

DETECÇÃO E CORREÇÃO DE ERROS EM LEVANTAMENTOS AGRÍCOLAS¹

FRANCISCO ALBERTO PINO²

RESUMO - Apresenta-se neste trabalho um procedimento prático e sua fundamentação estatística para detectar e corrigir erros que ocorrem durante o preenchimento de questionários em levantamentos de dados estatísticos. Embora tais erros ocorram em pequena proporção, seu efeito freqüentemente é desproporcional à sua freqüência. Sugere-se um modelo de correção, que é um caso de modelo de superpopulação associado ao estimador razão.

Termos para indexação: imputação, erros de resposta, modelo de superpopulação, estimador razão.

THE DETECTION AND CORRECTION OF ERRORS IN AGRICULTURAL SURVEYS

ABSTRACT - A practical procedure and the corresponding statistical aspects are presented in order to detect respondent errors. Although it occurs in a small proportion of answered questions, its effect is often disproportionate to its frequency. A model for correction is suggested which is shown to be a case of superpopulation model associated to the ratio estimator.

Index terms: imputation, respondent errors, superpopulation model, ratio estimator.

INTRODUÇÃO

Os erros de observação ou de resposta são inevitáveis nos levantamentos de dados estatísticos, principalmente nos países em desenvolvimento, onde os serviços estatísticos às vezes são precários, os entrevistadores qualificados são poucos e a população resiste a responder perguntas. Embora não se possa exigir correção absoluta, também não se pode trabalhar com dados de má qualidade, especialmente em pesquisa científica. Daí ser imprescindível a crítica em todas as fases do trabalho e a correção de todos os erros que possam comprometer a análise dos resultados. Entretanto, como a verificação de erros tem de ser cuidadosa para que não se introduzam erros ainda maiores e vieses, justifica-se a preocupação em estudar procedimentos que sejam, ao mesmo tempo, práticos e estatisticamente válidos.

Esforços têm sido feitos no sentido de minimizar o efeito dos erros. Um texto básico sobre o assunto é o de Zarkovich (1968). No Instituto de Economia Agrícola (IEA) vários trabalhos têm sido feitos para verificar e melhorar a qualidade

das estatísticas agrícolas do Estado de São Paulo. Sendin & Carmo (1970) utilizaram intervalos de confiança para verificar a qualidade de dados de preços recebidos pelos produtores. Pino & Jimenez Ossio (1975) descreveram um procedimento para a depuração de dados de produção agrícola utilizando computador. Tal procedimento foi testado em 1974, em levantamentos por amostragem de dados para previsão e estimativa de safras da Secretaria da Agricultura do Estado do Paraná. No mesmo ano foi adotado nos levantamentos por amostragem para previsão e estimativa de safras do Estado de São Paulo, sendo até hoje utilizado no IEA. A ocorrência de erros nestes levantamentos do IEA foi estudada por Pino & Caser (1984), que identificaram tipos e causas de erros.

O objetivo do presente trabalho foi o de desenvolver procedimentos para detectar e corrigir erros de resposta que ocorrem nos levantamentos de dados em geral, em particular nos levantamentos de dados agrícolas, a partir do método descrito em Pino & Jimenez Ossio (1975). O trabalho restringe-se aos aspectos teóricos da questão; avaliações numéricas são feitas em Pino & Caser (1984).

TERMINOLOGIA

Por razões didáticas, apresentam-se, a seguir, alguns conceitos e termos usados neste trabalho, a maioria deles familiares aos que trabalham com Estatística.

¹ Aceito para publicação em 28 de julho de 1986.

² Eng. - Agr., M.Sc., Instituto de Economia Agrícola, Caixa Postal 8114, CEP 01000 São Paulo, SP.

Chamamos de dados observados (ou valores observados, ou observações) os valores numéricos de uma variável aleatória obtidos para cada elemento de dada população ou de uma amostra dela. Chamamos de estatística uma função qualquer das observações de uma variável aleatória, que não contenha parâmetros desconhecidos, por exemplo, a média amostral, o total amostral, o desvio-padrão amostral.

Levantamento de dados é o ato de obter dados observados para os elementos de uma população. Diz-se que há um censo ou levantamento censitário quando todos os elementos da população são levantados, e levantamento por amostragem, quando apenas uma parte dos elementos da população são levantados.

Os dados obtidos num levantamento são chamados, às vezes, de dados primários (ou dados básicos, ou dados originais) e raramente podem ser considerados uma informação prontamente utilizável. É mais comum que se utilizem estatísticas calculadas a partir dos dados observados. Tais estatísticas são chamadas, às vezes, de dados elaborados ou dados secundários. Estatísticas calculadas a partir de dados secundários de diferentes levantamentos são chamadas, às vezes, de dados terciários. A rigor, dados observados só se transformam em informação estatística depois de passados por um processo de inferência estatística.

Num levantamento, os dados podem ser obtidos de duas formas: a) por medição; b) por entrevista. No levantamento por medição, usa-se um instrumento de medida (ou aparelho, ou equipamento) para obter diretamente os dados desejados; por exemplo, num levantamento topográfico. Num levantamento por entrevista, uma pessoa (chamada entrevistador ou enumerador) preenche um formulário (ou questionário), entrevistando outra pessoa (chamada informante). Eventualmente, o próprio informante pode preencher o questionário, dispensando o entrevistador. Os levantamentos por entrevista são comuns na obtenção de dados sócio-econômicos.

A expressão erro ou desvio estatístico refere-se à medida da diferença entre um valor verdadeiro da população e o valor da estatística usada para estimá-lo. Os erros estatísticos podem ser classificados em: a) erros amostrais; b) erros não-amostrais. Os erros amostrais devem-se à variabilidade natural na população e ao esquema amostral utilizado. Podem ser calculados (se a amostra for mensurável) e controlados dentro do próprio esquema amostral. Já os erros não-amostrais são os que não dependem do esquema amostral, podendo ocorrer até nos censos. Incluem vieses, erros de observação e outros. Geralmente, não podem ser calculados somente a partir dos dados observados, nem controlados pelo esquema amostral: são necessárias informações de fora do levantamento para seu cálculo e controle. Uma estatística diz-se acurada se o erro total (incluindo erros amostrais e não-amostrais) for pequeno; e diz-se precisa, se o erro amostral for pequeno.

Os erros de levantamento (ou de observação, ou de resposta) são erros não-amostrais que ocorrem durante a

execução do levantamento dos dados ou durante seu processamento. Incluem erros de medida, erros no preenchimento de questionários, erros de transcrição (inclusive de digitação, quando se usa processamento em computador) e similares.

Classificação dos erros de levantamento

Quanto ao causador, os erros podem ser: a) instrumentais; b) pessoais. Os instrumentais são causados por imprecisão ou por falhas do instrumento de medida (aparelho ou questionário). Os pessoais são causados pelas pessoas envolvidas no levantamento, principalmente entrevistador e informante.

Quanto à previsibilidade, podem ser: a) previsíveis; b) imprevisíveis. Os previsíveis podem e devem ser evitados. É o caso de erros devidos a entrevistadores não qualificados ou mal treinados.

Quanto à variabilidade, podem ser: a) sistemáticos; b) variáveis. Os sistemáticos ocorrem sempre; por exemplo, os erros devidos à má formulação das perguntas ou os erros induzidos pela parcialidade do entrevistador frente ao informante.

Quanto à determinabilidade, podem ser: a) detectáveis; b) não detectáveis. Alguns erros são facilmente detectáveis, enquanto outros não o são, por exemplo, se o informante fornecer um dado incorreto, mas, dentro do que se considerar razoável.

Quanto à corrigibilidade, podem ser: a) passíveis de correção; b) não passíveis de correção.

Causas dos erros de levantamento

As causas mais freqüentes de erros de levantamento centralizam-se: a) no questionário (ou instrumento de medida); b) no entrevistador (ou na pessoa que efetua a medida); c) no informante; d) na digitação.

Algumas das causas de erros que surgem no questionário são: erros conceituais nas perguntas, má formulação das perguntas (por exemplo, falta de clareza), e imperfeições em geral (Pino 1981). Quanto aos aparelhos de medida, as principais causas são os defeitos no aparelho e sua imprecisão.

No entrevistador (ou na pessoa que faz as medições) centram-se causas como: falta de qualificação para o serviço, falta de treinamento para o serviço, não entendimento da pergunta, interesse nas respostas, parcialidade frente ao informante, falta de atenção, negligência, má-vontade, má-fé, tendência pessoal e leitura errada de aparelhos.

Quanto ao informante, algumas das causas são: resistência a responder perguntas, desconhecimento da importância e da utilidade do levantamento, falta de registros, ignorância, não-entendimento da pergunta, interesse nas respostas, negligência, má-vontade, má-fé, tendência pessoal, desconfiança, vaidade.

Na preparação de arquivos de dados para processamento em computador, algumas das causas são: falta de qualificação, falta de treinamento para o serviço, falta de atenção e negligência.

Além de tudo isso, alguns erros podem se dever a causas geográficas (resultantes de diferenças regionais e de dificuldades de acesso e transporte), a causas institucionais (como a desordem nos serviços de algumas instituições) e a causas culturais.

Verificação da qualidade dos dados

Chamamos de detecção de erros o processo de rever os dados procurando por erros (ou possíveis erros). Chamamos de correção de erros o processo de correção de um erro detectado. Ao conjunto dos dois processos, detecção e correção, chamamos de depuração: diz-se depurar os dados de erros de levantamento. Também se diz imputação, referindo-se ao ato de qualificar um dado como errado e atribuir-lhe um novo valor.

Zarkovich (1968) classifica os métodos de detecção de erros em técnicas a priori e técnicas a posteriori.

Técnicas a priori

É a verificação por amostragem, que consiste em repasses na mesma área através de subamostras da amostra principal, levantadas por entrevistadores diferentes daqueles da primeira passagem. Também se podem incluir perguntas no questionário que sirvam de provas. Embora cara, esta técnica é particularmente útil quando se tem pouco conhecimento sobre o assunto levantado.

Técnicas a posteriori

São quatro técnicas:

1. comparação dos resultados obtidos com valores esperados ou previstos antes de se iniciar o levantamento;
2. comparação dos resultados obtidos com dados procedentes de fontes independentes;
3. estudo de consistência interna (testes dentro do próprio conjunto de dados levantados);
4. estudo de sobrevivência de grupos (em levantamentos periódicos).

O objetivo do estudo de consistência interna e da comparação com valores esperados é comparar os dados levantados com algum conhecimento geralmente aceito sobre o assunto levantado. Nos questionários com muitas perguntas sobre o mesmo assunto, aparecem boas oportunidades de estudos de consistência interna. Geralmente, a comparação com valores esperados é feita sobre as estatísticas ou dados elaborados, enquanto o estudo de consistência interna é feito diretamente sobre os dados observados, sendo esta a única técnica a posteriori que pode ser utilizada neste caso.

As técnicas a posteriori apresentam alguns inconvenientes, como a necessidade da existência de dados obtidos anteriormente sobre o mesmo assunto para comparação. Além disso, sua aplicação é difícil nos levantamentos em que as características mudam irregularmente em curtos períodos de tempo.

Correção

Uma vez detectado um possível erro, é preciso decidir se a correção será mesmo feita e como será feita. A correção deve ser cuidadosa, para não se causar erro mais grave que o original e para não se alterar a distribuição da variável que está sendo levantada. Bom senso é essencial, lembrando-se sempre de que corrigir o certo é mais grave do que não corrigir o errado.

Para Backstrom & Hursh (1963), quando forem detectados erros sistemáticos todos os questionários daquela área ou daquele entrevistador deverão ser verificados. Lembram, também, que a correção deve ser feita no campo, sendo permitida no escritório somente quando houver motivos fortes, como a existência de probabilidades irrefutáveis. Para evitar reincidência de erros nos levantamentos periódicos, deve-se chamar a atenção do entrevistador e a do informante para as falhas encontradas.

Ao serem detectados erros em vários testes de consistência, é necessário decidir quais dos campos (variáveis) envolvidos nos testes serão corrigidos. Para isso, Fellegi & Holt (1976) apresentam métodos que minimizam o número de campos a serem corrigidos (podendo a solução não ser única).

DETECÇÃO DE ERROS DE LEVANTAMENTO

Apresenta-se, nesta seção, um procedimento para detectar erros de levantamento pela técnica de consistência interna. O método baseia-se num resultado simples de probabilidade: seja X uma variável aleatória tal que $E|X| < \infty$; então, dado $0 < \epsilon \leq 1$, existem números reais a e b tais, que

$$P(x \notin [a, b]) < \epsilon \quad (1)$$

Isto significa que podemos construir um intervalo finito tal que seja muito pequena a probabilidade de essa variável ter valores fora desse intervalo. Então, podemos considerar que observações fora desse intervalo de tolerância sejam valores estranhos ("outliers") ou valores possivelmente errados.

Seja \underline{X} um vetor aleatório $m \times 1$, constituído das m variáveis do levantamento. Se f e g forem funções reais de \underline{X} (inclusive constantes), então, definimos o valor de um teste por

$$T(\underline{X}) = f(\underline{X}) / g(\underline{X}) \quad (2)$$

com $T(\underline{X}) \in [L_1, L_3]$, onde L_1 e L_3 são números reais, $L_1 < L_3$. Se f e g tiverem significado prático, chamaremos T de um teste de consistência interna.

O resultado (1) garante a existência de L_1 e L_3 , desde que tenhamos $E|T(\underline{X})| < \infty$, o que é equivalente a $E|f(\underline{X})| < \infty$ e $E|g(\underline{X})| \neq 0$.

Dado um teste de consistência interna, dizemos que há uma situação de erro (ou um possível erro) de consistên-

cia interna se ocorrer pelo menos uma das seguintes situações:

- Valor baixo, quando $T(\bar{X}) < L_i$ (menor que o limite inferior);
- Valor alto, quando $T(\bar{X}) > L_s$ (maior que o limite superior);
- Valor diferente do esperado, quando $L_i = L_s = L$ e $T \neq L$; esta situação pode ser considerada caso particular das duas primeiras;
- Valor esquecido, quando $g(\bar{X}) = 0$ e $f(\bar{X}) \neq 0$ (denominador nulo) ou quando $g(\bar{X}) \neq 0$ e $f(\bar{X}) = 0$ (numerador nulo).

Dizemos "possível erro" porque, como mostrado na Fig. 1, é sempre necessário decidir se a correção será feita ou não, isto é, quais as causas e se está mesmo errado ou não. Em geral, tal decisão cabe a um especialista no assunto levantado, mas, eventualmente, a decisão pode ser tomada por um computador e depois verificada por um especialista (é o que chamamos método automático de correção).

O teste apresentado em (2) pode ser utilizado na detecção de erros de levantamento. Praticamente, qualquer teste de consistência interna pode ser escrito como em (2). Além disso, as funções f e g costumam ser funções simples, em geral, lineares.

Determinação do intervalo de tolerância

O intervalo de tolerância de um teste deve ser definido *a priori*, antes de se aplicar o teste aos dados. Há três tipos de intervalos:

- intervalo lógico (quando os limites são estabelecidos pela própria lógica do teste);
- intervalo histórico (quando os limites são estabelecidos por um conhecimento prévio do assunto);
- intervalo calculado (quando for possível calcular um intervalo de confiança para a estatística, baseado nos dados levantados, uma vez que sua distribuição seja conhecida).

A experiência recomenda que os intervalos históricos sejam transigentes e elásticos. De fato, se o número de erros detectados for muito grande, só haverá duas explicações: ou o levantamento foi mal feito, ou o intervalo de tolerância do teste é excessivamente rígido.

A maneira usual de construir intervalos de confiança para estatísticas de variáveis com distribuição normal,

$$T \pm t_{\alpha} \cdot s(T) \quad (3)$$

onde $s(T)$ é o erro padrão de T e t_{α} é o valor da variável t de Student ao nível de significância α , pode resultar ruim se a distribuição da variável T afastar-se muito da normal. Neste caso, recomenda-se utilizar uma transformação apropriada antes de calcular intervalos de confiança, como a proposta por Box & Cox (1964):

$$X_t = \begin{cases} (X + k)^{\lambda}, & \text{se } \lambda \neq 0 \\ \log(X + k), & \text{se } \lambda = 0, \end{cases} \quad (4)$$

onde X é o dado original, X_t é o dado transformado, e k e λ são os parâmetros da transformação.

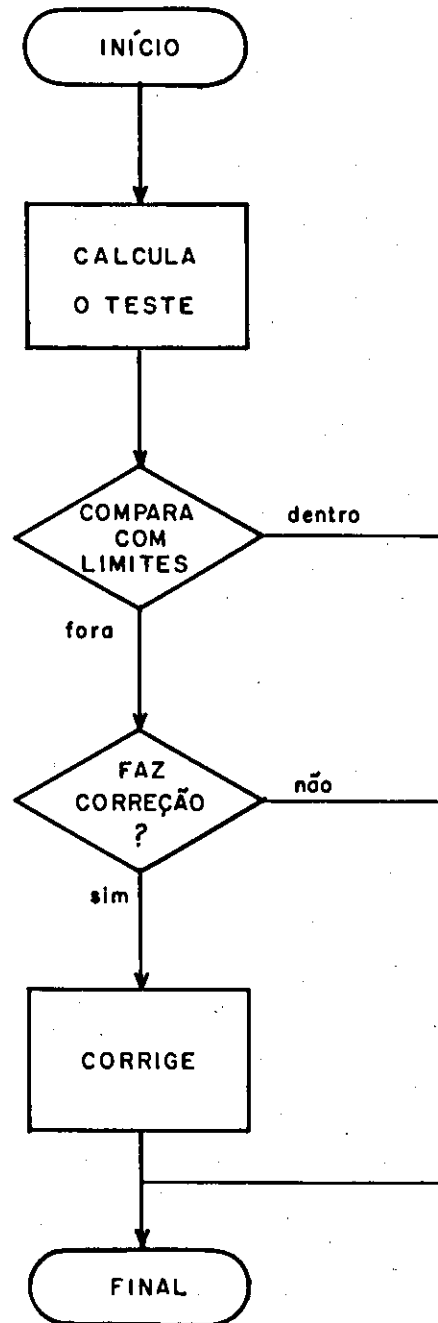


FIG. 1. Esquema de depuração de dados com teste de consistência interna.

CORREÇÃO DOS ERROS DETECTADOS

Apresenta-se, nesta seção, um procedimento para correção dos erros de levantamento detectados. A correção pode basear-se:

a) em critérios lógicos (por exemplo: se existem parcelas e falta o resultado da soma, a correção é óbvia);

b) em consulta ao campo, ou repasse, ou novo levantamento;

c) no histórico do informante (em levantamentos periódicos, a correção pode ser feita levando-se em conta dados do mesmo informante obtidos em levantamentos anteriores; este procedimento garante, inclusive, a sobrevivência de grupos);

d) nos outros dados do próprio levantamento (sob certas condições, um modelo calculado com os dados corretos pode ser utilizado na correção dos dados errados).

A ordem dos critérios apresentados acima é exatamente a ordem de preferência com que devem ser usados. De fato, o procedimento (d) só deve ser usado quando nenhum dos três anteriores for possível. Mesmo assim, conforme Fellegi & Holt (1976), ele deve ser utilizado de modo a modificar o menor número possível de itens (campos) do questionário.

Não há muito que comentar a respeito dos três primeiros procedimentos, uma vez que são naturais e auto-explicativos. Detenhamo-nos, portanto, no procedimento (d).

Seja uma amostra de n elementos e m variáveis e chamemos $A = \{1, \dots, n\}$ o conjunto de elementos da amostra. Analisemos o caso em que os valores das variáveis $X_1 = (X_1, \dots, X_h)$ estão corretos para todos os elementos de A , enquanto os valores das variáveis $X_2 = (X_{h+1}, \dots, X_m)$ podem estar errados para alguns elementos de A . Seja B o conjunto dos elementos da amostra A tal que os valores de X_2 estão corretos.

Então, a amostra A pode ser vista como uma população finita, e a subamostra B , como uma amostra de A . Os valores de X_1 correspondentes aos elementos de B , podem ser utilizados para estimar um modelo especificado. Depois, o modelo estimado aplicado aos valores de X_1 correspondentes aos elementos de $A-B$ pode ser utilizado para prever os valores de X_2 respectivos, isto é, o procedimento (d) pode ser considerado um caso de modelo de superpopulação (na verdade, associado ao estimador razão, como se poderá perceber adiante, como em Royall & Cumberland (1981).

Consideremos a definição (2) e analisemos dois casos de interesse.

Correção de $T(X)$ e $f(X)$

O primeiro caso é aquele em que $T(X)$ e $f(X)$ devem ser corrigidos, ou, equivalentemente, $T(X)$ e $g(X)$. Sejam:

$$E[f(X_i)] = T \cdot g(X_i) \quad e \quad (5)$$

$$e_i = f(X_i) - E[f(X_i)] \quad (6)$$

para $i \in A$. Então,

$$\hat{f}(X_i) = \hat{T} \cdot g(X_i) \quad (7)$$

Supondo-se que os erros e_i tenham variância constante e covariância nula, então, T poderá ser estimado ajustando-se uma reta de regressão linear simples de $f(X)$ contra $g(X)$. Se usarmos estimadores de mínimos quadrados, obteremos:

$$\hat{T} = \frac{\sum_{i \in B} f(X_i)}{\sum_{i \in B} g(X_i)} \quad (8)$$

Correção de X_2

O segundo caso é aquele em que $T(X)$ é conhecido e $f(X)$ só inclui elementos de X_2 .

Suponhamos que f seja linear, isto é, que exista um vetor de números reais a tal que $f(X) = a'X$ e que

$$a_j X_{ij} = c_{ij} T(X_i) g(X_i), \quad (9)$$

para $j = h+1, \dots, m, \quad i = 1, \dots, n$ e

$$\sum_j c_{ij} = 1.$$

Então,

$$g' X_i = \sum_j a_j X_{ij} = T(X_i) g(X_i). \quad (10)$$

Sejam

$$E(X_{ij}) = c_j T(X_i) g(X_i) / a_j \quad e \quad (11)$$

$$e_{ij} = X_{ij} - E(X_{ij}) \quad (12)$$

para $i \in A$ e $j = h+1, \dots, m$. Então,

$$\hat{X}_{ij} = \hat{c}_j T(X_i) g(X_i) / a_j. \quad (13)$$

Supondo-se que os erros e_{ij} tenham média nula, variância constante e covariância nula, então, os c_j poderão ser estimados ajustando-se retas de regressão linear simples de X_{ij} contra $T(X_i) g(X_i)/a_j$. Se usarmos estimadores de mínimos quadrados obteremos:

$$\hat{c}_j = a_j \frac{\sum_{i \in B} X_{ij}}{\sum_{i \in B} T(X_i) g(X_i)} \quad (14)$$

APLICAÇÃO EM LEVANTAMENTOS AGRÍCOLAS

Apresentam-se, a seguir, três exemplos típicos em levantamentos de produção agropecuária:

a) Produtividade

Sejam as variáveis

X(1): Área plantada com algodão,
X(2): Produção de algodão.

O teste será dado por $f(X) = X(2)$ e $g(X) = X(1)$, sendo $T(X)$ a produtividade ou rendimento agrícola. Neste caso, os limites podem ser históricos (produtividade normalmente obtida na região) ou calculados sobre os próprios dados ou sobre dados obtidos num levantamento anterior.

Se a produção estiver errada, a correção por modelo será dada por

$$\hat{X}(2) = \hat{T} X(1), \text{ com}$$

$$\hat{T} = \frac{\sum_{i \in B} X_i(2)}{\sum_{i \in B} X_i(1)},$$

isto é, \hat{T} será a produtividade média dos que responderam corretamente.

b) Peso de abate

Seja a variável

X(1): Peso médio de abate de bovinos.

O teste será dado por $f(X) = X(1)$ e $g(X) = 1$ (constante). Os limites podem ser históricos ou calculados.

A correção por modelo será dada por

$$\hat{X}(1) = \hat{T}, \text{ com}$$

$$\hat{T} = \bar{X}(1),$$

calculado sobre os elementos corretos, isto é, \hat{T} será o peso médio de abate dos que responderam corretamente. Casos como este, em que resultam correções pela média, podem distorcer as distribuições. Por isso, Fellegi & Holt (1976) recomendam que a correção seja aleatorizada, isto é, que um dos valores corretos seja sorteado para substituir o valor errado.

c) Soma

Sejam as variáveis

X(1): Área plantada sem financiamento,

X(2): Área plantada com financiamento,

X(3): Área plantada total.

O teste será dado por $f(X) = X(1) + X(2)$ e $g(X) = X(3)$. Os limites lógicos são $L_1 = L_3 = 1$.

Se as parcelas estiverem erradas, a correção por modelo será dada por

$$\hat{X}(j) = \hat{c}_j X(3), \text{ com}$$

$$\hat{c}_j = \frac{\sum_{i \in B} X_i(j)}{\sum_{i \in B} X_i(3)}, \quad j = 1, 2.$$

isto é, \hat{c}_j será a proporção média de $X(j)$ na soma.

OBSERVAÇÕES FINAIS

Programas de computador usando os procedimentos apresentados neste trabalho encontram-se descritos em Pino & Jimenez Ossio (1975). Os resultados obtidos no Instituto de Economia Agrícola têm sido considerados satisfatórios.

Finalmente, convém lembrar que o método aqui apresentado não consegue detectar erros de levantamento que não fujam do razoável. De fato, ele se presta para, numa primeira depuração, eliminar os erros grosseiros e aqueles devidos à falta de atenção do entrevistador. Por exemplo, no estudo de caso apresentado por Pino & Caser (1984), a quase-totalidade dos erros incluíam-se nessa classe. Para detectar erros que não fogem do razoável, é necessário utilizar outras técnicas, como o estudo de sobrevivência de grupos, os repasses, e a comparação com resultados de outras fontes. Entretanto, uma vez eliminados os erros grosseiros, os vieses resultantes de erros mais sutis serão certamente menores.

REFERÊNCIAS

- BACKSTROM, C.H. & HURSH, G.D. Survey research. Minneapolis, Northwestern Univ. Press, 1963.
- BOX, G.E.P. & COX, D.R. An analysis of transformations. *J. R. Stat. Soc. Ser. B.*, 26:211-52, 1964.
- FELLEGI, I.P. & HOLT, D. A systematic approach to automatic edit and imputation. *J. Am. Stat. Assoc.*, 71(353):17-35, 1976.
- PINO, F.A. *Elaboração de questionários para levantamentos de campo*. São Paulo, Inst. Econ. Agríc., 1981. (Boletim interno, 2)
- PINO, F.A. & CASER, D.N. *Análise de erros não amostrais em levantamentos para previsão e estimativa de safras do Estado de São Paulo*. São Paulo, Inst. Econ. Agríc., 1984. (Relatório de pesquisa, 10)
- PINO, F.A. & JIMENEZ OSSIO, J.H. *Um método para a depuração de erros não amostrais em dados obtidos por levantamento de campo*. São Paulo, Inst. Econ. Agríc., 1975. Trabalho apresentado na XIII Reunião da Sociedade Brasileira de Economia Rural, Curitiba, PR, jul. 1975.

ROYALL, R.M. & CUMBERLAND, W.G. An empirical study of the ratio estimator and estimators of its variance. *J. Am. Stat. Assoc.*, 76(373):66-77, 1981.

SENDIN, P.V. & CARMO, M.S. Análise da qualidade dos preços médios recebidos pelos produtores de milho

no Estado de São Paulo, 1969. *Agric. SP*, 17(7/8): 1-17, 1970.

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