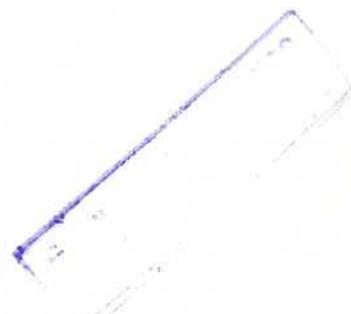


TEXTOS PARA DISCUSSÃO, ISSN 0103-6661

REVISÃO DA PNAD
APURAÇÃO DAS INFORMAÇÕES
MÓDULO III DO ANTEPROJETO

NÚMERO 56

NOVEMBRO DE 1991



Presidente da República
Fernando Collor de Mello

Ministro da Economia, Fazenda e Planejamento
Marcílio Marques Moreira

FUNDAÇÃO INSTITUTO
BRASILEIRO DE GEOGRAFIA
E ESTATÍSTICA - IBGE

Presidente
Eduardo Augusto Guimarães

Diretor-Geral
José Guilherme Almeida dos Reis

ÓRGÃOS TÉCNICOS SETORIAIS

Diretoria de Pesquisas
Lenildo Fernandes Silva

Diretoria de Geociências
Mauro Pereira de Mello

Diretoria de Informática
Nuno Duarte da Costa Bittencourt

Centro de Documentação e Disseminação de Informações
Nelson de Castro Senra



Fundação Instituto Brasileiro de Geografia e Estatística - IBGE
Diretoria de Pesquisas - DPE

REVISÃO DA PNAD

APURAÇÃO DAS INFORMAÇÕES

MÓDULO III DO ANTEPROJETO

ANTONIO CARLOS FERNANDES DE MENEZES
Estatístico, Gerente do Projeto de Reformulação
e Modernização, do DEREN

LUIZ FERNANDO RAMOS DE MELLO
Estatístico, Analista Especializado do DEREN

PEDRO LUIZ PINTO FELICÍSSIMO
Estatístico, Analista Especializado do DEREN

RIO DE JANEIRO

1991

FUNDAÇÃO INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE

Av. Franklin Roosevelt, 166 - Centro
CEP 20 021 - Rio de Janeiro, RJ - Brasil

DIRETOR DE PESQUISAS

LENILDO FERNANDES SILVA

DIRETORA-ADJUNTA DE PESQUISAS

MÁRCIA BANDEIRA DE MELLO LEITE

CHEFE DO DEPARTAMENTO DE EMPREGO E RENDIMENTO

MARIA MARTHA MALARD MAYER

CHEFE DA DIVISÃO DE ESTUDOS E ANÁLISES

PAULO CESAR DE SOUSA QUINTSLR

CHEFE DE PESQUISA MENSAL

MARILENI SILVA MANSOLDO

CHEFE DE PESQUISA ANUAL

TEREZINHA BATISTA TAVARES COUTINHO

© 1990,91

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: Núcleo de Documentação e Disseminação da Diretoria de Pesquisas.
(NDI/DPE)

Menezes, Antonio Carlos Fernandes de

Revisão da PNAD : apuração das informações : módulo III do anteprojeto / Antonio Carlos Fernandes de Menezes, Luiz Fernando Ramos de Mello, Pedro Luiz Pinto Felicíssimo. — Rio de Janeiro : IBGE, Diretoria de Pesquisas, 1991.

36 f. — (Textos para discussão / IBGE, Diretoria de Pesquisas, ISSN 0103-6661; n.56)

ISBN 84-240-0428-2

1. Pesquisa Nacional por Amostra de Domicílios. I. Mello, Luiz Fernando Ramos de. II. Felicíssimo, Pedro Luiz Pinto. III. IBGE. Diretoria de Pesquisas. IV. Título. V. Título: Pesquisa Nacional por Amostra de Domicílios: apuração das informações. VI. Série.

IBGE.CDDI. Dep. de Documentação e Biblioteca
RJ-IBGE/92-03

CDU 314.6(81)

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

Dando seqüência à elaboração do Anteprojeto da Pesquisa Nacional por Amostra de Domicílios - PNAD, apresentamos o Módulo III - A Apuração das Informações, que é precedido pelos Módulos I - A Questão Conceitual e II - A Questão Amostral, publicados na Série Textos para Discussão, Nos. 46,54.

Este módulo anuncia as questões a serem estudadas com o objetivo de definir-se o processo de apuração dos dados para a PNAD da década de 90.

O texto, ora apresentado, foi elaborado por Antonio Carlos Fernandes de Menezes (Gerente do Projeto de Reformulação e Modernização), Luiz Fernando Ramos de Mello e Pedro Luiz Pinto Felicissimo (técnicos do Projeto de Reformulação e Modernização).

Nas discussões que precederam esta versão final, foram muito importantes as contribuições dos técnicos Terezinha Batista Tavares Coutinho (Chefe da Divisão de Pesquisa Anual), Rodrigo Mariano Resende de Brito (Técnico da Divisão de Pesquisa Anual), Paulo Cesar de Souza Quintslr (Chefe da Divisão de Estudos e Análises), Rosângela Antunes Pereira Almeida (Chefe da Equipe de Estudos e Análises Estatísticas), Luiz Mauricio Rezende Mansur (Gerente do Projeto DEATE/DIDEM/DI) e Pedro Luiz do Nascimento Silva (Chefe do Núcleo de Metodologia da DPE).

Com esta divulgação, espera-se a leitura e as críticas e sugestões de todas as áreas do IBGE, interessadas na discussão do tema.

Maria Martha Malard Mayer

Chefe do Departamento de Emprego e Rendimento

SUMÁRIO

Apresentação

Introdução

1 - O Processo Atual de Apuração da PNAD

1.1 Descrição Geral

1.2 A Microconsistência dos Dados

1.2.1 Critica Visual e Codificação

1.2.2 Empastamento

1.2.3 Digitação dos Dados

1.2.4 Critica Semi-Automatizada

1.3 A Macroconsistência dos Dados

1.3.1 Critica Visual de Frequência de Variáveis

1.3.2 Critica Visual dos Dados Agregados em Tabelas

1.3.3 Cálculo e Análise dos Erros de Amostragem

2 - A Problematização do Processo

2.1 Aspectos Gerais

2.2 A Microconsistência dos Dados

2.3 A Macroconsistência dos Dados

3 - Considerações Finais

Bibliografia

Anexos

Diagrama de Fluxo de Dados

Executar a Apuração da Pnad

Executar a Microconsistência dos Dados

Executar a Macroconsistência dos Dados

INTRODUÇÃO

O processo de apuração das informações é, com certeza, das etapas de pesquisa que demanda aos técnicos do IBGE, a necessidade de reflexão cuidadosa. O que se tem hoje na Instituição, de maneira geral, são rotinas de apuração cuja concepção está baseada na tradição e experiência de realização de levantamentos censitários e que, em alguns casos, permanecem praticamente invariáveis em sua estruturação e operacionalização. Existe uma tendência a utilizar-se procedimentos idênticos, independente do porte das pesquisas, além do que estes costumam caracterizar-se pela repetição de atividades, verificações das tarefas já realizadas, documentação inadequada etc. Dado que boa parte das técnicas utilizadas decorrem de reflexões da década de 60, existe, no geral, um progresso limitado na incorporação de novas metodologias.

Várias questões contribuem para este quadro e, certamente, uma das mais relevantes é a resistência na aceitação de mudanças que se fazem necessárias para a implantação de processos mais modernos. Pode-se dizer que boa parte dos técnicos da Instituição mostra-se apreensiva quando, por exemplo, na atividade de crítica das informações, pensa-se na substituição de métodos manuais de correção por rotinas de imputação automática de dados.

Todos esses aspectos levam a uma outra questão, que tem a ver com o volume de recursos alocados à etapa de apuração das pesquisas, que normalmente representam boa parte do custo desses levantamentos. Esta situação pode também decorrer da idéia equivocada de que na apuração, a fase de detecção e correção de inconsistências, invariavelmente melhora a qualidade das informações da pesquisa. No IBGE, o que se tem

normalmente são rotinas, geralmente automatizadas, com grande poder de detecção de erros, mas associadas a critérios de correção, muitas vezes manuais, que não incorporam a confirmação das informações no campo. Assim, o que se faz na realidade, é uma adequabilidade dos registros incorretos a um conjunto de funções de crítica pré-estabelecido. Nesse caso, a meta principal deve ser a estruturação dos levantamentos, de modo que sejam dirigidos recursos adequados para a etapa de coleta das informações, ensejando que a qualidade dos dados seja verdadeiramente assegurada.

Deve-se perseguir na apuração, e particularmente na detecção e correção de inconsistências, o desenvolvimento de tarefas que, simplesmente, façam pequenos ajustes nos dados e que não tragam impactos às estatísticas divulgadas. O que se tem hoje bem caracterizado, é que os processos de apuração utilizados no IBGE têm sido, em grande medida, responsáveis pela demora na divulgação dos dados de muitas das pesquisas da Instituição, o que vem reduzindo, substancialmente, a utilidade desses levantamentos.

A Pesquisa Nacional por Amostra de Domicílios (1) pode ser colocada como exemplo desta e de outras dificuldades. Para este levantamento, revisar o processo de apuração de informações é imperioso, tanto pelo lado estritamente técnico da rotina atual, quanto pela sua adequabilidade à realidade de investigação de pesquisas no IBGE.

1 - O PROCESSO ATUAL DE APURAÇÃO DA PNAD

1.1 DESCRIÇÃO GERAL

Nesta parte do documento visa-se descrever, de maneira geral, a apuração das informações da PNAD, de modo que, a seguir, seja possível apresentar algumas dificuldades que vêm sendo encontradas para execução desta atividade no processo da pesquisa.

A apuração da pesquisa vem mantendo, basicamente, a mesma estrutura utilizada nas PNADs anteriores, concentrando-se em duas grandes etapas que são a Microconsistência e a Macroconsistência dos dados (ver diagramas em anexo).

A etapa de Microconsistência trabalha os dados a nível de cada unidade domiciliar e de cada pessoa entrevistada (microdado), registrados nos questionários aplicados às unidades da amostra selecionadas para a pesquisa, objetivando a adequabilidade destas informações, sendo realizada, descentralizadamente, através das seguintes atividades:

- Crítica Visual e Codificação;
- Empastamento;
- Digitação dos Dados ; e
- Crítica Semi-Automatizada

Antes de 1984, as Unidades Regionais - URs - participavam somente da fase de coleta da PNAD, sendo a fase de apuração centralizada no Rio de Janeiro. A partir dessa ocasião, a Crítica Visual, a

Codificação e o Empastamento começaram a ser levados para as URs, iniciando-se o processo de descentralização.

Com a implantação dos Centros de Processamento de Dados - CPDs - nas Unidades Regionais do Pará, Rio Grande do Norte, Ceará, Pernambuco, Minas Gerais, Espírito Santo, Rio de Janeiro, São Paulo, Paraná, Santa Catarina e Rio Grande do Sul, foi também descentralizada a Crítica Semi-Automatizada. As demais Unidades passaram então a encaminhar seu material para estes CPDs, num processo de adoção de áreas, que se completou em 1989. A seguir mostra-se a situação por ocasião da PNAD 1990.

PARÁ (ACRE, AMAZONAS, RORAIMA E AMAPÁ)

RIO GRANDE DO NORTE

CEARÁ (MARANHÃO, PIAUÍ)

PERNAMBUCO (ALAGOAS, PARAÍBA)

MINAS GERAIS (SERGIPE, GOIÁS)

ESPIRITO SANTO (BAHIA)

RIO DE JANEIRO (MATO GROSSO)

SÃO PAULO (RONDÔNIA)

PARANÁ

SANTA CATARINA (MATO GROSSO DO SUL)

RIO GRANDE DO SUL (DISTRITO FEDERAL)

A etapa de Macroconsistência, ainda não descentralizada, é realizada tendo como base os dados agregados por Unidade da Federação, Região Metropolitana, Grande Região e Brasil (macrodados), sendo concluída após as atividades de :

- Critica Visual de Frequência de Variáveis;
- Critica Visual dos Dados Agregados em Tabelas; e
- Cálculo e Análise dos Erros de Amostragem

1.2 - A MICROCONSISTENCIA DOS DADOS

1.2.1 - CRITICA VISUAL E CODIFICAÇÃO

A Critica Visual tem por finalidade verificar a compatibilidade das informações a nível de questionário e a observância das normas e critérios para o seu preenchimento.

Através desta critica são captadas algumas divergências, como a falta de informações, erro de sequência na entrevista, etc., retornando-se, quando necessário, à unidade domiciliar para os devidos esclarecimentos e acertos dos quesitos inconsistentes.

Já na Codificação, o principal objetivo é adequar certas informações ao programa de entrada de dados, onde os quesitos passíveis de codificação dentro do questionário (ocupação, atividade, etc) são analisados, para em seguida serem aplicados os códigos.

1.2.2 EMPASTAMENTO

Operação realizada após a Critica Visual e Codificação, objetivando facilitar a localização dos questionários e agilizar a sua movimentação nas etapas subsequentes, evitando que estes instrumentos sofram danos ou extravios.

As Unidades Regionais recebem do DEREN as pastas já identificadas pelo seu número, o número do primeiro e do último Número de Controle e o total de questionários que a pasta deverá conter. Além disso, as pastas seguem carimbadas, na parte superior à direita da capa, com o número da pasta e, na lombada, com o ano da pesquisa e o tipo de questionário - Básico ou Suplementar -. Ainda para a fase de Empastamento, o DEREN emite o relatório 'Registro de Controle do Material', com informações que identificam os questionários a serem empastados e campos em branco para serem preenchidos nas URs, com informações sobre o número de unidades investigadas, o total de unidades listadas e o tipo de entrevista. Esse relatório é devolvido ao DEREN para que seja atualizado o Arquivo de Controle da Amostra. Para maiores esclarecimentos a respeito destas duas fases deve-se consultar o texto referência (2) na Bibliografia.

1.2.3 - DIGITAÇÃO DOS DADOS

Esta fase é executada nas Unidades Regionais adotantes, através de um programa de entrada de dados, recorrendo-se à dupla digitação para reduzir-se os erros do processo. Os dados digitados são armazenados em dois arquivos em meio magnético, um contendo as informações do questionário básico da pesquisa e outro as do questionário suplementar.

1.2.4 - CRÍTICA SEMI-AUTOMATIZADA

Uma vez gerados os arquivos na fase de Digitação dos Dados, estas informações são lidas por meio de um programa reformatador, que faz as transformações dos registros anteriormente criados em dois tipos:

o primeiro contendo as informações que dizem respeito à Unidade Domiciliar e o outro as informações referentes às pessoas.

Esta fase do processo de Microconsistência (3) é executada por um conjunto de programas (10 ao todo) e tem como componentes principais a Crítica Quantitativa (QT) e a Crítica Qualitativa (QL). Estes programas, apesar de submetidos pelos CPDs regionais, tem o seu processamento realizado no ambiente do CPD da Diretoria de Informática DI. Após sua execução, gerados os resultados da QT e QL, estes são transmitidos aos CPDs regionais, onde impressos, serão trabalhados pelas equipes técnicas da PNAD (4 e 5).

A QT investiga os dados dos arquivos reformatados, apontando as incorreções devidas a falta ou excesso de informações. Identifica a omissão de investigação para moradores e, tendo como referência o Arquivo de Controle da Amostra, aponta também divergências na quantidade de unidades domiciliares investigadas nos setores/controles selecionados.

A QL analisa o microdado levando em conta um conjunto de funções de crítica (plano de críticas) que investiga a compatibilidade entre as informações dos quesitos para cada pessoa entrevistada e entre os dados para o domicílio (críticas intra registros), realizando também comparações entre os dados de pessoa da mesma família e de famílias conviventes (críticas entre registros). O resultado da QL aponta para a massa de dados três tipos de ocorrências: a impossibilidade indicando, para cada variável, os registros que não são admissíveis; a incompatibilidade que indica os registros impossíveis para uma ou mais variáveis em função de outras; e as advertências indicando, para

determinadas variáveis, os registros que ultrapassam limites considerados normais.

A Critica Semi-Automatizada é executada em passagens ou ciclos de critica. Os técnicos das URs de posse dos relatórios de QT e QL, fazem a análise das incorreções, tendo como referência os dados registrados nos questionários. Esclarecidas as situações de erro, são apontadas as correções, que depois de digitadas, constituem um arquivo magnético, posteriormente lido por um programa atualizador, que providencia a substituição dos valores incorretos nos arquivos formatados, anteriormente criados. Em seguida submete-se essa nova massa de dados aos programas de QT e QL, ocorrendo assim outra passagem de critica, com a emissão de novos relatórios para análise e correção dos erros remanescentes. Este procedimento é repetido até que os dados sejam considerados "limpos" ou livres das impossibilidades e incompatibilidades, e com as advertências esclarecidas.

Na QT, para correção de qualquer ocorrência, é indispensável que o questionário seja digitado, total ou parcialmente. No caso da QL, os acertos podem ser feitos no próprio relatório de critica ou em planilhas de correção, sendo estes os documentos remetidos para a digitação.

1.3 - A MACROCONSISTENCIA DOS DADOS

1.3.1 - CRÍTICA VISUAL DE FREQUENCIA DE VARIÁVEIS

De posse das listagens emitidas pela DI, os técnicos do DEREN executam essa atividade (6), cujo procedimento inicial é fazer a comparação dos totais de homens e mulheres com os registrados nos relatórios finais da Crítica Semi-Automatizada, para verificar se está sendo lido o arquivo que contém os microdados consistentes.

Esta fase tem prosseguimento com a inspeção das frequências apuradas para os itens do questionário, orientada por uma rotina que reúne relações de igualdade e desigualdade entre os macrodados.

Qualquer erro detectado, implica em novo exame da etapa de Microconsistência, sendo necessário reabrir a fase de Crítica Semi-Automatizada.

1.3.2 - CRÍTICA VISUAL DOS DADOS AGREGADOS EM TABELAS

A crítica do plano tabular é realizada, também, através do exame orientado por uma rotina que reúne relações matemáticas, tendo como referência as células das diversas tabelas.

As informações, segundo os níveis geográficos, são emitidas primeiramente sem ponderação - dados da amostra - sendo então aplicada a rotina para o estabelecimento de sua consistência (7). Da mesma forma que na fase anterior, quando detectada uma inconsistência, é necessário o exame da etapa de Microconsistência.

Dando seguimento a esta fase, faz-se o exame do plano tabular contendo os dados expandidos (8), verificando-se a ponderação, comparando esses valores com as estimativas de população projetadas pelo DEPOP, para a data de referência da pesquisa; por fim, aplica-se outra rotina de verificação entre as células destas tabelas.

1.3.3 - Cálculo e Análise dos Erros de Amostragem

A tarefa de calcular os erros amostrais é de responsabilidade dos técnicos do DEREN e da Diretoria de Informática, e seu resultado são os valores dos coeficientes de variação associados às estimativas divulgadas pela pesquisa.

Como a PNAD investiga, em sua grande maioria, variáveis qualitativas, é possível ajustar uma forma matemática que associe ao valor das estimativas divulgadas o seu coeficiente de variação. Dessa maneira, de posse dos parâmetros da função é possível obter-se estimativas dos coeficientes para a quase totalidade das células do plano tabular.

A rotina consiste em selecionar para cada nível geográfico de divulgação da PNAD - Unidade da Federação e Região Metropolitana - segundo a situação do domicílio, uma amostra de células das tabelas dos planos tabulares, separadamente para o corpo básico e suplementos por variável - de pessoa ou de domicílio. A seguir, solicita-se à DI a programação para cálculo dos respectivos coeficientes de variação.

De posse dos pares ordenados - estimativas e coeficientes - executa-se uma rotina de ajustamento, estabelecendo-se assim os

parâmetros das funções, uma para cada nível geográfico, segundo a situação do domicílio e tipo de variável.

A seguir é realizada uma análise estatística de cada ajustamento e por fim é feita uma verificação da magnitude dos coeficientes de variação estimados. Só depois disso, os coeficientes são liberados e encaminhados à divulgação, juntamente com as tabelas do plano tabular, encerrando-se assim a apuração das informações da PNAD.

2 - A PROBLEMATIZAÇÃO DO PROCESSO

2.1 ASPECTOS GERAIS

Para tratar os problemas do processo de apuração da PNAD é conveniente, num primeiro momento, identificar as duas fontes principais que contribuem para essas ocorrências; são elas a carência de recursos e a pouca eficiência de algumas das rotinas de trabalho. A primeira contribui para que o processo de apuração das informações não seja, no seu todo, executado convenientemente, ou de acordo com o que está determinado; a outra tem a ver com o desempenho do processo, estando diretamente ligada à concepção do próprio sistema de apuração.

A exiguidade de recursos pode ser melhor entendida ao considerar-se os seguintes aspectos. Até o final da década de 70, a PNAD pôde contar com recursos financeiros bastante significativos. Implantada durante o período do "milagre econômico", conseguiu-se por muito tempo atender aos requerimentos financeiros para as suas diversas fases de execução.

No entanto, a década de 80 mostrou exatamente o contrário, verificando-se, ao longo do tempo, uma redução real das verbas destinadas à pesquisa. Isto, em grande parte, explica a dificuldade crescente em constituir-se as equipes da PNAD, principalmente a partir de 1982. Com a descentralização gradativa das atividades de pesquisa foi ficando cada vez mais patente o fato de que os recursos disponíveis eram insuficientes, não só para a apuração das informações, mas para outras tarefas do levantamento.

Outra questão importante é que, desde o lançamento da pesquisa, o processo de apuração das informações não incorporou mudanças no que diz respeito aos seus aspectos fundamentais ou à sua essência. Assim, como já foi dito na Introdução, a PNAD carrega até hoje uma visão desatualizada do processo de apuração de informações que, por exemplo, determina a inspeção integral do material coletado, a execução repetida de rotinas e a verificação sistemática das tarefas executadas.

Estes e outros pontos que serão a seguir tratados, fazem que para a execução do processo de apuração, seja requerido um montante significativo de recursos. Deve-se ter em conta também, que a descentralização da etapa de Microconsistência, basicamente, transferiu tarefas, antes realizadas no DEREN, para as Unidades Regionais, praticamente não acarretando redução de custos. A única providência específica para diminuir os custos do levantamento, foi o estudo (10) realizado em 1986, para a redução da amostra da PNAD. Este trabalho indicou a viabilidade de uma diminuição de, aproximadamente, 44% no número de domicílios a serem selecionados para investigação da pesquisa do ano seguinte. Deve-se considerar no entanto, que ao nível das Unidades da Federação, esse percentual variou e que a amostra reduzida manteve inalteradas as unidades de primeiro e segundo estágios - municípios e setores - selecionadas no início da década. Assim, houve uma redução diferenciada de custos, e pôde-se somente trabalhar o último estágio de seleção da amostra, que não é o mais importante, quando se trata da demanda de recursos.

Pelo exposto, fica claro que, para a PNAD a partir dos anos 80, a redução na oferta de recursos financeiros não correspondeu a uma diminuição adequada pelo lado da demanda; e a apuração, por representar

parcela importante dessa demanda, teve problemas para o desenvolvimento de suas atividades.

Tratada a questão de recursos, serão a seguir apresentados alguns problemas relativos à eficiência das rotinas de trabalho, fazendo-se referência à cada etapa do processo de apuração da PNAD.

2.2 A MICROCONSISTENCIA DOS DADOS

O objetivo principal para a descentralização da Crítica Visual de Microconsistência é fazer com que esta atividade seja desenvolvida o mais próximo possível das fontes de informação, favorecendo a tarefa de volta ao informante para a correção de erros. No entanto, a falta de pessoal em muitas Agências de Coleta, faz com que o supervisor, terminado o trabalho na etapa de entrevista, não possa realizar as tarefas de Crítica Visual. Isto faz com que o material seja enviado à sede da UR para ser criticado, o que dificulta sobremaneira o retorno ao campo. Por outro lado, deve-se registrar que a atitude de algumas URs de reverem o trabalho de Crítica Visual, aponta a pouca eficiência dessa tarefa quando realizada ao nível das agências.

Para a Codificação prevalece o mesmo raciocínio desenvolvido anteriormente quanto ao aspecto de recursos humanos, mas deve-se considerar que, em condições normais, a necessidade de volta ao informante por erros nesta fase é bem menor que na anterior. Isto faz com que possa ser desenvolvida centralizadamente na UR, afastado das fontes de informação.

Acontece que a atividade de Codificação exige para a sua execução técnicos experimentados, caso contrário podem acontecer erros

frequentes. Isto faz com que as equipes das coordenações estaduais tenham também motivos para rever esta tarefa, quando realizada nas agências. Sobressai também, o fato de que o processo de Codificação é bastante trabalhoso e existe dificuldades para realização de um trabalho homogêneo em qualidade. A relação de códigos das ocupações vai-se desatualizando com o tempo, e o processo de incorporação de novos códigos, em cada UR, é de difícil disseminação a nível nacional, acontecendo o mesmo para outras decisões ao longo da atividade. Espera-se que a estratégia da codificação automática/assistida, a ser desenvolvida no Censo Demográfico de 1991, venha eliminar ou minimizar bastante estes problemas nas futuras PNADs.

Concluindo, deve ser citado que tanto a Crítica Visual como a Codificação não contribuem para o aprimoramento futuro destas atividades, na medida em que não se dispõe de registros que informem a respeito da qualidade do trabalho realizado e, por exemplo, possam orientar a tarefa de treinamento das equipes.

A fase de Digitação das informações da PNAD constitui-se numa atividade que agrega aos dados uma quantidade de erros bastante superior ao que se poderia admitir. Isto se verifica também, devido ao fato de o programa de entrada de dados, que orienta esta atividade, não incorporar qualquer crítica que rejeite informações digitadas incorretamente.

Deve-se reconhecer que qualquer função de crítica executada em conjunto com a Digitação, reduz a velocidade desta tarefa. No entanto, é necessário considerar que a menor agilidade da entrada de dados será aceitável, na medida em que mais do que compense o tempo e o esforço gasto na correção destes erros pelas fases seguintes da apuração. Na PNAD, é razoável admitir que o tipo de crítica que rejeite

informações incompatíveis com os valores possíveis das variáveis do questionário, possa ser incorporado à Digitação, trazendo vantagens ao processo de apuração como um todo.

No que se refere à fase de Crítica Semi-Automatizada, as situações que acarretam a perda de eficiência de suas rotinas, vão desde os equívocos de concepção, passam por dificuldades ligadas ao sistema informatizado que lhes dá suporte, indo até a problemas associados a operacionalização do processo.

Primeiramente deve ser apontado que a dificuldade da comunicação entre os técnicos do DEREN e da DI, é uma fonte importante de problemas para as atividades subsequentes do sistema de Crítica Semi-Automatizada da pesquisa. Isto acontece quando o DEREN explicita aos técnicos da DI, em linguagem corrente, as funções de crítica que devem compor o plano de detecção de erros - planos de QT e QL - e estes precisam transformá-las em 'statements' de programas, em linguagem adequada ao processamento informatizado dos dados. Por tratar-se de uma pesquisa de investigação complexa, a solicitação dos técnicos envolve um grande número de variáveis e, conseqüentemente de funções de crítica, cuja compreensão não é nada trivial. O mais importante, no entanto, é que os técnicos do DEREN desconhecem as linguagens de processamento, o que imediatamente dificulta a comunicação e o entendimento necessário à validação do trabalho realizado.

Atualmente, este processo de explicitação das funções de crítica em dois estágios pode ser eliminado completamente, pela utilização de 'softwares' (12) que dão a oportunidade aos técnicos diretamente responsáveis pela pesquisa - no caso o DEREN - de uma única vez, definirem as críticas (11). Isto se faz por meio de

declarações em formas bastante simples, que são diretamente processadas pelos programas e, é claro, permitem a inspeção e validação direta por quem as definiu.

Outro ponto importante é que o resultado da aplicação do plano de crítica da PNAD às informações coletadas, é distinto daquele que inicialmente foi estabelecido como meta durante a tarefa de definição das funções de crítica. Isto acontece, basicamente, pelo fato de não se conseguir controlar o comportamento efetivo do plano, sem uma análise lógica do seu todo. Das funções de crítica de QT e QL, aproximadamente 800 no questionário básico, decorrem outras tantas, não percebidas, que determinam um nível de redundância desconhecido, e podem gerar conflitos em relação aos objetivos inicialmente estabelecidos, pois a conjugação de duas ou mais funções, invariavelmente gera outras inteiramente novas.

Para que se tenha segurança, é muito importante que o plano de críticas das informações da PNAD revisada seja logicamente analisado. O IBGE dispõe, para uso imediato, de rotina informatizada (11) que pode realizar este trabalho, apontando com exatidão todas as funções de crítica redundantes ou conflitantes, bem como as derivadas. Estes programas, com base na relação de códigos possíveis para o conjunto de itens do questionário, também identificam grande parte dos erros de transcrição das funções de crítica, seja por grafia ou os introduzidos pela digitação.

A validação integral do plano de Crítica Semi-Automatizada da PNAD, é, até hoje, uma questão duvidosa. Esta situação é característica de todo o plano de crítica que depende do estabelecimento de massas de teste, para avaliar seu comportamento. Assim, essa avaliação será tão boa quanto for boa a massa de teste, e sabe-se que numa pesquisa

complexa é praticamente impossível prever-se todas as situações de erro, podendo acarretar que o plano não avalie com precisão a massa real de dados da pesquisa.

Devido a forma como estão programadas, algumas rotinas do sistema responsável pela QL exigem que, a cada realização da pesquisa, sejam repassados parâmetros para atualização de algumas funções de crítica sobre os dados de rendimento e escolaridade. Esta operação, de descrição bastante simples, assume, na prática, conotação diferente, pois torna-se até necessário, que o DEREN elabore massas de teste para validar essa atualização.

Sob o ponto de vista operacional da detecção e correção de erros, podem ser comentados, dentre outros, os seguintes pontos sobre a Crítica Semi-Automatizada de Microconsistência.

Na Crítica Quantitativa, boa parte das ocorrências dizem respeito a enganos nos itens de Identificação dos questionários. A partir do momento que foi posto em prática o sistema de etiquetas, houve uma redução considerável desse tipo de erro, que ocorre, na grande maioria das vezes, quando da transcrição dessas informações. No entanto, esse problema continua presente nos domicílios onde se faz necessário a utilização de folhas internas, para as quais não existe condição de emissão prévia de etiquetas. Para a PNAD futura deve-se estudar a melhor forma de criticar-se os itens de identificação, e já durante a Digitação, através do programa de entrada de dados, eliminar quase que totalmente essas irregularidades.

Na Crítica Qualitativa, deve-se registrar que todas as funções de crítica operacionalizadas na fase de Crítica Visual e muitas vezes revisadas pelos técnicos das coordenações estaduais, são repetidas de

forma automatizada. Em grande parte, os erros ainda detectados não podem mais ser esclarecidos pela volta ao campo, principalmente no caso das áreas que remetem seu material para ser trabalhado em outras Unidades Regionais. Devido ao tempo decorrido entre a data da entrevista e a realização da QL, até o contato telefônico pode não resolver, pois toda a estrutura montada para a realização das entrevistas, ao nível das agências, não raro, já se encontra desativada. Dessa forma a providência que se toma é a imputação manual de informações que, por mais cuidadosa, levanta sempre a questão da subjetividade e conseqüentemente da homogeneidade de tratamentos das situações, principalmente em pesquisas de âmbito nacional, apuradas descentralizadamente, como a PNAD; isto, em hipótese alguma, deve ser compreendido como um discurso contra a descentralização da apuração da pesquisa.

Ainda na QL, tem-se a questão bastante séria do elevado número de passagens para que os dados sejam considerados livres de incorreções, o que tem levado à demora na divulgação das informações e a um aumento considerável do custo da fase de apuração. Isto ocorre, principalmente, pelo fato de que a correção dos erros apontados nos relatórios de uma passagem da crítica, podem gerar a detecção de outros enganos na passagem seguinte. Esta situação, durante muito tempo, vem dificultando a etapa de apuração em pesquisas mais complexas, e deve-se considerar que mesmo alguns métodos de correção automática tidos como avançados, conseguem minimizar, mas não eliminar, este problema. Mais recentemente, já é possível contar com softwares que integram logicamente a detecção e correção de erros, e promovem imputações tendo como referência todo o registro de informações para cada unidade de observação investigada - no caso da PNAD trata-se da pessoa - e o conjunto das funções de crítica.

Estes recursos, já disponíveis no IBGE, através de critérios matemáticos e estatísticos, modificam o menor número possível de informações, para que o registro satisfaça, após uma única passagem, à todas as funções de crítica estabelecidas.

Outra dificuldade é a quase impossibilidade de o processo de Crítica Semi-Automatizada da pesquisa gerar informações sobre a qualidade dos dados que são apresentados à crítica, e dos efeitos que esta atividade tem sobre os resultados da pesquisa. Basicamente, isto ocorre pela quase ausência do levantamento sistematizado de estatísticas, e pela dificuldade de utilização das atualmente apresentadas nos Relatórios de QT e QL, para qualquer tentativa de avaliação.

Aliado a estas questões, tem-se, por exemplo, outras bastante particulares como a que acontece com a verificação das mensagens de advertência na Crítica Qualitativa. Esse tipo de ocorrência que, como se sabe, não necessariamente trata-se de um erro, deveria ser verificada e esclarecida durante a passagem de crítica que a identificou. No entanto, as advertências permanecem registradas até a passagem final da crítica e por isso estão sempre sujeitas à verificações. A questão é mais séria, considerando que ao final da Crítica Semi-Automatizada, quando os questionários são enviados pelos CPDs regionais ao DEREN, este, de pronto, verifica novamente todas as advertências e, muitas vezes, é obrigado a reabrir a QL. Mais lógico seria verificar-se, uma única vez e descentralizadamente, essas ocorrências, criando-se mecanismos de controle da qualidade dessa verificação.

2.3 - A MACROCONSISTENCIA DOS DADOS

A problematização desta etapa do processo de apuração pode ser feita em grande parte, globalmente, pois algumas questões impactam tanto a fase de Critica Visual de Frequência de Variáveis como a Critica Visual dos Dados Agregados em Tabelas.

A questão central é: quase todas as atividades, da forma como são desenvolvidas, da etapa de Macroconsistência dos dados da PNAD, têm a finalidade, ou de detectar erros remanescentes da Microconsistência, principalmente da Critica Semi-Automatizada, ou de corrigir possíveis enganos - inclusive erros de programação - no processamento de dados, na própria Macroconsistência.

É fácil entender que grande parte da etapa atual da Macroconsistência não se justifica, a não ser pela fragilidade das rotinas desenvolvidas anteriormente. Ou seja, se o microdado estivesse, ao fim da Critica Semi-Automatizada, realmente livre de irregularidades, qualquer distorção detectada nas frequências de variáveis ou nos dados tabulados, seriam devidas ao processamento destas atividades.

Assim, as verificações do total de pessoas por sexo, realizada na Critica Visual de Frequência de Variáveis, têm por finalidade conferir, unicamente, se as fitas magnéticas lidas foram, realmente, as correspondentes às da última passagem da Critica Semi-Automatizada de Microconsistência. As verificações de ponderação se resumem, efetivamente, a saber se os valores das projeções independentes foram passados corretamente para os programas de expansão dos dados. Em seguida, os procedimentos da Critica Visual dos Dados Agregados em Tabelas são feitos para os dados da amostra e a seguir para os

expandidos ou seja, essa repetição tem a finalidade de verificar se o programa de expansão leu as fitas corretas.

Quanto ao tratamento dado aos erros amostrais, duas questões devem ser citadas. Em primeiro lugar, o processo de cálculo desses erros precisa ser incorporado à rotina de emissão do plano tabular dos dados expandidos, de modo que, ao serem emitidas as tabelas, tenha-se ao mesmo tempo os coeficientes de variação. O outro aspecto é a necessidade de, além de verificar a magnitude dos coeficientes, como já é feito, realizar-se uma análise estatística das estimativas divulgadas.

Admitindo-se que a etapa de Microconsistência não limpa os dados, coisa que é verdade, poder-se-ia ter grande parte da Macroconsistência realizada simplesmente pela inspeção das mensagens do sistema controlador de "jobs", para ter-se a certeza da leitura das fitas corretas, e as rotinas de verificação de frequências e de dados tabulados seriam realizadas através de programas informatizados.

Como pode ser visto, o que se tem a perseguir para a PNAD revisada, é a eliminação de quase todo o trabalho feito hoje na Macroconsistência, substituindo-o por outro mais produtivo, voltado para a análise dos dados obtidos, buscando atender, de maneira mais rica, os usuários da pesquisa.

3 - CONSIDERAÇÕES FINAIS

Não é difícil admitir-se que o processo de apuração das informações da PNAD, baseia-se, grandemente, em rotinas que tentam validar outras atividades realizadas anteriormente no próprio processo. Assim, vê-se o trabalho muito carregado de verificações, o que aponta uma quase total desconfiança nestes procedimentos, e também nas equipes que os operacionalizam; é certo que estas atividades, ao invés de somente corrigir, funcionam também como introdutoras de erros.

Um outro aspecto do que foi citado no parágrafo anterior, é a total impossibilidade de atribuir-se responsabilidade pelos enganos cometidos, pois só assim pode-se justificar tantas repetições ou 'controles'. Acredita-se que este fato deve-se, em muito, a pouca eficiência de algumas rotinas do processo de apuração. Tem-se como claro o fato de que, mesmo que funcione obedecendo ao que foi estabelecido, a apuração da PNAD ainda carregaria muitos dos problemas até aqui descritos.

Por fim, a inadequabilidade do processo contribuiu para a sua inadaptabilidade às possibilidades do próprio IBGE, no que se refere a recursos financeiros e humanos. Ou seja, deve-se ter como meta para a PNAD futura, um processo de apuração moderno, mas que necessariamente se ajuste às condições oferecidas pelo IBGE; de nada valerá contar-se com a técnica mais apurada, se esta não se mostrar viável do ponto de vista da Instituição.

Para a consecução destes objetivos os técnicos do DEREN - do Projeto de Reformulação e Modernização e da Divisão de Pesquisa Anual- vêm , em conjunto com técnicos do Núcleo de Metodologia - NME da Diretoria de Pesquisas e da Diretoria de Informática, examinando rotinas alternativas para apuração de pesquisas, estando ou não suas metodologias disponíveis, neste momento, no IBGE.

A idéia é optar por um modelo de apuração depois de um levantamento cuidadoso das opções e, principalmente, de uma reflexão atenta, tendo como apoio o amplo debate técnico.

No entanto, a decisão por esta ou aquela metodologia de apuração, irá depender ainda de uma atividade de simulação, prevista no cronograma de revisão da PNAD, tendo como referência informações da última pesquisa, contando com a participação das áreas já citadas e dos técnicos das Unidades Regionais que, desde o início, vêm participando das atividades de revisão.

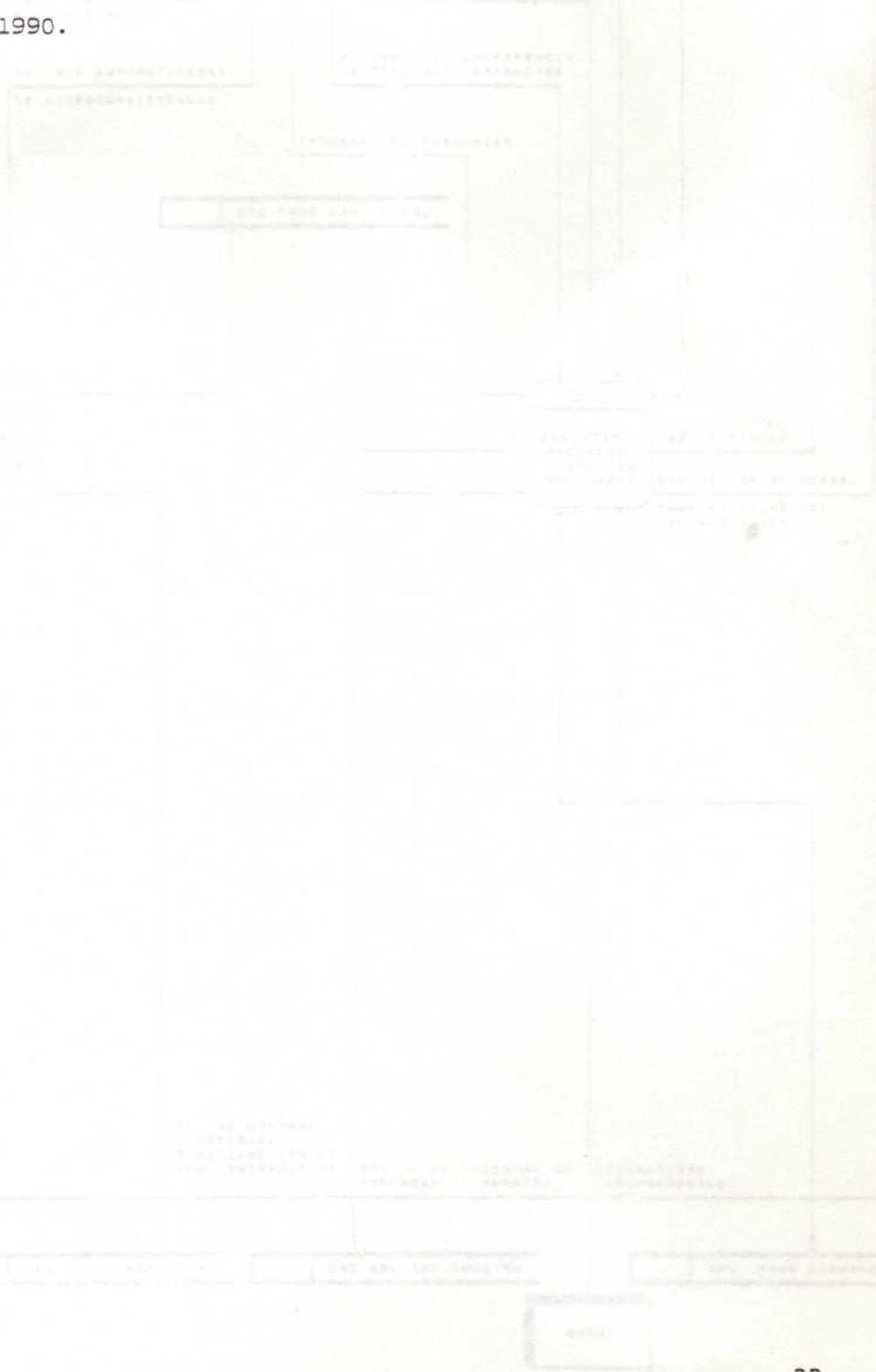
BIBLIOGRAFIA

- 1 - Metodologia da Pesquisa Nacional por Amostra de Domicílios na Década de 70. Rio de Janeiro: IBGE, 1981 (Séries Relatórios Metodológicos, 1)
- 2 - Manual de Crítica Visual de Microconsistência, Codificação e Empastamento - PNAD 1990. Rio de Janeiro: IBGE, Departamento de Emprego e Rendimento, 1990.
- 3 - Mansur, Luiz Maurício Rezende; Relatório de Avaliação do Sistema Atual PNAD, Rio de Janeiro: IBGE, DI, DIDEM, DEATE.
- 4 - Manual de Crítica Semi-Automatizada de Microconsistência - PNAD 1990. Rio de Janeiro: IBGE, Departamento de Emprego e Rendimento, 1990.
- 5 - Manual de Crítica Semi-Automatizada de Microconsistência - Manual do Usuário - PNAD 1990. Rio de Janeiro: IBGE, Departamento de Emprego e Rendimento, 1990.
- 6 - Rotina de Crítica Visual de Frequência das Variáveis da Pesquisa. Rio de Janeiro: IBGE, Departamento de Emprego e Rendimento, 1990
- 7 - Rotina para conferência das Tabelas da Amostra. Rio de Janeiro: IBGE, Departamento de Emprego e Rendimento, 1990.
- 8 - Rotina para conferência da Tabulação Expandida. Rio de Janeiro: IBGE, Departamento de Emprego e Rendimento, 1990.
- 9 - O Processo de Apuração da PNAD na Década de 80. Rio de Janeiro: IBGE, Departamento de Emprego e Rendimento, 1990
- 10 - Menezes, Antonio Carlos Fernandes de; et all. Pesquisa Nacional por Amostra de Domicílios: Redução do Número de Domicílios na Amostra e Proposta de Novos Estimadores. Rio de Janeiro: IBGE, DEREN, 1986

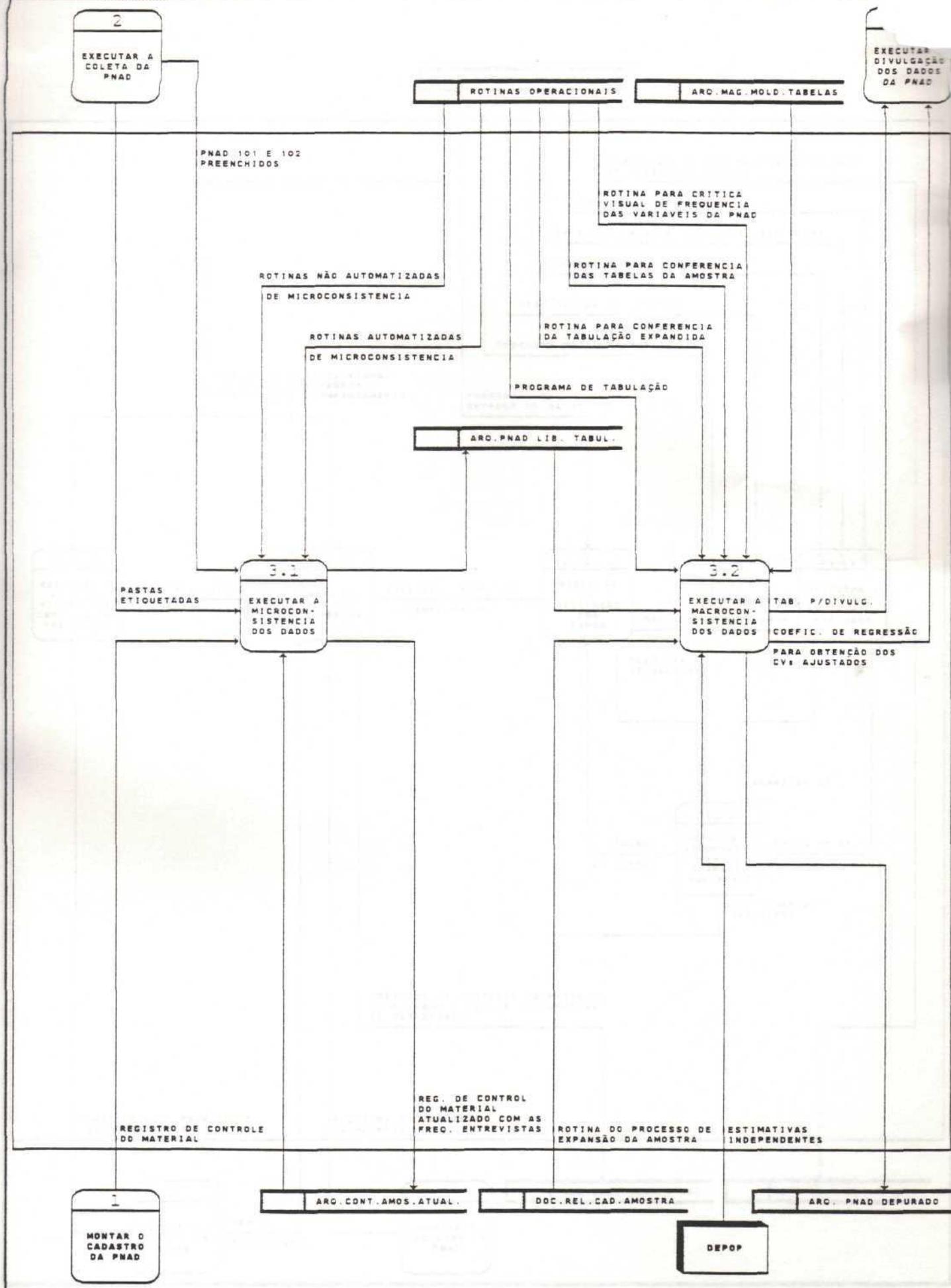
11 - Rubio, Emiliano Garcia; criado, Ildefonso Villan. Sistema DIA.

Madrid: Instituto Nacional de Estadística, 1988.

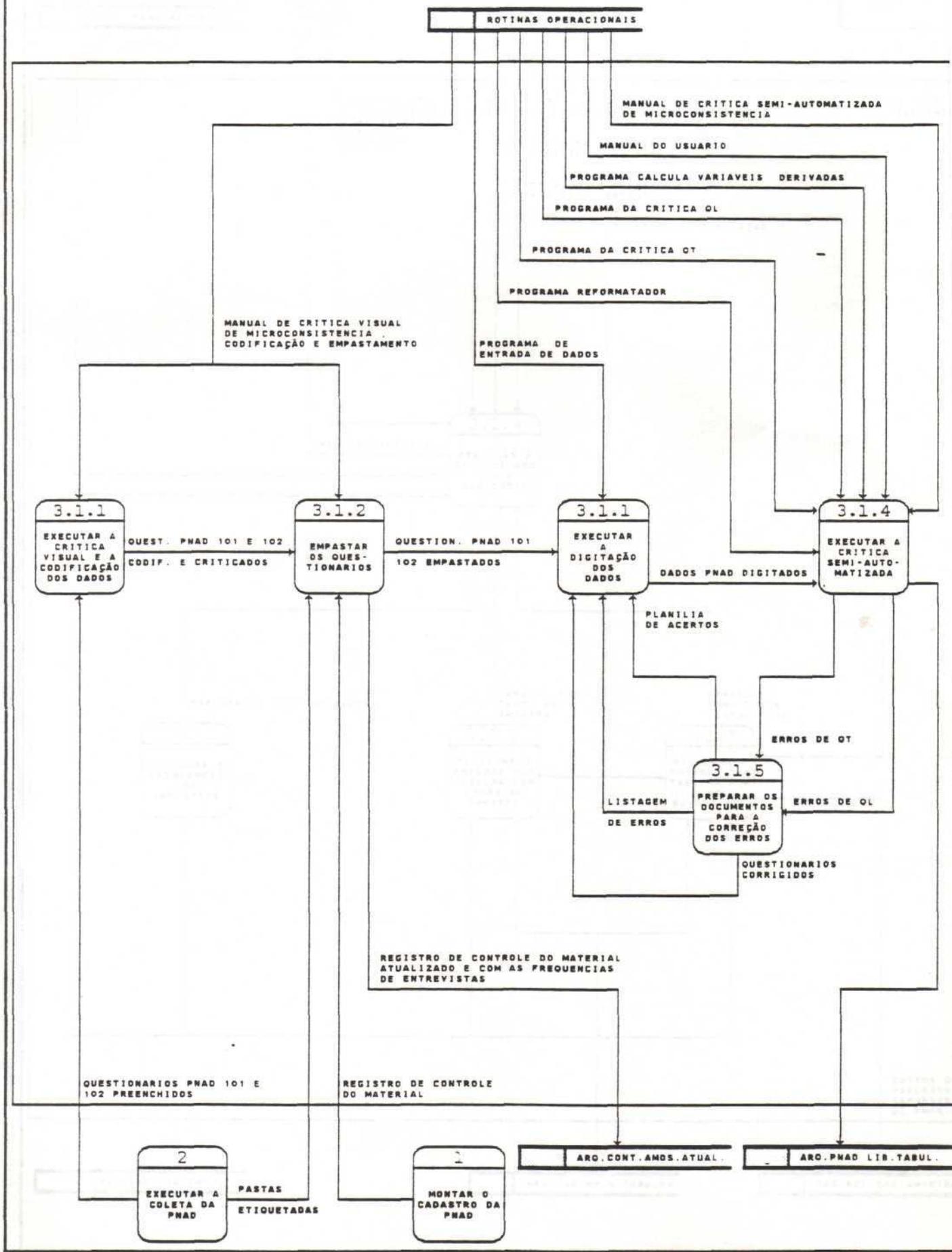
12 - Hanono, M. R; Barbosa, R. M. D. Uma Experiência de Desenvolvimento de Sistema Gerador de Aplicações Utilizando recursos de Engenharia de Software. Rio de Janeiro:IBGE, DI, Artigo Técnico Número 15, 1990.



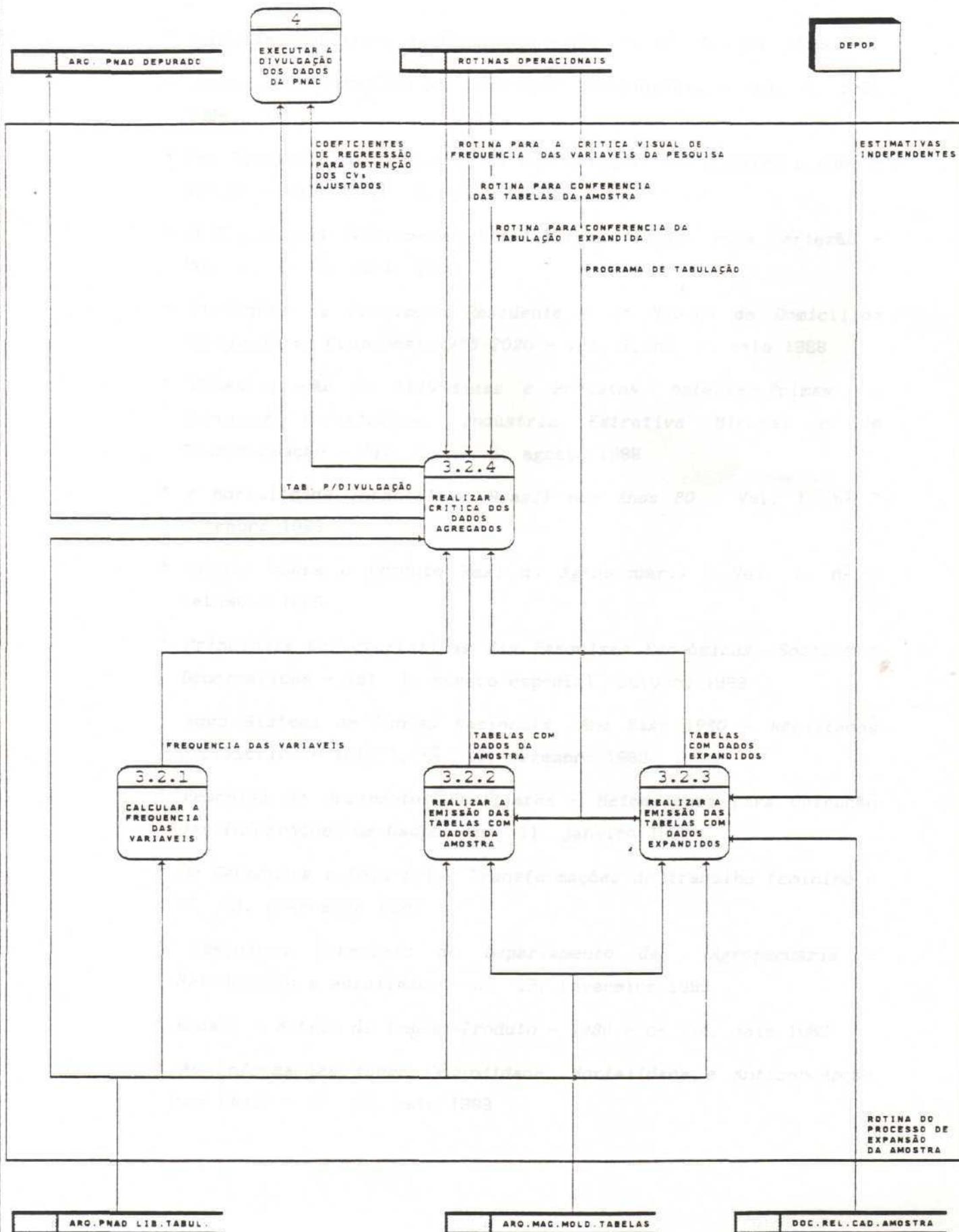
3 - EXECUTAR A APURAÇÃO DA PNAD



3.1 - EXECUTAR A CRITICA DE MICROCONSISTENCIA



3.2 - EXECUTAR A MACROCONSISTENCIA DOS DADOS



Textos para Discussão já publicados

- * *Pesquisas Contínuas da Indústria* - Vol. 1, nº 1, jan. 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 - n^o 16, junho 1989*
- * *Brasil - Sistema de Contas Nacionais Consolidadas - n^o 17, agosto 1989*
- * *Brasil - Produto Interno Bruto Real Trimestral - Metodologia - n^o 18, agosto 1989*
- * *Estatísticas e Indicadores Sociais para a Década de 90 - n^o 19, setembro 1989*
- * *Uma análise do Cotidiano da Pesquisa no DEREN (As Estatísticas do Trabalho) - n^o 20, outubro 1989*
- * *Coordenação Estatística Nacional - Reflexões sobre o caso brasileiro - n^o 21, novembro 1989*
- * *Pesquisa Industrial Anual 1982/84 - Análise dos Resultados - n^o 22, novembro 1989*
- * *O Departamento de Comércio e Serviços e a III Conferência Nacional de Estatística - n^o 23, dezembro 1989*
- * *Um projeto de Integração para as Estatísticas Industriais - n^o 24, dezembro 1989*
- * *Cadastro de Informantes de Pesquisas Econômicas - n^o 25, janeiro 1990*
- * *Ensaio sobre a Produção de Estatística - n^o 26, janeiro 1990*
- * *O Espaço das Pequenas Unidades Produtivas: Uma tentativa de integração - n^o 27, fevereiro 1990*
- * *Uma Nova Metodologia para Correção Automática no Censo Demográfico Brasileiro: Experimentação e primeiros resultados - n^o 28, fevereiro 1990*
- * *Notas Técnicas sobre o Planejamento de Testes e Pesquisas Experimentais - n^o 29, março 1990*
- * *Estatísticas, Estudos e Análises Demográficas - Uma visão do Departamento de População - n^o 30, abril 1990*
- * *Crítica de Equações de Fechamento de Empresas no Censo Econômico de 1985 - n^o 31, maio 1990*

- * *Efeito de Conglomeração da Malha Setorial do Censo Demográfico de 1980 - n^o 32, maio 1990*
- * *A Redução da Amostra e a Utilização de Duas Frações Amostrais no Censo Demográfico de 1990 - n^o 33, junho de 1990*
- * *Estudos e Pesquisas de Avaliação de Censos Demográficos - 1970 a 1990 - n^o 34, julho de 1990*
- * *A Influência da Migração no Mercado de Trabalho das Capitais do Centro-Oeste - 1980 - n^o 35, agosto de 1990*
- * *Pesquisas de Conjuntura: Discussão sobre Variáveis a Investigar - n^o 36, setembro de 1990*
- * *Um Modelo para Estimar o Nível e o Padrão da Fecundidade por Idade com Base em Parturições Observadas - n^o 37, outubro de 1990*
- * *A Estrutura Operacional de Uma Pesquisa por Amostra - n^o 38, novembro de 1990*
- * *Produção Agrícola, Agroindustrial e de Máquinas e Insumos Agrícolas no Anos 80: Novas Evidências Estatísticas - n^o 39, dezembro de 1990*
- * *A Inflação Medida pelo Índice de Preços ao Consumidor - n^o 40, janeiro de 1991*
- * *A Participação Política Eleitoral no Brasil - 1988, Análise Preliminar - n^o 41, fevereiro de 1991*
- * *Ensaio sobre Estatísticas do Setor Produtivo - n^o 42, março de 1991*
- * *A Produção Integrada de Estatística e as Contas Nacionais: Agenda para Formulação de um Novo Plano Geral de Informações Estatísticas e Geográficas - n^o 43, março de 1991*
- * *Matriz de Fluxos Migratórios Intermunicipais - Brasil - 1980 - n^o 44, abril de 1991*
- * *Fluxos Migratórios Intrametropolitanos - Brasil - 1970-1980 - n^o 45, abril de 1991*

- * *A Revisão da PNAD, A Questão Conceitual e Relatório das Contribuições* - n^o 46, maio de 1991
- * *A Dimensão Ambiental no Sistema de Contas Nacionais* - n^o 47, maio de 1991
- * *Estrutura das Contas Nacionais Brasileiras* - n^o 48, junho de 1991
- * *Mercado do Couro e Resultados da Pesquisa Anual do Couro* - n^o 49, junho de 1991
- * *As Estatísticas e o Meio Ambiente* - n^o 50, julho de 1991
- * *Novo Sistema de Contas Nacionais Séries Correntes : 1981-85 Metodologia, Resultados Provisórios e Avaliação do Projeto* - n^o 51, julho de 1991 (2 Volumes : Volume 1-Metodologia, Resultados Provisórios e Avaliação do Projeto; Volume 2-Tabelas)
- * *O Censo Industrial de 1985 — Balanço da Experiência* - n^o 52, agosto de 1991
- * *Análise da Inflação* - n^o 53, agosto de 1991
- * *Módulo II do Anteprojeto da Pesquisa Nacional por Amostra de Domicílios - A Questão Amostral* n^o 54, setembro de 1991
- * *A Força de Trabalho e os Setores de Atividade - Uma Análise da Região Metropolitana de São Paulo - 1986-1990* n^o 55, outubro de 1991