

# CONTROLE DE QUALIDADE DE DADOS ESTATÍSTICOS: O Levantamento Censitário de Unidades de Produção Agrícola<sup>1</sup>

---

Vera Lúcia Ferraz dos Santos Francisco<sup>2</sup>

Maria de Lourdes Sumiko Sueyoshi<sup>3</sup>

Francisco Alberto Pino<sup>4</sup>

Antonio José Torres<sup>5</sup>

## RESUMO

*Descrevem-se aqui os procedimentos para detecção de erros de resposta (ou falta de resposta parcial) e para imputação de dados no levantamento censitário de unidades de produção agrícola do Estado de São Paulo, em 1995/96. Utilizaram-se algoritmos computacionais para testar o conjunto de dados com intervalos tanto lógicos como históricos. A maioria dos erros detectados foram corrigidos pelo pessoal de levantamento de campo, mas alguns foram imputados com modelos não-estocásticos. Os procedimentos mostraram-se particularmente úteis sob as condições adversas existentes durante o trabalho do levantamento censitário.*

**Palavras-chave:** *detecção de erros de resposta, imputação de dados, controle de qualidade, levantamento censitário.*

## STATISTICAL DATA QUALITY CONTROL: The Agricultural Census Survey

## SUMMARY

*This study describes the procedures for response error (or partial non-response) detection and for data imputation in the 1995/96 agricultural census survey in the state of Sao Paulo, Brazil. Computational algorithms were used to test the data set with both logical and historical intervals. Most detected errors were corrected by the field survey staff, but some were imputed with non-stochastic models. The procedures showed to be particularly useful under the adverse conditions existing during the census survey work.*

**Key-words:** *response error detection, data imputation, quality control, census survey.*

---

<sup>1</sup>Os autores agradecem ao Pesquisador Científico Flavio Condé de Carvalho a leitura de versão preliminar do texto e as sugestões.

<sup>2</sup>Estatístico, Pesquisador Científico do Instituto de Economia Agrícola.

<sup>3</sup>Matemático, Pesquisador Científico do Instituto de Economia Agrícola.

<sup>4</sup>Engenheiro Agrônomo, Dr., Pesquisador Científico do Instituto de Economia Agrícola.

<sup>5</sup>Engenheiro Agrônomo, Assistente Agropecuário da Coordenadoria de Assistência Técnica Integral.

## 1 - INTRODUÇÃO

A partir de 1993, a Secretaria de Agricultura e Abastecimento do Estado de São Paulo (SAA) passou a fornecer dados sobre área cultivada, em nível de município, necessários ao cálculo do índice de participação percentual dos municípios do Estado de São Paulo no Imposto sobre Operações Relativas à Circulação de Mercadorias e sobre Prestações de Serviços de Transporte Interestadual e Intermunicipal e de Comunicação (ICMS)<sup>6</sup>. Os dados então existentes em cadastros e censos, como os da Fundação Instituto Brasileiro de Geografia e Estatística (IBGE), Instituto Nacional de Colonização e Reforma Agrária (INCRA) e do Imposto Territorial Rural (ITR), não atendiam àquela necessidade, tampouco os dados subjetivos então levantados pela própria SAA.

Para cumprimento dessa lei, foi criado grupo de trabalho<sup>7</sup> para organizar e coordenar a execução do projeto de Levantamento Censitário das Unidades de Produção Agrícola do Estado de São Paulo, que se tornou conhecido por Projeto LUPA. O projeto foi executado em 1995-96, em parceria, pelo Instituto de Economia Agrícola (IEA) e pela Coordenadoria de Assistência Técnica Integral (CATI), órgãos da SAA. De modo geral, constou de sete grandes fases: organização, levantamento de campo, digitação, depuração, processamento dos dados finais, armazenamento e divulgação. Para compreender melhor o desenvolvimento do Projeto, pode-se consultar o questionário do Projeto (Anexo 1 e Anexos 2 a 5) e os manuais de instruções (PINO et al., 1995, 1996a, 1996b, 1997a; FRANCISCO et al., 1996; PINO e FRANCISCO, 1997b).

A execução do levantamento de campo, a digitação e parte da depuração de dados couberam à CATI, sob responsabilidade das diretorias das Divisões Regionais Agrícolas (DIRAs)<sup>8</sup>. Já a

equipe do IEA responsabilizou-se por grande parte da depuração, processamento final e armazenamento eletrônico. Em todas as etapas houve intercâmbio constante entre as equipes dos dois órgãos.

O procedimento de depuração dos dados é o objeto principal do presente trabalho. Foi realizada, parte pela CATI e parte pelo IEA, na quarta fase das sete executadas no Projeto LUPA. Atribui-se a esse tratamento parcela considerável de qualidade, confiabilidade e fidedignidade dos dados. A coerência na depuração requereu, também, conhecimentos encadeados sobre agricultura, estatística e informática.

## 2 - OBJETIVOS

O objetivo geral deste trabalho é detalhar os procedimentos utilizados na depuração dos dados numéricos e alfanuméricos do Levantamento Censitário de Unidades de Produção Agropecuária.

Os objetivos específicos são apresentar: a) o tratamento estatístico das totalizações e testes utilizados para verificação de consistência dos dados e b) os procedimentos de imputação de dados.

## 3 - MATERIAL

Estabeleceram-se, por questões de sigilo das informações, dois tipos de arquivos independentes: um de dados numéricos (contendo os dados estatísticos propriamente ditos) e outro de dados alfanuméricos (contendo a identificação dos informantes), devidamente codificados (Anexos 1 a 5).

O total de dados brutos depurados foi da ordem de 300 Megabytes, correspondentes a 277.124 Unidades de Produção Agropecuária (UPAs), pertencentes aos 645 municípios do Estado de São Paulo. De modo geral, uma UPA correspondia a um imóvel rural (PINO, 1997, para detalhes).

O primeiro volume depurado, por município,

<sup>6</sup>Para atender à Lei Estadual 8.510, de 29/12/93. Para detalhes, ver PINO (1994).

<sup>7</sup>Resolução SAA-15, de 26/04/95.

<sup>8</sup>A partir de 16 de abril de 1997, as 14 Divisões Regionais Agrícolas foram substituídas por 40 Regionais, chamadas Escritórios

de Desenvolvimento Rural, pelo Decreto Estadual nº 41.718, que organizou o Sistema Estadual Integrado de Agricultura e Abastecimento.

foi o de dados numéricos referentes à utilização do solo, aspectos sócio-econômicos, benfeitorias e máquinas; e o segundo, foi o de alfanuméricos, dados cadastrais de propriedade e localização da UPA.

Os erros considerados foram os de resposta, cometidos durante o preenchimento dos questionários, podendo-se atribuir tanto ao entrevistador quanto ao entrevistado, além de eventuais erros de digitação (foi feita uma redigitação dos dados, o que diminuiu sensivelmente este último tipo de erro).

PINO et al. (1996c; 1996d; 1996e) definiram, *a priori*, os possíveis erros que poderiam ocorrer, com o objetivo de elaborar os programas da depuração dos dados. Dos 44 tipos de erros enumerados, 20 foram detectados e corrigidos nas duas primeiras fases da depuração realizadas nas DIRAs; os erros restantes foram detectados no IEA, nas demais fases da depuração e corrigidos nas DIRAs.

A transmissão de dados, via *modem*, foi realizada pelo sistema de rede *dial-up*, com pontos de conexão no IEA/São Paulo, CATI/Campinas e CATI/Araçatuba. Esse sistema está contido no programa *WINDOWS 95* e possibilita acessar informações compartilhadas em outro computador. Utilizaram-se, também, disquetes para a remessa de dados.

## 4 - MÉTODO

Chama-se **detecção de erros** (alguns chamam de crítica dos dados) o processo de rever os dados procurando por erros e **correção de erros** o processo de corrigi-los. Ao conjunto dos dois processos, detecção e correção, denomina-se **depuração dos dados**. O ato de qualificar um dado como errado e atribuir-lhe um novo valor chama-se **imputação de dados**.

### 4.1 - Detecção de Erros

Os métodos de detecção de erros consistem em técnicas *a priori* e técnicas *a posteriori* (ZAR-KOVICH, 1968; PINO, 1986). No Projeto LUPA, aplicaram-se duas técnicas *a posteriori* para o

controle de erros no levantamento: a) análise de consistência interna, ou seja, testes dentro do próprio conjunto de dados; b) análise de consistência externa, ou seja, comparação dos resultados obtidos com dados procedentes de fontes independentes. Para a consistência interna, os intervalos de tolerância foram definidos *a priori*, considerando-se dois tipos de intervalo, histórico e lógico, conforme PINO (1986). A consistência interna com intervalo lógico é baseada em limites estabelecidos pela própria lógica dos testes. Foram utilizados para verificar a validade de códigos e de respostas, associação de respostas em variáveis diferentes, somas e similares. A análise de consistência interna com intervalo histórico é baseada em limites estabelecidos por conhecimento prévio sobre o assunto. Foram utilizados em relações como lotação de pastagem, densidade de cultivo, bem como em variáveis do tipo tamanho de explorações animais e quantidades de máquinas, veículos, equipamentos e benfeitorias. Por outro lado, a consistência externa consiste em comparar valores obtidos no levantamento em questão com valores de outros levantamentos independentes. Este tipo de consistência foi utilizado para comparar a soma das áreas totais das UPAs com a área territorial do respectivo município obtida através do Instituto Geográfico e Cartográfico (IGC).

Em todos os testes aplicados aos dados do Projeto LUPA, considerou-se que a hipótese verdadeira era a exatidão dos dados, enquanto que a hipótese alternativa era utilizada com a finalidade de detecção de erros de dados.

### 4.2 - Correção dos Dados

Quando o resultado de um determinado teste indicava um possível erro, retornava-se ao questionário de levantamento para averiguação da informação ou consultavam-se os manuais de códigos, para verificar se se tratava mesmo de um erro. Quando a correção mostrava-se realmente necessária, em alguns casos ela era mais ou menos óbvia e imediata, mas, na maioria dos casos, o material voltou ao pessoal de campo, sendo às vezes necessária a confirmação de valores junto à UPA.

Outro procedimento de correção utilizado foi a imputação de dados.

#### 4.3 - Imputação de Dados

A imputação de dados, que tem sido muito utilizada nas áreas “*de agricultura, melhoramento animal, economia e dados censitários, iniciou-se na área de agricultura e foi proposta por ALLAN e WISHART (1930)*”, tendo mais tarde se desenvolvido “*também na área de agricultura, com YATES (1933)*” (citados por BARROSO, 1995). Os procedimentos de imputação de dados constituem uma das maneiras de tratar dados incompletos, principalmente para o caso de não-resposta parcial, para completar os dados ausentes ou substituir os dados rejeitados. Essa técnica consiste em prever os valores perdidos e então analisar os dados completados (BARROSO, 1995). Embora com muitas críticas, os procedimentos de imputação também podem ser aplicados ao tratamento da não-resposta total, conforme BANNISTER (1980), citado por BARROSO (1995).

Entende-se por método de imputação todo procedimento de previsão de valores individuais, isto é, por qualquer método que atribua um valor específico a cada dado ausente ou rejeitado. Por exemplo, se  $y_i$  indica a observação perdida da  $i$ -ésima unidade amostral, no caso, unidade censitária (UPA), em um método de imputação, ela pode ser descrita como função de variáveis auxiliares  $z$  como  $y_i = f(z) + e_i$ , onde  $e_i$  é um resíduo aleatório. Em estudos censitários, a imputação, além de uma possível técnica, é uma necessidade (BARROSO, 1995). PINO (1986) refere-se a tais técnicas como correção por modelos. Modelos de imputação também podem ser encontrados em FELLEGI e HOLT (1976).

#### 4.4 - Algoritmos

A lógica da construção de algoritmos pode ser encontrada em KNUTH (1975). Os programas computacionais utilizados foram especialmente desenvolvidos para o Projeto LUPA por membros do

Grupo de Trabalho de Coordenação do Projeto<sup>9</sup>.

### 5 - RESULTADOS E DISCUSSÃO

Os resultados do presente trabalho consistem num conjunto de testes para detecção de erros e num conjunto de modelos para imputação de dados, que são discutidos quanto à sua eficácia para a finalidade a que se destinam. Os algoritmos apresentados a seguir foram construídos com o objetivo de fornecer o embasamento para a elaboração de dezenas de programas computacionais utilizados na depuração de dados do Projeto LUPA.

#### 5.1 - Consistência Interna com Intervalo Lógico

Obtiveram-se 31 testes com intervalos lógicos, dos quais os 23 primeiros funcionaram a contento, enquanto que os sete finais indicaram problemas, no sentido de detectar erro em muitas situações em que os dados estavam corretos (nesses casos, os testes foram desconsiderados). Apresentam-se, inicialmente, dois testes que dizem respeito ao controle geral do serviço.

##### 5.1.1 - Operação não realizada

A detecção de um erro neste teste significa que uma das operações (digitação, verificação, detecção de erros ou correção de erros) não foi realizada. O programa utilizado em cada operação introduzia um marcador no arquivo, indicando que a operação tinha sido realizada. Se, por distração ou negligência, uma operação deixara de ser realizada, este fato era detectado por este teste e o arquivo era devolvido aos responsáveis para completar o serviço.

A forma matemática do teste (isto é, a situação esperada, que se caracteriza como correta) é a seguinte:

$$N_{ij} \in L$$

onde:

<sup>9</sup> Vera Lúcia Ferraz dos Santos Francisco, Antonio José Torres e Francisco Alberto Pino.

$N_{ij}$  = valor da  $i$ -ésima operação na  $j$ -ésima UPA;  
 $L$  = conjunto de códigos para operações realizadas.

Qualquer situação diferente da esperada caracteriza um possível erro.

#### 5.1.2 - Número de unidades digitadas

A detecção de um erro neste teste significa que o número de UPAs gravadas no disquete diferia do número de UPAs que a respectiva equipe informava ter levantado. Possivelmente tratava-se de erro na digitação ou na manipulação de questionários.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$\frac{\sum_j X_j}{Y_j} = 1$$

onde:

$X_j$  = número de UPAs calculadas no arquivo do  $j$ -ésimo município;

$Y_j$  = número de UPAs levantadas no campo no  $j$ -ésimo município.

A seguir, apresentam-se três testes que dizem respeito a erros na codificação, que representam falhas na identificação de cada registro.

#### 5.1.3 - Código da UPA duplicado

A detecção de um erro neste teste significa que, dentro de um mesmo município, duas ou mais UPAs foram identificadas com o mesmo código.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$N_i \neq N_k \quad i \neq k$$

onde:

$N_i$  = número de identificação da  $i$ -ésima UPA;

$N_k$  = número de identificação da  $k$ -ésima UPA, para  $i \neq k$ .

#### 5.1.4 - Código de variável duplicado

A detecção de um erro neste teste significa que, dentro de uma mesma UPA, duas ou mais variáveis (campos) foram identificadas com o mesmo

código.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$N_{ij} \neq N_{kj} \quad i \neq k$$

onde:

$N_{ij}$  = número de identificação da  $i$ -ésima posição na  $j$ -ésima UPA;

$N_{kj}$  = número de identificação da  $k$ -ésima posição na  $j$ -ésima UPA, para  $i \neq k$ .

#### 5.1.5 - Código de variável inválido

A detecção de um erro neste teste significa que uma variável (campo) foi identificada com código inválido.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$N_{ij} \in L$$

onde:

$N_{ij}$  = código da  $i$ -ésima posição na  $j$ -ésima UPA;

$L$  = conjunto de códigos válidos.

Os demais testes, apresentados a seguir, referem-se aos erros ocorridos durante o preenchimento do questionário no campo (devidos tanto ao entrevistador quanto ao entrevistado). Os testes foram aplicados em cada UPA de cada município e dizem respeito à consistência interna dos próprios dados (valores) levantados. Inicia-se a apresentação com dois testes referentes à localização da UPA.

#### 5.1.6 - Universal Transversa de Mercator (UTM) nula

As coordenadas UTM indicam a localização aproximada da UPA no mapa. A detecção de um erro neste teste significa que o entrevistador não preencheu esse campo, resultando em valor nulo, que é inválido.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$X_{101,j} \neq 0$$

onde:

$X_{101,j}$  = valor do código 101 (UTM) na  $j$ -ésima UPA.

### 5.1.7 - Distância da UPA à sede do município inválida

A detecção de um erro neste teste significa que a distância da UPA à sede do município informada é nula ou negativa, o que, em princípio, não deveria acontecer. Na verdade, as distâncias inferiores a 500m foram aproximadas para baixo, resultando em 0km. Logo, cada vez que um erro foi detectado, verificou-se a possibilidade de o dado estar correto, antes da correção.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$X_{102,j} \geq 0$$

onde:

$X_{102,j}$  = valor do código 102 (distância da UPA à sede do município) na j-ésima UPA.

Na sequência, apresentam-se dois testes sobre a área da UPA.

### 5.1.8 - Área total da UPA inválida

A detecção de um erro neste teste significa que a área total informada para a UPA não é positiva. É evidente que, se a UPA não tiver área, ela não existe.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$X_{110,j} > 0$$

onde:

$X_{110,j}$  = valor do código 110 (área total) na j-ésima UPA.

### 5.1.9 - Soma das áreas de ocupação do solo diferente da área total

A ocupação do solo refere-se a áreas com: culturas perenes, semi-perenes e anuais; pastagens cultivadas e não cultivadas; vegetação natural; reflorestamento; e áreas inaproveitável, inaproveitada e complementar. O teste verifica se a soma dessas áreas é diferente da área total da UPA.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$\frac{\sum_{n=111}^{120} X_{nj}}{X_{110,j}} = 1$$

onde:

$X_{nj}$  = área da ocupação do solo correspondente ao código  $n=111, \dots, 120$  (respectivamente, áreas de perenes, semi-perenes, anuais, pastagem cultivada, pastagem não cultivada, vegetação natural, reflorestamento, inaproveitável, inaproveitada e complementar) na j-ésima UPA;

$X_{110,j}$  = valor do código 110 (área total) na j-ésima UPA.

A seguir, apresentam-se oito testes relacionados ao quadro de perguntas sobre explorações vegetais (Anexo 1).

### 5.1.10 - Código de área homogênea inválido

Uma área homogênea foi definida como um conjunto de talhões com as mesmas características produtivas (os talhões podem não ser contíguos, mas, devem estar dentro do mesmo imóvel rural). Dentro de uma UPA, eles foram numerados sequencialmente 1, 2, etc. A detecção de um erro neste teste significa que o código da área homogênea não é positivo.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$N_{ij} > 0$$

onde:

$N_{ij}$  = i-ésima área homogênea na j-ésima UPA.

### 5.1.11 - Área de uma área homogênea inválida

A detecção de um erro neste teste significa que a área (ha) de uma área homogênea não é positiva.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$X_{ij} > 0$$

onde:

$X_{ij}$  = área da i-ésima área homogênea na j-ésima UPA.

### 5.1.12 - Código de cultura inválido

A detecção de um erro neste teste significa que uma cultura foi identificada com código inválido.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$N_{ij} \in L$$

onde:

$N_{ij}$  = código da cultura na i-ésima posição na j-ésima UPA;

$L$  = conjunto de códigos válidos para culturas.

### 5.1.13 - Data de plantio inválida

A detecção de um erro neste teste significa a inexistência de data de plantio ou data inválida.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$1 \leq X_{ij} \leq 12$$

e

$$Y_{ij} > 0$$

onde:

$X_{ij}$  = mês de plantio da i-ésima cultura na j-ésima UPA;

$Y_{ij}$  = ano de plantio da i-ésima cultura na j-ésima UPA.

### 5.1.14 - Data de colheita inválida

A detecção de um erro neste teste significa a inexistência de data de colheita ou data inválida.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$1 \leq X_{ij} \leq 12$$

onde:

$X_{ij}$  = data de colheita da i-ésima cultura na j-ésima UPA.

### 5.1.15 - Tipo de cultura inválido

A detecção de um erro neste teste significa resposta para tipo de cultura diferente dos valores possíveis (0, 1 ou 2).

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$0 \leq X_{ij} \leq 2$$

onde:

$X_{ij}$  = tipo da i-ésima cultura na j-ésima UPA.

### 5.1.16 - Resposta inválida para irrigação

A detecção de um erro neste teste significa resposta para irrigação diferente dos valores possíveis (0 ou 1).

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$0 \leq X_{ij} \leq 1$$

onde:

$X_{ij}$  = valor para irrigação da i-ésima cultura na j-ésima UPA.

### 5.1.17 - Ocupação do solo e áreas homogêneas

Cada área homogênea foi classificada como sendo de uma das seguintes categorias: cultura perene, cultura semi-perene, cultura anual ou reforestamento (PINO et al. 1997).

A detecção de um erro neste teste significa que a área de uma ou mais categorias de culturas no quadro da ocupação do solo (Anexo 1) difere da respectiva soma de áreas homogêneas no quadro de explorações vegetais (Anexo 1).

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$Y_{ij} = \sum_{k=1}^{n_{ij}} X_{ijk}$$

onde:

$Y_{ij}$  = área com ocupação do solo, no quadro 4,

da  $j$ -ésima categoria na  $i$ -ésima UPA (Anexo 1);  
 $X_{ijk}$  = área com ocupação da  $j$ -ésima categoria, no quadro 7, na  $i$ -ésima UPA, na  $k$ -ésima área homogênea (Anexo 1);

com

$j=1$  (perene), 2 (semi-perene), 3 (anual) e 4 (reflorestamento),

$n_{ij}$  = número de áreas homogêneas com ocupação da  $j$ -ésima categoria na  $i$ -ésima UPA.

Seguem-se dois testes sobre explorações animais.

#### 5.1.18 - Vacas em lactação sem bezerros

A existência de vacas em lactação implicaria existência de bezerro(a)s até 12 meses. A detecção de um erro neste teste significa, portanto, a existência de vacas em lactação sem bezerro(a)s.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$\sum_i X_{ij} > 0 \Rightarrow \sum_i Y_{ij} > 0$$

onde:

$X_{ij}$  = número de vacas em lactação da  $i$ -ésima categoria na  $j$ -ésima UPA;

$Y_{ij}$  = número de bezerro(a)s da  $i$ -ésima categoria na  $j$ -ésima UPA;

com

$i$  = correspondendo às categorias bovino de corte, bovino de uso misto, bovino de leite e bubalinos.

#### 5.1.19 - Sericicultura e cultura de amoreira

A detecção de um erro neste teste significa que existe a atividade de sericicultura, mas não há cultura da amoreira, ou vice-versa.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$X_{316,j} \cdot Y_{19,j} > 0, \text{ para } X_{316,j} + Y_{19,j} > 0$$

onde:

$X_{316,j}$  = valor do código 316 (gramas de larvas) na

$j$ -ésima UPA;

$Y_{19,j}$  = valor do código 19 (área da amoreira) na  $j$ -ésima UPA.

A seguir, apresenta-se um teste sobre mão-de-obra.

#### 5.1.20 - Inexistência de mão-de-obra na UPA

A detecção de um erro neste teste significa inexistência de qualquer tipo de mão-de-obra na UPA, o que pode ser um erro ou não.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$\sum_{i=160}^{162} X_{ij} > 0$$

onde:

$X_{ij}$  = resposta à pergunta de código  $i$  da  $j$ -ésima UPA, com  $i = 160, 161, 162$ , correspondendo, respectivamente, a número de familiares do proprietário que trabalham na UPA, número de trabalhadores permanentes e número máximo diário de trabalhadores temporários (Anexo 1).

Seguem-se dois testes sobre os indicadores.

#### 5.1.21 - Resposta inválida para os indicadores

A detecção de um erro neste teste significa resposta diferente dos valores possíveis (0 ou 1) nas questões de códigos 171 a 192, 194 e 197 (Anexo 1). Trata-se de perguntas que exigem respostas binárias do tipo sim/não.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$0 \leq X_{ij} \leq 1$$

onde:

$X_{ij}$  = valor da resposta à  $i$ -ésima questão na  $j$ -ésima UPA, com

$i=171, 172, \dots, 192, 194, 197$ .

#### 5.1.22 - Resposta inválida para o nível de instrução do proprietário

A detecção de um erro neste teste significa



resposta diferente dos valores possíveis (0, 1, 2, 3 ou 4) na questão de código 193 (Anexo 1), referente ao nível de instrução do proprietário da UPA.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$0 \leq X_{193,j} \leq 4$$

onde:

$X_{193,j}$  = valor da resposta à questão 193 na j-ésima UPA.

Segue-se um teste sobre casas de moradia.

### 5.1.23 - Casas de moradia

A detecção de um erro neste teste significa que o número de casas de moradia habitadas é maior do que o número total de casas de moradia.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$\frac{X_{511,j}}{Y_{512,j}} \leq 1 \quad \text{para} \quad Y_{512,j} > 0$$

onde:

$X_{511,j}$  = número de casas de moradia habitadas na j-ésima UPA;

$Y_{512,j}$  = número total de casas de moradia na j-ésima UPA.

A seguir, apresentam-se sete testes referentes ao cruzamento dos dados de benfeitorias com outros dados. Estes testes não funcionaram bem, tendo, em alguns casos, ocasionado novos erros durante a correção. Isso aconteceu porque todos eles referem-se a uma pequena possibilidade (e não à certeza) de erro.

### 5.1.24 - Sericicultura e barracão/sirgaria

A detecção de um erro neste teste significa que existe a atividade, mas não há barracão para bicho da seda/sirgaria, ou vice-versa.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$X_{316,j} \cdot Y_{507,j} > 0, \quad \text{para} \quad X_{316,j} + Y_{507,j} > 0$$

onde:

$X_{316,j}$  = valor da variável sericicultura na j-ésima UPA,

$Y_{507,j}$  = valor da variável barracão para bicho da seda/sirgaria na j-ésima UPA.

### 5.1.25 - Avicultura e barracão para granja

A detecção de um erro neste teste significa que existe a atividade, mas não há barracão para granja, ou vice-versa.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$\sum_{i=303}^{305} X_{ij} \cdot Y_{508,j} > 0, \quad \text{para} \quad \sum_{i=303}^{305} X_{ij} + Y_{508,j} > 0$$

onde:

$X_{ij}$  = número de cabeças de aves na j-ésima UPA, com  $i=303, 304, 305$ , para avicultura de corte, para ovos e ornamental, respectivamente.

$Y_{508,j}$  = valor da variável barracão para granja na j-ésima UPA.

### 5.1.26 - Cultura de café e terreiro ou depósito/tulha

A detecção de um erro neste teste significa que existe a cultura de café, mas não há terreiro nem depósito/tulha, ou vice-versa.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$\frac{Y_{514,j} + Y_{537,j}}{X_{41,j}} > 0, \quad \text{para} \quad X_{41,j} > 0$$

onde:

$X_{41,j}$  = área com café na j-ésima UPA;

$Y_{514,j}$  = número de depósito/tulha na j-ésima UPA;

$Y_{537,j}$  = área com terreiro na j-ésima UPA.

### 5.1.27 - Plasticultura e estufa

A detecção de um erro neste teste significa que existe utilização de plasticultura sem estufa, ou vice-versa.

A forma matemática do teste (isto é, a

situação esperada) é a seguinte:

$$\frac{Y_{518,j}}{X_{187,j}} > 0 \text{ para } X_{187,j} > 0$$

onde:

$X_{187,j}$  = indicador de plasticultura na j-ésima UPA  
(0=não utiliza; 1=utiliza rotineiramente);

$Y_{518,j}$  = área de estufa/plasticultura na j-ésima UPA.

### 5.1.28 - Fábrica de ração

A detecção de um erro neste teste significa que existe fábrica de ração na UPA, mas não se utiliza a ração para confinamento, ou criação de aves e ou de suínos, ou vice-versa.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$\frac{X_{189,j} + X_{303,j} + X_{304,j} + X_{305,j} + X_{317,j}}{X_{519,j}} > 0, \text{ para } X_{519,j} > 0$$

onde:

$X_{189,j}$  = indicador de confinamento de bovinos na j-ésima UPA (0=não utiliza; 1=utiliza);

$X_{303,j}$  = número de cabeças de aves para avicultura de corte na j-ésima UPA;

$X_{304,j}$  = número de cabeças de aves para avicultura para ovos na j-ésima UPA;

$X_{305,j}$  = número de cabeças de aves para avicultura ornamental na j-ésima UPA;

$X_{317,j}$  = plantel de suínos;

$X_{519,j}$  = número de fábricas de ração na j-ésima UPA.

### 5.1.29 - Equínos e haras/baia/cocheira

A detecção de um erro neste teste significa que existe a benfeitoria, mas não há criação de equínos.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$\frac{X_{310,j}}{X_{521,j}} > 0 \text{ para } X_{521,j} > 0$$

onde:

$X_{310,j}$  = número de cabeças de equínos na j-ésima UPA;

$X_{521,j}$  = número de haras/baia/cocheira na j-ésima UPA.

### 5.1.30 - Instalações e prática de confinamento

A detecção de um erro neste teste significa que existem instalações para confinamento, mas não se utiliza o confinamento de bovinos.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$\frac{X_{189,j}}{X_{523,j}} > 0, \text{ para } X_{523,j} > 0$$

onde:

$X_{189,j}$  = indicador de confinamento de bovinos na j-ésima UPA (0=não utiliza; 1=utiliza);

$X_{523,j}$  = número de instalações para confinamento na j-ésima UPA.

### 5.1.31 - Suinocultura e pocilga

A detecção de um erro neste teste significa que existem pocilgas, mas não há suinocultura.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$X_{317,j} \cdot X_{530,j} > 0, \text{ para } X_{317,j} + X_{530,j} > 0$$

onde:

$X_{317,j}$  = plantel de suínos na j-ésima UPA;

$X_{530,j}$  = número de pocilgas na j-ésima UPA.

## 5.2 - Consistência Interna com Intervalo Histórico

Obtiveram-se três grupos de testes com intervalos históricos, isto é, baseados em limites estabelecidos pelo conhecimento prévio do assunto.

### 5.2.1 - Rebanho e pastagem

A detecção de um erro neste teste significa que existe um número muito grande de bovídeos (bovinos e bubalinos) por unidade de área. Os limi-

tes considerados aceitáveis pelos especialistas consultados são de 0 a 5 cabeças por hectare.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$0 \leq \frac{\sum_{i=121}^{159} X_{ij}}{\sum_{k=114}^{115} Y_{kj}} \leq 5$$

onde:

$X_{ij}$  = número de cabeças do rebanho na categoria  $i=121, 122, \dots, 159$ , na  $j$ -ésima UPA;

$Y_{kj}$  = área de pastagens cultivada ( $k=114$ ) e natural ( $k=115$ ) na  $j$ -ésima UPA.

### 5.2.2 - Densidade de cultivo (adensamento)

Este teste foi aplicado a cada uma das culturas perenes. A detecção de um erro neste teste, para uma dada cultura, significa que existe um número muito grande (ou muito pequeno) de plantas (pés) por unidade de área. Os limites considerados aceitáveis foram fornecidos por especialistas, pelo pessoal de campo e pela literatura. Em geral, utilizaram-se limites bem amplos para evitar corrigir algum dado correto. Mesmo assim, cada erro detectado foi investigado para verificar se estava realmente incorreto.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$I_i \leq \frac{X_{ij}}{Y_{ij}} \leq S_i$$

onde:

$X_{ij}$  = número de plantas da  $i$ -ésima cultura na  $j$ -ésima UPA;

$Y_{ij}$  = área ocupada pela  $i$ -ésima cultura na  $j$ -ésima UPA;

$I_i$  = limite inferior para a densidade de cultivo da  $i$ -ésima cultura,

$S_i$  = limite superior para a densidade de cultivo da  $i$ -ésima cultura.

### 5.2.3 - Valores aceitáveis

Este teste foi aplicado aos dados sobre explorações animais (exceto bovídeos), bem como aos dados de máquinas, veículos, equipamentos, benfeitorias e instalações. Para esses itens foram estabelecidos, por especialistas, níveis de quantidade ou tamanho considerados aceitáveis. A aplicação desse tipo de teste deve ser cuidadosa, para não gerar ainda mais erros durante a correção. Eles permitiram detectar, entre outros, alterações na unidade da resposta. Por exemplo, solicitava-se o tamanho (em  $m^2$ ) de certas instalações, mas o informante respondia o número de instalações (ou vice-versa). A detecção de um erro neste teste, para uma dada variável, significa que seu valor é pequeno ou grande demais diante do conhecimento geralmente aceito sobre o assunto.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$I_i \leq X_{ij} \leq S_i$$

onde:

$X_{ij}$  = valor da  $i$ -ésima variável na  $j$ -ésima UPA;

$I_i$  = limite inferior para a  $i$ -ésima variável;

$S_i$  = limite superior para a  $i$ -ésima variável.

### 5.3 - Consistência Externa

Apenas um teste de consistência externa foi realizado, mesmo porque os dados de outros levantamentos confiáveis, como o censo agropecuário anterior do IBGE, eram já bastante antigos. Ele consistiu em comparar a soma das áreas totais das UPAs em cada município com a área territorial do município fornecida pelos órgãos oficiais. A detecção de um erro neste teste, para um dado município, significa que a área total das UPAs é pequena ou grande demais diante da área territorial do município. Os limites do intervalo foram arbitrados como de 10% a 95% da área territorial.

A forma matemática do teste (isto é, a situação esperada) é a seguinte:

$$0,1 \leq \frac{\sum_{j=1}^{n_i} X_{110,ij}}{A_i} \leq 0,95$$

onde:

$X_{110,ij}$  = área total da j-ésima UPA do i-ésimo município;

$A_i$  = área territorial do i-ésimo município;

$n_i$  = número de UPAs no i-ésimo município.

A maioria dos municípios não acusou problemas neste teste, mas mesmo para estes foi verificada se a relação calculada no teste era razoável. Por exemplo, para um município sabidamente agrícola, espera-se uma relação alta entre a soma das áreas das UPAs e a área territorial, enquanto que para uma metrópole industrial espera-se uma relação baixa.

Não se encontraram erros para baixo, mas, os erros para cima detectados contribuíram para a correção dos dados em diversos municípios. Também se verificou que os dados oficiais de área territorial apresentavam problemas em alguns poucos municípios, principalmente naqueles de criação recente ou que passaram por desmembramentos.

#### 5.4 - Imputação de Dados

A correção automática ou imputação de dados foi realizada norteando-se pelo desenho lógico do questionário de levantamento. O teste descrito na seção 5.1.17 foi o único que possibilitou esse tipo de correção, caracterizando-se, portanto, como caso de não-resposta parcial.

Quando o teste descrito na seção 5.1.17 acusa erro, faz-se um teste adicional, cuja forma matemática (isto é, a situação esperada) é a seguinte:

$$\sum_{j=1}^4 Y_{ij} = \sum_{j=1}^4 \sum_{k=1}^{n_{ij}} X_{ijk}$$

utilizando a mesma notação da seção 5.1.17.

Se este teste acusar erro, a correção deve ser como nos demais testes; caso contrário, é possível fazer a imputação. Neste caso,  $Y_{ij}$  indica a observação perdida da j-ésima categoria de ocupação do solo, na i-ésima unidade censitária (UPA), que pode ser descrita como função não-estocástica das va-

riáveis auxiliares  $X_{ijk}$ , com parâmetros  $\alpha$  e  $\beta$  e resíduo aleatório nulo  $e_{ij} = 0$ . Portanto, a forma matemática do modelo não-estocástico de imputação é a seguinte:

$$\begin{aligned} \hat{Y}_{ij} &= \alpha_j + \sum_{k=1}^{n_{ij}} \beta_{jk} X_{ijk} \\ &= \sum_{k=1}^{n_{ij}} X_{ijk} \end{aligned}$$

onde:

$\hat{Y}_{ij}$  = valor imputado (isto é, correção a ser usada no lugar de  $Y_{ij}$ ), para  $j=1,2,3,\dots,4$ .

#### 6 - CONSIDERAÇÕES FINAIS

Destaca-se, sobretudo, a relevância dos procedimentos estatísticos no equacionamento de problemas e na utilização adequada da informática como instrumental poderoso para o controle de qualidade de dados estatísticos obtidos mediante levantamento de campo. O planejamento dos trabalhos permitiu a construção de algoritmos para detecção de erros e imputação de dados, em sua maioria eficazes, tendo se mostrado especialmente úteis nas condições adversas em que se realizou o Projeto.

Convém salientar, embora não se exibam dados, mas como advertência para levantamentos futuros, que a minimização de custos no controle de qualidade dos dados está diretamente relacionada ao cuidado das três fases que a precedem: organização do levantamento, preenchimento de questionário no campo e digitação. Neste contexto, em que pese todo o planejamento inicial do Projeto LUPA, a experiência mostrou que a restrição imposta pela contenção de gastos com treinamento de pessoal (ainda que se dispusessem de vários manuais de instruções detalhados, principalmente o de coleta de dados) implicou acréscimos diretos nos custos, com despesas extras de telefone, fax, transmissão de dados, diárias, programação e processamento de dados, sobretudo acréscimos em horas despendidas na depuração, acarretando atraso considerável na divulgação dos resultados.

O principal resultado do presente trabalho foi possibilitar o tratamento estatístico adequado

de controle de qualidade nos dados do Projeto LUPA (PINO e FRANCISCO 1997b). Os resultados específicos deste trabalho foram tornar público a metodologia utilizada e disponibilizar todos os algoritmos utilizados na depuração e imputação do Projeto LUPA, possibilitando o aprimoramento em trabalhos futuros.

**LITERATURA CITADA**

ALLAN, F. E.; WISHART, J. A method of estimating yield of missing plot in field experimental work. **Journal of Agricultural Science**, v.20, p.399-406, 1930.

BANNISTER, J. Use and abuse of census editing and imputation. In: **The Asia and Pacific Census Forum**, 6. Honolulu: East-West Population Institute, 1980.

BARROSO, Lúcia P. **Imputação de dados em painéis para populações finitas**. São Paulo: USP/IME, 1995. 179p. (Tese de Doutorado em Estatística).

FELLEGI, I. P.; HOLT, D. A systematic approach to automatic edit and imputation. **Journal of the American Statistical Association**, Washington, v.71, n.353, p.17-35, Mar. 1976.

FRANCISCO, Vera L. F. S. et al. **Projeto LUPA**: manual de digitação de dados alfanuméricos. São Paulo: SAA, ago. 1996. 32p.

KNUTH, Donald E. **The art of computer programming**: fundamental algorithms. 2.ed. Reading: Addison-Wesley, 1975. v.1. 634p.

PINO, Francisco A. Detecção e correção de erros em levantamentos agrícolas. **Pesquisa Agropecuária Brasileira**, Brasília, v.21, n.9, p.979-985, 1986.

\_\_\_\_\_. Participação dos municípios paulistas no ICMS. **Informações Econômicas**, SP, v.24, n.5, p.9-21, 1994.

\_\_\_\_\_. (Org.) **Levantamento censitário de unidades de produção agrícola do Estado de São Paulo**. São Paulo: IEA/CATI/SAA, 1997. 4v.

PINO, Francisco A.; FRANCISCO, Vera L. F. S. **Projeto LUPA**: manual do livro de estatísticas. São Paulo: SAA, jul. 1997b. 41p.

PINO, Francisco A. et al. \_\_\_\_\_.: digitação de dados. São Paulo: SAA, abr. 1996b. 30p.

\_\_\_\_\_. \_\_\_\_\_.: diretrizes gerais. São Paulo: SAA, abr. 1996a. 31p.

\_\_\_\_\_. \_\_\_\_\_.: manual da segunda depuração de dados. São Paulo: SAA, set. 1996d. 12p.

\_\_\_\_\_. \_\_\_\_\_.: manual de depuração de dados. São Paulo: SAA, jul. 1996c. 27p.

\_\_\_\_\_. \_\_\_\_\_.: manual de instruções de campo. São Paulo: SAA, set. 1995. 58p.

\_\_\_\_\_. \_\_\_\_\_.: manual de organização de dados. São Paulo: SAA, out. 1996e. 19p.

\_\_\_\_\_. \_\_\_\_\_.: manual de organização de dados. São Paulo: SAA, fev. 1997a. 26p. (Versão 2).

YATES, F. The analysis of replicated experiments when the field results are incomplete. **Empire Journal of Experimental Agriculture**, v.1, p.129-142, 1933.

ZARKOVICH, S. S. **Calidad de los datos estadísticos**. Roma: FAO, 1968.

*Agricultura em São Paulo, SP, 45(1):33-58, 1998.*

Recebido em 05/01/98. Liberado para publicação em 10/03/98.

**CONTROLE DE QUALIDADE DE DADOS ESTATÍSTICOS:**  
**O levantamento Censitário de Unidades de Produção Agrícola**

**Anexo 1**

(continua)

<b>GOVERNO DO ESTADO DE SÃO PAULO</b>  <b>SECRETARIA DE AGRICULTURA E ABASTECIMENTO</b> <b>LEVANTAMENTO CENSITÁRIO DE UNIDADES DE PRODUÇÃO AGROPECUÁRIA</b> <b>PROJETO LUPA</b>	<b>1995</b>  Código: <input type="text"/> <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
---	--

Nome ou razão social do proprietário: _____	
Endereço para correspondência: _____	
Município: _____	CEP: _____ Tel.: (____) _____ Fax: (____) _____

Nome da UPA (imóvel): _____	
Localização: _____	
Município: _____	Setor: _____

<b>Q3- Localização da UPA</b> (p.14)	
UTM: (101) <input type="text"/> <input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> <input type="text"/>	Distância da UPA à sede do município (km) (102) <input type="text"/> <input type="text"/> <input type="text"/>

<b>Q4-Ocupação do solo (área em ha, com uma decimal):</b> (p.15)	
Total: (110) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	Pastagem natural: (114) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> Inaproveitada: (118) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
Cultura perene: (111) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	Pastagem cultivada: (115) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> Inaproveitável: (119) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
Cultura semi-perene: (112) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	Reflorestamento: (116) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> Complementar: (120) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
Cultura anual: (113) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	Vegetação natural: (117) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>

<b>Q5-Rebanho bovínico (número de cabeças):</b> (p.17)				
Categoria animal	Bovinos de corte	Bovinos de uso misto	Bovinos de leite	Bubalinos
Touro (>36 meses):	(121) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(131) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(141) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(151) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
Vaca em lactação (> 36 meses):	(122) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(132) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(142) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(152) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
Vaca seca (> 36 meses):	(123) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(133) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(143) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(153) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
Bezerro(a) (até 12 meses):	(124) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(134) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(144) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(154) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
Novilha (12 a 36 meses):	(125) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(135) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(145) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(155) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
Garrote/tourinho (12 a 24 meses):	(126) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(136) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(146) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(156) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
Novilho (24 a 36 meses):	(127) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(137) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(147) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(157) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
Boi magro (> 36 meses):	(128) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(138) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(148) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(158) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
Boi gordo (> 36 meses):	(129) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(139) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(149) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	(159) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>

<b>Q6-Outras explorações animais (código e quantidade):</b> (p.18)				
<input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	• <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	• <input type="text"/> <input type="text"/> <input type="text"/>
<input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	• <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	• <input type="text"/> <input type="text"/> <input type="text"/>
<input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	• <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	• <input type="text"/> <input type="text"/> <input type="text"/>

(continua)



(continua)

**Q8-Mão-de-obra:(p.25)**Número de familiares do proprietário que trabalham na UPA: **(160)** Número de trabalhadores permanentes: **(161)** Número máximo diário de trabalhadores temporários: **(162)** **Q9 - Máquinas, veículos e equipamentos (código e quantidade): (p.25)**

<input type="text"/>	<input type="text"/>	•	<input type="text"/>	<input type="text"/>	•	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	•	<input type="text"/>	<input type="text"/>	•	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	•	<input type="text"/>	<input type="text"/>	•	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	•	<input type="text"/>	<input type="text"/>	•	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	•	<input type="text"/>	<input type="text"/>	•	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

**Q10 - Benfeitorias e instalações (código e quantidade): (p. 26)**

<input type="text"/>	<input type="text"/>	•	<input type="text"/>	<input type="text"/>	•	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	•	<input type="text"/>	<input type="text"/>	•	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	•	<input type="text"/>	<input type="text"/>	•	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	•	<input type="text"/>	<input type="text"/>	•	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	•	<input type="text"/>	<input type="text"/>	•	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

**Q11-Indicadores (0=não utiliza; 1=utiliza rotineiramente): (p.26)**

É cooperado: <b>(171)</b> <input type="text"/>	Dispõe de energia elétrica: <b>(179)</b> <input type="text"/>	Plasticultura: <b>(187)</b> <input type="text"/>
	(para uso residencial)	
É associado: <b>(172)</b> <input type="text"/>	Dispõe de energia elétrica: <b>(180)</b> <input type="text"/>	Inseminação artificial: <b>(188)</b> <input type="text"/>
	(para uso na atividade agrícola)	
É sindicalizado: <b>(173)</b> <input type="text"/>	Usa computador na agropecuária: <b>(181)</b> <input type="text"/>	Confinamento de bovinos: <b>(189)</b> <input type="text"/>
Assistência técnica oficial: <b>(174)</b> <input type="text"/>	Análise de solo: <b>(182)</b> <input type="text"/>	Pastejo intensivo: <b>(190)</b> <input type="text"/>
Assistência técnica privada: <b>(175)</b> <input type="text"/>	Calagem: <b>(183)</b> <input type="text"/>	Mineralização: <b>(191)</b> <input type="text"/>
Tem comunicação telefônica: <b>(176)</b> <input type="text"/>	Adubação orgânica/verde: <b>(184)</b> <input type="text"/>	Vermifugação: <b>(192)</b> <input type="text"/>
Crédito rural: <b>(177)</b> <input type="text"/>	Práticas de conservação do solo: <b>(185)</b> <input type="text"/>	
Escrituração agrícola: <b>(178)</b> <input type="text"/>	Semente melhorada: <b>(186)</b> <input type="text"/>	

Nível de instrução do proprietário: **(193)** 

0=outros; 1=primário completo; 2=1º grau ou ginasial completo; 3=colegial completo; 4=superior completo

O proprietário reside no imóvel (0=não; 1=sim) **(194)** Número de arrendatários que cultivam no imóvel: **(195)** Número de assentados que cultivam no imóvel: **(196)** O imóvel estende-se por mais de um município (0=não; 1=sim) **(197)**

(conclusão)

**Rascunho**

Croqui da UPA, relação de confrontantes ou roteiro de acesso

Nome do entrevistador:		
Cargo/função:		
Data: ____/____/____	Tel.:(____)_____	Fax.:(____)_____
Revisor	Data da revisão	Data da correção
	____/____/____	____/____/____
	____/____/____	____/____/____
	____/____/____	____/____/____
	____/____/____	____/____/____
	____/____/____	____/____/____

## Anexo 2

## CÓDIGOS DE CULTURAS

(continua)

Código	Grupo <sup>1</sup>	Nome comum
1	P	Abacate
2	S	Abacaxi (ou ananás)
3	P	Abiu (ou caimito)
4	A	Abóbora Nota: inclui abóbora rasteira, abobrinha-italiana, abóbora-de-árvore, abóbora de moita, mogango, moranga (ou jerimum)
5	A	Açafrão Nota: inclui o açafrão-europeu (ou verdadeiro) e o açafrão-da-índia (ou da terra, ou curcuma)
6	A	Acelga
7	A	Agrião Nota: inclui agrião d'água e da terra (ou de terra enxuta)
8	A	Aipo Nota: aipo-rábano e salsão são variedades de raízes grossas e comestíveis
9	S	Alcachofra
10	A	Alface
11	S	Alfafa
12	A	Algodão
13	A	Alho
14	A	Alho-poró (ou porro)
15	A	Almeirão (ou chicória amarga)
16	A	Alpiste (ou alpista)
17	P	Ameixa Nota: inclui ameixa-européia e ameixa-japonesa
18	A	Amendoim
19	P	Amoreira Nota: para bicho-da-seda
20	P	Amora Nota: para fruto
21	P	Anona Inclui: fruta-de-conde (ou fruta-do-conde, ou pinha, ou ata, ou nona, ou anona), graviola (ou ata, ou condessa, ou pinha, ou araticum), cherimoia, atemoia
22	P	Antúrio Nota: de corte
23	A	Araruta
24	A	Arroz
25	P	Aspargo
26	A	Aveia
27	A	Azevém (ou erva-castelhana, ou joio-castelhano, ou relva)
28	P	Bambu

<sup>1</sup>A=cultura anual; S=cultura semi-perene; P=cultura perene; C=pastagem; F=florestal.

(continua)

Código	Grupo <sup>1</sup>	Nome comum
29	P	Banana
30	A	Bardana (ou <i>gobô</i> )
31	A	Batata-doce
32	A	Batata-inglesa (ou batata, ou batatinha)
33	P	Baunilha
34	A	Berinjela
35	A	Beterraba
36	A	Branquinha
37	C	Braquiária
38	A	Brócolos (ou brócolis)
39	A	Bucha
40	P	Cacau
41	P	Café
42	P	Caju
43	A	Camomila
44	S	Cana-de-açúcar
45	A	Colza
		Nota: inclui as variedades canola e níger
46	P	Capim-cidrão
47	C	Capim-colonião
48	C	Capim-gordura
49	C	Capim-jaraguá
50	C	Capim-napier (ou capim-elefante)
51	P	Caqui
52	A	Cará (ou acará)
53	P	Carambola
54	A	Catalonha
55	A	Cebola
56	A	Cebolinha
57	A	Cenoura
58	A	Centeio
59	P	Cereja (ou cereja-da-europa)
60	P	Cereja-das-antilhas (ou cereja-do-pará, ou acerola)
61	A	Cevada
62	P	Chá
63	A	Chicória (ou chicória de folha crespa)
64	A	Chuchu
65	P	Cidra
66	P	Coco-da-baía
67	A	Coentro
68	A	Cogumelo
69	S	Couve (ou couve-crespa)

<sup>1</sup>A=cultura anual; S=cultura semi-perene; P=cultura perene; C=pastagem; F=florestal.

(continua)

Código	Grupo <sup>1</sup>	Nome comum
70	A	Couve-de-bruxelas
71	A	Couve-chinesa (ou repolho-chinês)
72	A	Couve-flor
73	A	Couve-nabo (ou rutabaga)
74	A	Couve-rábano
75	A	Couve-tronchuda (ou couve-troncha, ou troncha)
76	A	Cravo
77	P	Cravo-da-índia Nota: condimento
78	A	Crisântemo
79	A	Crotalária
80	P	Damasco (ou abricó)
81	P	Dendê
82	P	Erva-mate
83	A	Ervilha Nota: inclui ervilha-torta, ervilha-seca, ervilha-verde
84	S	Erva-cidreira Nota: medicinal
85	A	Escarola (ou endívia, ou chicória de folha lisa)
86	A	Espinafre (ou espinafre-verdadeiro ou espinafre-europeu)
87	A	Espinafre-da-nova-zelândia
88	A	Estévia
89	F	Eucalipto
90	A	Fava
91	A	Feijão (ou feijão-de-cores)
92	A	Feijão-de-lima
93	A	Feijão-fradinho (ou feijão-frade, ou fradinho, ou feijão-europeu, ou feijão-asparago, ou feijão-de-corda, ou macaçar)
94	P	Figo (ou figo-da-europa)
95	P	Figo-da-índia
96	P	Framboesa
97	S	Funcho (ou erva-doce de cabeça, ou anis-doce)
98	A	Gengibre
99	A	Gergelim
100	A	Girassol
101	A	Gladíolo
102	P	Goiaba
103	A	Grão-de-bico
104	P	Groselheira
105	P	Guaraná
106	A	Guandu (ou andu ou feijão-guando)
107	S	Hortelã Nota: inclui menta e hortelã-pimenta
108	A	Inhame
109	P	Jabuticaba
110	P	Jaca

<sup>1</sup> A=cultura anual; S=cultura semi-perene; P=cultura perene; C=pastagem; F=florestal.

(continua)

Código	Grupo <sup>1</sup>	Nome comum
111	A	Jiló
112	P	Juta
113	P	Kinkan
114	P	Kiwi
115	P	Laranja Nota: inclui as variedades de laranja doce normais (pêra, hamlin, seleta, barão, valência, branca, rosa, lisa, china ou caipira, abacaxi, washington, rubi, maltesa, westin, serra-d'água, joão nunes, orvalho-de-mel, champagne, feijão-cru, mato grosso, doblefina, moro, mangaratiba, elisa, torocco, natal, sabará), de umbigo (baía, baianinha, thompson navel), sangüíneas (sangüínea, sanguinelli, doppio sanguigoso, dr. amaral) e sem ácidos (lima, piralima)
116	P	Laranja-azeda (ou laranja-amarga ou laranja-da-terra) Nota: inclui variedades agro-sevilhana, bergamota, amargo-doce, apepu ou agridoce
117	A	Lentilha
118	P	Lichia (ou lechia)
119	P	Lima Nota: inclui variedades doces (lima-da-pérsia, lima-de-umbigo, palestina, colúmbia, kusaie) e ácidas (mexicana, sutil), exceto falsos limões
120	P	Limão Nota: inclui os limões ( <i>C.limon</i> , variedades siciliano, eureka, lisboa, vila franca, gêno-va, rugoso, meier, âmbarou, ouro, feminello, interdonato, monachello, berna, mesero), os falsos limões ( <i>C.aurantifolia</i> , variedades limão-galego e thaiti) e o limão-cravo ( <i>C.reticulata</i> )
121	A	Linho
122	A	Lírio
123	S	Lúpulo
124	P	Maçã
125	P	Macadamia
126	S	Mamão
127	S	Mamona
128	S	Mandioca (ou aipim, ou macaxeira, ou maniva)
129	A	Mandioquinha (ou batata-baroa, ou mandioquinha-salsa, ou batata-salsa)
130	P	Manga
131	A	Mangarito
132	P	Mangustão (ou mangustão-amarelo, ou bacupari)
133	S	Manjerona
134	S	Maracujá
135	A	Margarida
136	P	Marmelo
137	A	Maxixe (ou pepino-das-antilhas)
138	A	Melancia
139	A	Melão
140	A	Milho

<sup>1</sup> A=cultura anual; S=cultura semi-perene; P=cultura perene; C=pastagem; F=florestal.

(continua)

Código	Grupo <sup>1</sup>	Nome comum
141	A	Morango
142	A	Mostarda
		Nota: inclui verdura e condimento
143	A	Mucuna
144	A	Nabiça
		Nota: híbrido de nabo e mostarda
145	A	Nabo (ou nabo redondo, doce)
146	P	Nectarina
147	P	Nêspera
148	P	Nogueira pecã
149	P	Oliveira (azeitona)
150	S	Orégano (ou orégão)
151	A	Painço
152	S	Palmito
153	A	Pepino
154	P	Pêra
155	P	Pêssego
156	A	Pimenta
157	P	Pimenta-do-reino
158	A	Pimentão
159	F	Pinhão
160	F	Pinus
161	P	Pitanga
162	S	Poejo
163	P	Pomelo (ou <i>grape-fruit</i> )
164	A	Quiabo
165	F	Quiri (ou kiri)
166	A	Rabanete
167	A	Rábano
168	P	Rami
169	A	Repolho
170	A	Repolho-crespo
171	P	Romã
172	P	Rosa
173	A	Rúcula (ou pinchão)
174	A	Ruibarbo
175	A	Salsa (ou salsinha)
176	S	Salvia
177	P	Seringueira
178	P	Sisal
179	A	Soja
180	A	Sorgo
181	S	Taioba

<sup>1</sup>A=cultura anual; S=cultura semi-perene; P=cultura perene; C=pastagem; F=florestal.

(conclusão)		
Código	Grupo <sup>1</sup>	Nome comum
182	P	Tâmara
183	P	Tamarindo
184	P	Tangerina Nota: inclui as variedades doces (mexerica, mexerica do rio, mexerica ipanema, clementina, poncã, dancy ou flórida, cravo, satsuma, montenegrina, cleópatra, imperador, onesco) e ácidas (limão cravo, rosa ou rangpur, calamondin)
185	P	Tangelo Nota: híbridos de tangerina e laranja-doce (murcote, temple, tangerona, kara, kinga)
186	P	Tangor Nota: híbrido de tangerina e pomelo
187	A	Tomate
188	S	Tomilho
189	P	Toranja
190	A	Trigo
191	A	Triticale Nota: híbrido de trigo e centeio
192	P	Tungue
193	P	Urucum (ou urucu)
194	P	Uva Nota: para comer e fabricar vinhos finos (itália, moscatel, etc.)
195	P	Uva Nota: para chupar e fabricar vinhos comuns (niagara, isabel, etc.)
196	A	Violeta africana
901		Outros capins
902		Outras culturas
903		Outras flores
904		Outras florestais
905		Outras olerícolas
906		Outras plantas medicinais

<sup>1</sup>A=cultura anual; S=cultura semi-perene; P=cultura perene; C=pastagem; F=florestal.



**Anexo 3****CÓDIGOS DE EXPLORAÇÕES ANIMAIS**

Tipo de exploração	Código	Quantidade solicitada
Apicultura	301	Número de colméias utilizadas na produção de mel de abelhas
Asininos e muares	302	Plantel total (número total de cabeças)
Avicultura de corte	303	Número de cabeças por ano
Avicultura para ovos	304	Plantel total de aves para produção de ovos
Avicultura ornamental, ou decorativa, ou exótica	305	Plantel total de aves
Caprinocultura	306	Plantel total de caprinos
Carcinocultura	307	Número de pós-larvas de crustáceos
Codornicultura	308	Plantel total
Cunicultura	309	Plantel total
Eqüinocultura	310	Plantel total de eqüinos
Helicicultura	311	Número de viveiros para helicicultura (refere-se a caracóis, principalmente o <i>escargot</i> ).
Minhocultura	312	Número de canteiros
Ovinocultura	313	Plantel total
Piscicultura	314	Área total de tanques (m <sup>2</sup> )
Ranicultura	315	Número de girinos
Sericicultura	316	Gramas de larvas
Suinocultura	317	Plantel total
Outra exploração animal	318	Plantel total ou similar

**Anexo 4****CÓDIGOS DE MÁQUINAS, VEÍCULOS E EQUIPAMENTOS**

Especificação	Código
Conjunto de fenação	401
Arado escarificador	402
Batedeira de cereais	403
Câmara fria	404
Caminhão, camioneta ou utilitário	405
Carregadeira de cana	406
Colhedeira acoplada	407
Colhedeira automotriz	408
Conjunto de irrigação pivot central	409
Conjunto de irrigação por gotejamento ou microaspersão	410
Conjunto de irrigação autopropelido	411
Conjunto de irrigação convencional	412
Desintegrador, picador, triturador	413
Ensiladeira	414
Microtrator	415
Misturador de ração	416
Ordenhadeira mecânica	417
Pulverizador tratorizado	418
Resfriador de leite, tanque expansão	419
Semeadeira/plantadeira para plantio direto	420
Subsolador	421
Terraceador	422
Trator de esteiras	423
Trator de pneus	424

## Anexo 5

## CÓDIGOS DE BENFEITORIAS E INSTALAÇÕES

Especificação	Código	Quantidade solicitada
Açude ou represa	501	unidade
Almoxarifado/oficina	502	unidade
Área de lazer	503	m <sup>2</sup>
Armazém para grãos ensacados	504	sacas
Balança para veículos	505	unidade
Balança para bovinos	506	unidade
Barracão para bicho da seda/sirgaria	507	unidade
Barracão para granja/avicultura	508	unidade
Barracão/galpão/garagem	509	unidade
Biodigestor	510	unidade
Casa de moradia (habitada)	511	unidade
Casa de moradia (total)	512	unidade
Curral/mangueira	513	unidade
Depósito/tulha	514	unidade
Engenho	515	unidade
Escola	516	unidade
Estábulo	517	unidade
Estufa/plasticultura	518	m <sup>2</sup>
Fábrica de ração	519	unidade
Fábrica de farinha	520	unidade
Haras/baia/cocheira para eqüinos	521	unidade
Igreja/capela	522	unidade
Instalação para confinamento	523	unidade
Laboratório para análise	524	unidade
Máquina de benefício	525	unidade
Moinho a vento	526	unidade
Olaria	527	unidade
<i>Packing house</i>	528	unidade
Pista de pouso	529	unidade
Pocilga	530	unidade
Poço semi-artesiano	531	unidade
Posto meteorológico	532	unidade
Roda d'água	533	unidade
Secador de grãos	534	unidade
Silo para silagem	535	t
Silo para grãos	536	t
Terreiro	537	m <sup>2</sup>
Usina/destilaria	538	unidade