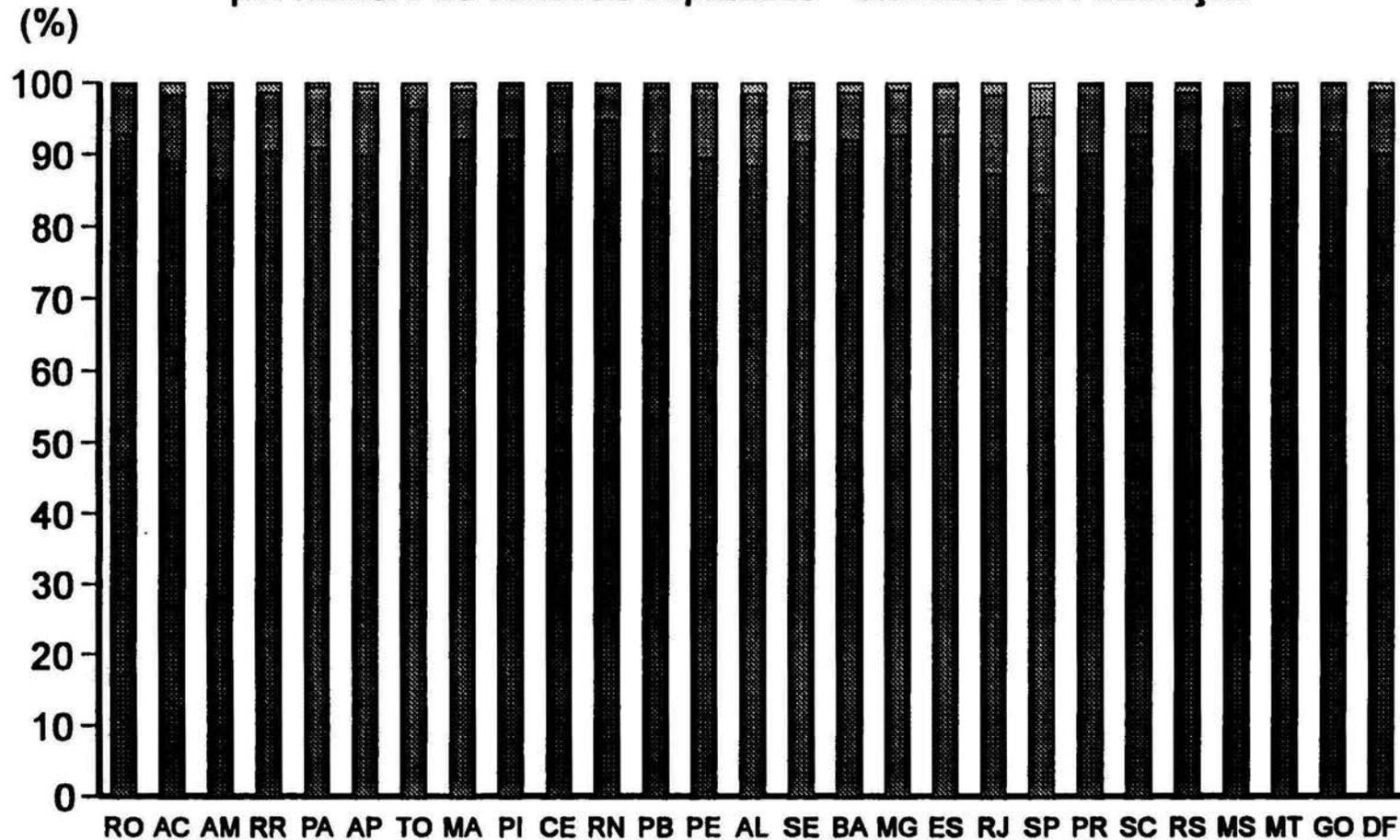
Legenda: Número de variáveis imputadas

■ 1    ■ 2    ■ 3 e 4    ■ 5 a 8    ■ 9

Fonte: Censo 91 - Resultados da Correção Automática do CD 1.01, usando o sistema DIA.

Gráfico A.7.7

Distribuição de registros com alguma imputação na aplicação de Chefe por número de variáveis imputadas - Unidades da Federação



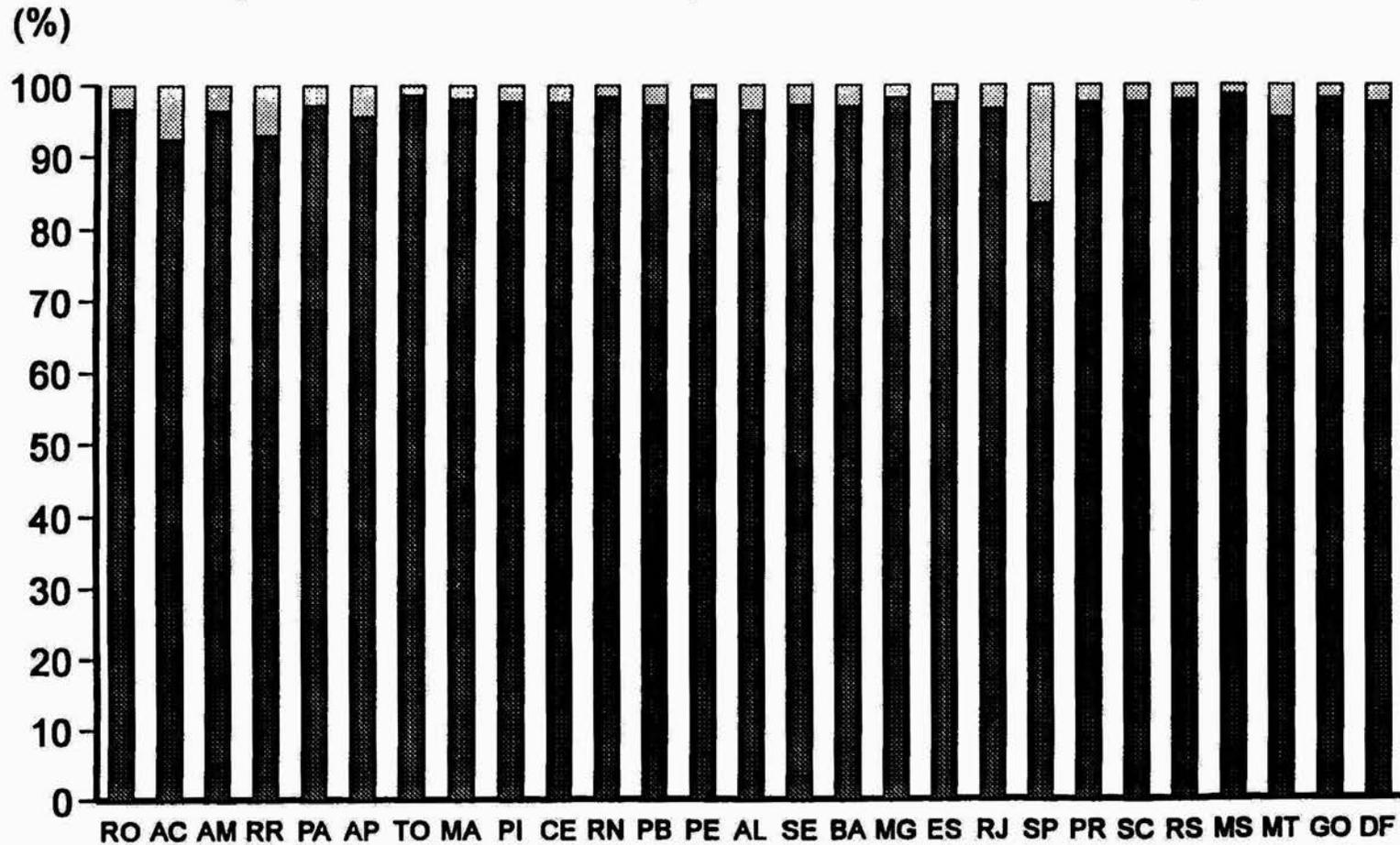
Legenda: Número de variáveis imputadas

■ 1    ■ 2    ■ 3 e 4    ■ 5

Fonte: Censo 91 - Resultados da Correção Automática do CD 1.01, usando o sistema DIA.

Gráfico A.7.8

Distribuição de registros com alguma imputação na aplicação de Não-chefe por número de variáveis imputadas - Unidades da Federação



Legenda: Número de variáveis imputadas

■ 1      □ 2

Fonte: Censo 91 - Resultados da Correção Automática do CD 1.01, usando o sistema DIA.

## **Textos para Discussão já publicados**

- ✓ *Pesquisas Contínuas da Indústria - Vol. 1, nº 1, janeiro 1988*
- ✓ *Pesquisas Agropecuárias Contínuas: Metodologia - Vol. I, nº 2, 1988*
- ✓ *Uma Filosofia de Trabalho: As experiências com o SNIPC e com o SINAPI - Vol. I, nº 3, março 1988*
- ✓ *O Sigilo das Informações Estatísticas: Idéias para reflexão - Vol. I, nº 4, abril 1988*
- ✓ *Projeções da População Residente e do Número de Domicílios Particulares Ocupados: 1985-2020 - Vol. I, nº 5, maio 1988*
- ✓ *Classificação de Atividades e Produtos, Matérias-Primas e Serviços Industriais: Indústria Extrativa Mineral e de Transformação - Vol. 1, nº 6, agosto 1988*
- ✓ *A Mortalidade Infantil no Brasil nos Anos 80 - Vol. I, nº 7, setembro 1988*
- ✓ *Ensaio sobre o Produto Real da Agropecuária - Vol. I, nº 9, setembro 1988*
- ✓ *Principais Características das Pesquisas Econômicas, Sociais e Demográficas - Vol. I, número especial, outubro 1988*
- ✓ *Novo Sistema de Contas Nacionais, Ano Base 1980 - Resultados Provisórios - Vol. I, nº 10, dezembro 1988*
- ✓ *Pesquisa de Orçamentos Familiares - Metodologia para Obtenção das Informações de Campo - nº 11, janeiro 1989*
- ✓ *De Camponesa a Bóia-fria: Transformações do trabalho feminino - nº 12, fevereiro 1989*
- ✓ *Pesquisas Especiais do Departamento de Agropecuária - Metodologia e Resultados - nº 13, fevereiro 1989*
- ✓ *Brasil - Matriz de Insumo-Produto - 1980 - nº 14, maio 1989*
- ✓ *As Informações sobre Fecundidade, Mortalidade e Anticoncepção nas PNADs - nº 15, maio 1989*
- ✓ *As Estatísticas Agropecuárias e a III Conferência Nacional de Estatística - nº 16, junho 1989*
- ✓ *Brasil - Sistema de Contas Nacionais Consolidadas - nº 17, agosto 1989*
- ✓ *Brasil - Produto Interno Bruto Real Trimestral - Metodologia - nº 18, agosto 1989*
- ✓ *Estatísticas e Indicadores Sociais para a Década de 90 - nº 19, setembro 1989*
- ✓ *Uma Análise do Cotidiano da Pesquisa no DEREN (As Estatísticas do Trabalho) - nº 20, outubro 1989*
- ✓ *Coordenação Estatística Nacional - Reflexões sobre o caso Brasileiro - nº 21, novembro 1989*

- ✓ *Pesquisa Industrial Anual 1982/84 - Análise dos Resultados* - nº 22, novembro 1989
- ✓ *O Departamento de Comércio e Serviços e a III Conferência Nacional de Estatística* - nº 23, dezembro 1989
- ✓ *Um projeto de Integração para as Estatísticas Industriais* - nº 24, dezembro 1989
- ✓ *Cadastro de Informantes de Pesquisas Econômicas* - nº 25, janeiro 1990
- ✓ *Ensaio sobre a Produção de Estatística* - nº 26, janeiro 1990
- ✓ *O Espaço das Pequenas Unidades Produtivas: Uma tentativa de delimitação* - nº 27, fevereiro 1990
- ✓ *Uma Nova Metodologia para Correção Automática no Censo Demográfico Brasileiro: Experimentação e primeiros resultados* - nº 28, fevereiro 1990
- ✓ *Notas Técnicas sobre o Planejamento de Testes e Pesquisas Experimentais* - nº 29, março 1990
- ✓ *Estatísticas, Estudos e Análises Demográficas - Uma visão do Departamento de População* - nº 30, abril 1990
- ✓ *Crítica de Equações de Fechamento de Empresas no Censo Econômico de 1985* - nº 31, maio 1990
- ✓ *Efeito de Conglomeração da Malha Setorial do Censo Demográfico de 1980* - nº 32, maio 1990
- ✓ *A Redução da Amostra e a Utilização de Duas Frações Amostrais no Censo Demográfico de 1990* - nº 33, junho 1990
- ✓ *Estudos e Pesquisas de Avaliação de Censos Demográficos - 1970 a 1990* - nº 34, julho 1990
- ✓ *A Influência da Migração no Mercado de Trabalho das Capitais do Centro-Oeste - 1980* - nº 35, agosto 1990
- ✓ *Pesquisas de Conjuntura: Discussão sobre Variáveis a Investigar* - nº 36, setembro 1990
- ✓ *Um Modelo para Estimar o Nível e o Padrão da Fecundidade por Idade com Base em Parturições Observadas* - nº 37, outubro 1990
- ✓ *A Estrutura Operacional de Uma Pesquisa por Amostra* - nº 38, novembro 1990
- ✓ *Produção Agrícola, Agroindustrial e de Máquinas e Insumos Agrícolas no Anos 80: Novas Evidências Estatísticas* - nº 39, dezembro 1990
- ✓ *A Inflação Medida pelo Índice de Preços ao Consumidor* - nº 40, janeiro 1991
- ✓ *A Participação Política Eleitoral no Brasil - 1988, Análise Preliminar* - nº 41, fevereiro 1991
- ✓ *Ensaio sobre Estatísticas do Setor Produtivo* - nº 42, março 1991

- ✓ *A Produção Integrada de Estatística e as Contas Nacionais: Agenda para Formulação de um Novo Plano Geral de Informações Estatísticas e Geográficas* - nº 43, março 1991
- ✓ *Matriz de Fluxos Migratórios Intermunicipais - Brasil - 1980* - nº 44, abril 1991
- ✓ *Fluxos Migratórios Intrametropolitanos - Brasil - 1970-1980* - nº 45, abril 1991
- ✓ *A Revisão da PNAD, A Questão Conceitual e Relatório das Contribuições* - nº- 46, maio 1991
- ✓ *A Dimensão Ambiental no Sistema de Contas Nacionais* - nº 47, maio 1991
- ✓ *Estrutura das Contas Nacionais Brasileiras* - nº 48, junho 1991
- ✓ *Mercado do Couro e Resultados da Pesquisa Anual do Couro* - nº 49, junho 1991
- ✓ *As Estatísticas e o Meio Ambiente* - nº 50, julho 1991
- ✓ *Novo Sistema de Contas Nacionais Séries Correntes : 1981-85 Metodologia, Resultados Provisórios e Avaliação do Projeto* - nº 51, julho 1991 ( 2 Volumes : Volume 1-Metodologia, Resultados Provisórios e Avaliação do Projeto; Volume 2-Tabelas)
- ✓ *O Censo Industrial de 1985 -- Balanço da Experiência* - nº 52, agosto 1991
- ✓ *Análise da Inflação Medida Pelo INPC 1989* - nº 53, agosto 1991
- ✓ *Revisão da PNAD : A Questão Amostral : Módulo II do Anteprojeto* nº 54, setembro 1991
- ✓ *A Força de Trabalho e os Setores de Atividade - Uma Análise da Região Metropolitana de São Paulo - 1986-1990* - nº 55, outubro 1991
- ✓ *Revisão da PNAD : Apuração das Informações : Módulo III do Anteprojeto* - nº 56, novembro 1991
- ✓ *Novos Usos para Pesquisa Industrial Mensal : A Evolução dos Salários Industriais, O Desempenho da Pecuária* - nº 57, novembro 1991
- ✓ *Revisão da PNAD : A Disseminação das Informações Módulo IV do Anteprojeto* - nº 58, dezembro 1991
- ✓ *Estatísticas Agropecuárias : Sugestões para o Novo Plano Geral de Informações* - nº 59, dezembro 1991
- ✓ *Análise Conjuntural e Pesquisa Industrial* - nº 60. janeiro 1992
- ✓ *Exploração dos Dados da Pesquisa Industrial Mensal de Dados Gerais* - nº 61, fevereiro 1992
- ✓ *Uma Proposta de Metodologia para a Expansão da Amostra do Censo Demográfico de 1991* - nº 62, outubro 1993
- ✓ *Expansão da Fronteira e Progresso Técnico no Crescimento Agrícola Recente* - nº 63, novembro 1993

- ✓ *Avaliação das Condições de Habitação com Base nos Dados da PNAD - nº 64, setembro 1993*
- ✓ *Análise da Taxa de Desemprego Feminino no Brasil - nº 65, dezembro 1993*
- ✓ *Aspectos da Metropolização Brasileira: Comentários sobre os Resultados Preliminares do Censo Demográfico de 1991 - nº 66, janeiro 1994*
- ✓ *Estimativas Preliminares de Fecundidade Considerando os Censos Demográficos, Pesquisas por amostragem e o Registro Civil - nº 67, janeiro 1994*
- ✓ *Apuração de Dados no IBGE: Problemas e Perspectivas - nº 68, fevereiro 1994*
- ✓ *Limeira - SP: Estimativas de Fecundidade e Mortalidade 1980/1988 - nº 69, março 1994*
- ✓ *Desemprego - Uma Abordagem Conceitual - nº 70, abril 1994*