

DIMENSIONAMENTO DE AMOSTRA PARA ESTIMATIVA E PREVISÃO DE SAFRA NO ESTADO DE SÃO PAULO (1)

Humberto de Campos (2)
Luiz Henrique de Oliveira Piva

O presente trabalho tem por finalidade determinar o tamanho da amostra para se proceder ao levantamento para estimativas e previsões de safras agrícolas no Estado de São Paulo.

Foi utilizado para tal fim o cadastro de propriedades rurais do INCRA, referente ao ano de 1972, com um total de 257.955 propriedades em todo o Estado, distribuídas por dez DIRAs (Divisões Regionais Agrícolas).

Em cada DIRA foram considerados doze estratos, de acordo com o tamanho das propriedades e abrangendo todas aquelas cuja área era acima de 3,0 hectares.

Os dados para o dimensionamento foram obtidos através de dois levantamentos realizados em janeiro e março de 1974, com uma amostra de 6.996 propriedades, distribuídas pelas DIRAs e estratos acima referidos.

$$\text{Critério 1: } d = 0,10 \bar{y}_{\text{est}}$$

$$\text{Critério 2: } d = 0,10 \bar{y}_h$$

$$\text{Critério 3: } d = 0,20 \bar{y}_h$$

onde:

d = semi-amplitude do intervalo de confiança da média, ao nível de 95% de probabilidade;

\bar{y}_{est} = média do Estado;

\bar{y}_h = média de cada DIRA h .

-
- (1) Relatório parcial do Projeto IEA/5 "Ampliação e Melhoria das Estatísticas Agrícolas", desenvolvido em caráter prioritário pelo Instituto de Economia Agrícola. Liberado para publicação em 19 de novembro de 1974.
- (2) Professor Adjunto do Departamento de Matemática da Escola Superior de Agricultura "Luiz de Queiroz"/USP.

Foram consideradas as oito culturas mais importantes do Estado, (algodão, amendoim, arroz, café, cana, feijão, milho e soja) e a amostra foi dimensionada pela partilha de Neyman para cada uma delas separadamente e através de 3 critérios distintos de precisão, ou seja:

Para o café, o dado considerado no dimensionamento foi o número de pés em produção, e para as demais culturas, a área plantada.

Considerando que se utilizou um cadastro único para todas as culturas, foram obtidas "amostras conciliatórias", tomando-se os resultados dos dois levantamentos isolada e conjuntamente. A conciliação constituiu em se determinar em cada Estrato e para cada DIRA a média dos tamanhos de amostra obtidos para cada uma das oito culturas.

Dentro das condições atuais de recursos materiais e humanos, embora não seja a ideal, mas a mais exequível, foi selecionada a composição da amostra conciliatória obtida através do critério 3, tomando-se conjuntamente os resultados obtidos pelos levantamentos de janeiro e de março, e que é constituída de 5.042 propriedades rurais.

1 - INTRODUÇÃO

Há muito tempo que a Secretaria da Agricultura do Estado de São Paulo, através de seus técnicos, preocupa-se com o uso da técnica da amostragem para obter estatísticas agrícolas e informações sócio-econômicas. Assim, no início da década de 50, os trabalhos, executados por SCHATTAN (2) e STEVENS (3) tornaram possíveis a implantação da técnica da amostragem para uso dos pesquisadores no campo da Economia Agrícola. Esse sistema, com poucas modificações, perdura até os dias atuais.

Até 1973 a meta desejada era obter estimativas para o Estado como um todo, com erros padrões que não ultrapassassem de $\pm 5\%$ do valor das referidas estimativas.

Devido ao aumento da demanda de estatísticas agrícolas para regiões específicas do Estado de São Paulo, surgiu a necessidade de se reformular os critérios de precisão procurando-se então, estimar as produções ao nível das Divisões Regionais Agrícolas (DIRAs), com um erro padrão não maior que $\pm 10\%$ — já que a amostra que

vinha sendo usada não cogitava da qualidade dos dados ao nível regional, mas somente ao nível do Estado. — As estimativas com esta precisão ao nível das DIRAs seriam suficientemente boas para as finalidades a que se destinam.

Norteando-se por esse critério, foi dimensionada uma amostra que teve como levantamento piloto as pesquisas feitas com a amostra antiga, de 2.282 elementos. O número de elementos de amostragem na nova amostra foi calculada em 6.996. Esta amostra permaneceu em uso por um ano, quando se pensou na sua reformulação; Nessa ocasião pretendeu-se fazer um trabalho mais elaborado e que fosse de alguma valia para outros pesquisadores que porventura esbarrassem com o problema de dimensionamento de amostra.

Essa melhoria da precisão dos levantamentos faz parte do projeto IEA/5, "Ampliação e Melhoria das Informações Agrícolas", projeto esse enquadrado no programa prioritário da Secretaria da Agricultura que objetiva alcançar o aperfeiçoamento do processo de tomada de decisões na agricultura

paulista. Sem informações e estatísticas acuradas dificilmente poderá ser alcançado esse objetivo. Daí a importância do projeto em questão sobretudo na atual fase de desenvolvimento do setor agrícola, cada vez mais sensível aos estímulos dos mercados de produtos e de fatores da produção.

1.1 - Objetivo do Trabalho

O presente trabalho tem por finalidade determinar tamanho de amostra a ser usada nas pesquisas de previsões e estimativas de safras agrícolas do Estado de São Paulo, de modo que as estimativas não apresentem erros padrões maiores que $\pm 10\%$ ao nível das DIRAs do Estado de São Paulo.

2 - MATERIAL E MÉTODOS

2.1 - Material

O rol utilizado no presente trabalho foi o do Instituto Nacional de Colonização e Reforma Agrária (INCRA), referente às propriedades rurais cadastradas no ano de 1972, cuja distribuição pode ser verificada no quadro 1.

Foram consideradas as dez Divisões Regionais Agrícolas (DIRA) do Estado, abrangendo as seguintes sub-regiões:

DIRA de São Paulo (2)

Capital
Bragança Paulista
Jundiaí
Registro
Santos
Mogi das Cruzes

DIRA do Vale do Paraíba (3)

São José dos Campos
Taubaté
Guaratinguetá

DIRA de Sorocaba (4)

Sorocaba
Tatuí
Itapetininga
Itapeva
Avaré
Botucatu

DIRA de Campinas (5)

Campinas
Piracicaba
Limeira
Rio Claro
São João da Boa Vista
Casa Branca

DIRA de Ribeirão Preto (6)

Ribeirão Preto
Franca
Orlândia
Barretos
Bebedouro
Araraquara
São Carlos
Taquaritinga

DIRA de Bauru (7)

Bauru
Lins
Jaú

QUADRO 1. - Distribuição das Propriedades Rurais do Estado de São Paulo, por DIRA e por Estrato, 1972

| DIRA | Estrato | | | | | | | | | | | | Total |
|-----------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|--------------|--------------|--------------|--------------|------------|----------------|
| | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | |
| São Paulo | 4.856 | 7.454 | 7.844 | 4.826 | 4.082 | 3.201 | 1.603 | 574 | 444 | 317 | 225 | 84 | 35.610 |
| Vale do Paraíba | 1.024 | 1.902 | 2.577 | 1.829 | 2.374 | 2.539 | 1.626 | 570 | 269 | 200 | 91 | 13 | 15.114 |
| Sorocaba | 4.114 | 7.656 | 9.866 | 6.370 | 6.349 | 5.532 | 3.056 | 1.122 | 865 | 623 | 323 | 104 | 45.979 |
| Campinas | 3.277 | 5.706 | 7.423 | 4.334 | 4.274 | 3.602 | 1.983 | 730 | 654 | 424 | 146 | 26 | 32.579 |
| Ribeirão Preto | 1.066 | 2.562 | 4.046 | 3.171 | 4.006 | 4.486 | 3.088 | 1.352 | 1.101 | 780 | 377 | 63 | 26.098 |
| Bauru | 581 | 1.131 | 2.085 | 1.587 | 1.170 | 1.859 | 1.304 | 558 | 463 | 367 | 172 | 32 | 11.859 |
| São José do Rio Preto | 1.208 | 3.043 | 7.085 | 5.141 | 5.410 | 4.900 | 2.737 | 983 | 706 | 431 | 206 | 25 | 31.875 |
| Araçatuba | 604 | 1.224 | 2.475 | 2.043 | 2.178 | 1.990 | 1.204 | 496 | 393 | 319 | 244 | 62 | 13.232 |
| Presidente Prudente | 1.501 | 3.312 | 6.067 | 3.991 | 3.458 | 2.537 | 1.392 | 584 | 418 | 338 | 324 | 99 | 24.021 |
| Marília | 919 | 2.186 | 4.676 | 3.649 | 3.512 | 3.148 | 1.802 | 578 | 531 | 372 | 192 | 23 | 21.588 |
| Total | 19.100 | 36.176 | 54.243 | 36.941 | 37.413 | 33.794 | 19.795 | 7.547 | 5.944 | 4.171 | 2.300 | 531 | 257.955 |

Fonte: Instituto Nacional de Colonização e Reforma Agrária (INCRA)

DIRA de São José do Rio Preto (8) Presidente Prudente
 Presidente Venceslau
 Dracena
 Osvaldo Cruz

São José do Rio Preto
 Catanduva
 Votuporanga
 Fernandópolis
 Mirassol

DIRA de Marília (11)

Marília
 Ourinhos
 Assis
 Tupã

DIRA de Araçatuba (9)

Araçatuba
 Andradina
 Penápolis

Em cada DIRA foram considerados
 doze estratos, de conformidade com
 o tamanho das propriedades.

DIRA de Presidente Prudente (10)

QUADRO 2. - Estratos, pelo Tamanho, das Propriedades Rurais do Estado de São Paulo,
 Utilizados nos Levantamentos para Estimativas e Previsões de Safras 1973 e 1974

| Estrato | Área (ha) |
|---------|-------------------|
| 03 | 3,1 a 5,0 |
| 04 | 5,1 a 10,0 |
| 05 | 10,1 a 20,0 |
| 06 | 20,1 a 30,0 |
| 07 | 30,1 a 50,0 |
| 08 | 50,1 a 100,0 |
| 09 | 100,1 a 200,0 |
| 10 | 200,1 a 300,0 |
| 11 | 300,1 a 500,00 |
| 12 | 500,1 a 1.000,0 |
| 13 | 1.000,1 a 3.000,0 |
| 14 | acima de 3.000,0 |

Fonte: Instituto de Economia Agrícola.

As propriedades rurais com área de até 3,0 hectares acham-se distribuídas nos estratos 0, 1 e 2, mas deixaram de ser consideradas por serem, em sua grande maioria, imóveis desprovidos de atividades agrícolas que não as hortícolas.

Para efeito de redimensionamento da amostra foram considerados os dados obtidos através de levantamentos realizados em janeiro e março de 1974 para as culturas relacionadas,

no quadro 3.

Nos casos do amendoim e do feijão foram consideradas as somas das áreas dos plantios das águas e da seca.

Os dados coletados para cada uma das culturas consideradas foram obtidos através de uma amostra única (geral para todas as culturas), de 6.996 propriedades, cuja distribuição numérica consta do quadro 4.

QUADRO 3. — Culturas e Levantamentos Considerados na Obtenção de Dados para o Redimensionamento da Amostra para Estimativas e Previsões de Safras do IEA

| Cultura | Levantamento de: | Variável considerada |
|---------------|------------------|------------------------|
| Café | Jan. e mar./74 | N.o de pés em produção |
| Arroz | Jan. e mar./74 | Área plantada |
| Algodão | Jan. e mar./74 | Área plantada |
| Milho | Jan. e mar./74 | Área plantada |
| Cana-forragem | Jan. e mar./74 | Área plantada |
| Amendoim | Jan. e mar./74 | Área plantada |
| Feijão | Jan./74 | Área plantada |
| Soja | Mar./74 | Área plantada |

QUADRO 4. - Número de Propriedades Rurais Utilizadas na Obtenção de Dados para o Redimensionamento da Amostra para Estimativas e Previsões de Safras, 1973/74

| DIRA | Estrato | | | | | | | | | | | | Total |
|-----------------------|---------|-----|-----|-----|-----|-------|-------|-----|-----|-----|-----|-----|-------|
| | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | |
| São Paulo | 8 | 18 | 38 | 40 | 54 | 72 | 74 | 48 | 26 | 12 | 6 | 4 | 400 |
| Vale do Paraíba | 16 | 32 | 52 | 44 | 64 | 74 | 58 | 24 | 18 | 12 | 6 | 2 | 402 |
| Sorocaba | 14 | 36 | 76 | 70 | 98 | 126 | 116 | 60 | 66 | 66 | 66 | 42 | 836 |
| Campinas | 16 | 40 | 82 | 70 | 94 | 122 | 100 | 52 | 68 | 68 | 44 | 12 | 768 |
| Ribeirão Preto | 10 | 24 | 70 | 76 | 126 | 204 | 224 | 136 | 154 | 166 | 158 | 44 | 1.392 |
| Bauru | 2 | 6 | 18 | 22 | 36 | 56 | 66 | 46 | 52 | 64 | 68 | 24 | 462 |
| São José do Rio Preto | 8 | 28 | 98 | 94 | 132 | 176 | 156 | 90 | 82 | 76 | 68 | 24 | 1.032 |
| Araçatuba | 4 | 10 | 32 | 36 | 46 | 58 | 54 | 30 | 30 | 42 | 46 | 26 | 414 |
| Presidente Prudente | 10 | 30 | 78 | 66 | 68 | 66 | 48 | 24 | 24 | 24 | 36 | 16 | 490 |
| Marília | 6 | 18 | 68 | 72 | 96 | 126 | 122 | 56 | 68 | 74 | 72 | 22 | 800 |
| Total | 94 | 242 | 612 | 590 | 814 | 1.082 | 1.018 | 566 | 588 | 604 | 570 | 216 | 6.996 |

Fonte: Instituto de Economia Agrícola.

2.2 - Métodos

2.2.1 - A partilha de Neyman

Por serem, neste caso, os custos fixos e não havendo limitação orçamentária, dentro do razoável, não se pensou na utilização da função de

custos, estando a determinação do tamanho da amostra baseada na "Partilha de Neyman", que, segundo COCHRAN (1), numa amostra estratificada, com L estratos, dimensiona o estrato h ($h = 1, 2, \dots, L$) através da fórmula:

$$n_h = n \frac{N_h S_h}{L \sum_{h=1}^L N_h S_h},$$

onde: n_h = Tamanho da amostra no estrato h,

n = Tamanho da amostra total,

N_h = População do estrato h,

S_h = Desvio padrão da população do estrato h.

COCHRAN (1) mostra também que o tamanho n da amostra é dado por:

$$n = \frac{\left(\sum_{h=1}^L W_h S_h \right)^2}{V}$$

onde:

$$W_h = \frac{N_h}{N},$$

$$N = \sum_{h=1}^L N_h$$

V = Variância pré-fixada

$$= \frac{d^2}{t^2}$$

sendo: d = Valor pré-fixado para a semi-amplitude do intervalo de confiança da média estratificada, a um nível $(1 - \alpha)$ de probabilidade,

t = Valor da tabela de t a um nível α de significância.

2.2.2 - Ajustes para o dimensionamento da amostra ao dimensionamento da amostra, considerou-se para o estrato j (j = 3, 4, ..., 14) e DIRA h (h = 2, 3, ..., 11)

A fim de se proceder, neste estudo, a seguinte notação:

N_{hj} = População do estrato,

n_{hj} = Número de propriedades amostradas no estrato.

S_{hj} = Estimativa do desvio padrão do estrato, obtida com os dados da amostra,

$N_h = \sum_{j=3}^{14} N_{hj}$ = População da DIRA h

$N = \sum_{h=2}^{11} N_h$ Total da População,

n_h = Tamanho da amostra na DIRA h

n = Tamanho total da amostra

Considerando-se que o cadastro utilizado foi único para todas as culturas em apreço, cada dimensionamento foi precedido das seguintes determinações: a) redução da população a nível de produtores — para isso, no estrato

j da DIRA h considerou-se

$$N'_{hj} = \frac{n'_{hj}}{n_{hj}} \cdot N_{hj} = \frac{N_{hj}}{n_{hj}} \cdot n'_{hj} = f_{hj} \cdot n'_{hj}$$

onde:

N'_{hj} = População de produtores no estrato.

N_{hj} = População do estrato.

n_{hj} = Dimensão do estrato na amostra.

n'_{hj} = Número de respostas obtidas no estrato.

f_{hj} = Fator de expansão do estrato.

consequentemente:

$$N'_h = \sum_{j=3}^{14} N'_{hj} = \text{População de produtores na DIRA } h,$$

$$n'_h = \sum_{j=3}^{14} n'_{hj} = \text{Números de produtores na amostra da DIRA } h,$$

$$N' = \sum_{h=2}^{11} N'_h = \text{Total da população de produtores,}$$

$$n' = \sum_{h=2}^{11} n'_h = \text{Tamanho total da amostra de produtores;}$$

b) determinação da média e da variância do estrato j na DIRA h – Estas duas medidas, em cada estrato,

foram calculadas, baseando-se na amostra de produtores, ou seja:

$$\bar{y}_{hj} = \frac{\sum_{i=1}^{n'_{hj}} y_{hji}}{n'_{hj}} = \frac{\hat{y}_{hj}}{n'_{hj}}$$

onde \hat{y}_{hj} é o total do estrato j da DIRA h, na amostra

$$s_{hj}^2 = \frac{\sum_{i=1}^{n'_{hj}} (y_{hji} - \bar{y}_{hj})^2}{n'_{hj} - 1}$$

- c) determinação da média e da variância da DIRA h — Estas medidas, onde se fizeram necessárias, foram obtidas conforme se segue:

$$\bar{y}_h = \frac{\sum_{j=3}^{14} N'_{hj} \bar{y}_{hj}}{N'_h} = \frac{\sum_{j=3}^{14} \hat{Y}_{hj}}{N'_h}$$

onde \hat{Y}_{hj} é a estimativa do total da população do estrato j na DIRA h

$$s_h^2 = \frac{\sum_{j=3}^{14} N'_{hj} s_{hj}^2}{N'_h}$$

Conforme se verifica, a variância, neste caso, é uma média ponderada das variâncias dos estratos;

cífico de produtores para cada cultura considerada.

- d) expansão da amostra — Em decorrência da redução da população a nível de produtores, os dimensionamentos foram feitos para uma amostra específica, isto é, admitindo-se a existência de um cadastro espe-

Assim sendo, pelo fato do cadastro utilizado ser geral para todas as culturas, em cada estrato torna-se necessário expandir a amostra, ou seja, ajustá-la para o cadastro geral.

Essa expansão, em cada estrato, é dada por:

$$n_{hj} = n'_{hj} \frac{N_{hj}}{N'_{hj}},$$

sendo:

n'_{hj} a dimensão obtida para o estrato

j da DIRA h, admitindo-se uma população específica de produtores;

n_{hj} a dimensão expandida do estrato j da DIRA h, admitindo-se que o cadastro é geral para todas as culturas; e

Portanto, no caso em estudo, uma vez obtido o dimensionamento da amostra, separadamente para cada uma das culturas, procede-se ao dimensionamento único, ou seja, a obtenção da amostra conciliatória pelo processo acima descrito.

e) amostra conciliatória – COCHRAN (1) afirma: “Uma vez que a melhor repartição para uma determinada especificação não o será, em geral, para uma outra, deve-se chegar a uma conciliação nos levantamentos com numerosas especificações”.

Tal conciliação, referida por esse autor, é conseguida, tomando-se em cada estrato a média dos dimensionamentos obtidos para cada especificação.

Afirma ainda: “A repartição conciliatória dá resultados quase tão precisos quanto os que seriam possíveis, usando-se repartições ótimas, separadas para cada especificação”.

2.2.3 – Dimensionamento da amostra

Conforme já visto em 2.2.1, a dimensão da amostra, no presente caso, é função da semi-amplitude d , pré-fixada, do intervalo de confiança da média estratificada.

Foram estabelecidos, para fins de dimensionamento da amostra, três critérios, ou seja:

Critério 1 – Admitindo-se o Estado como uma população única e, consequentemente, uma estratificação geográfica (DIRAs) e uma estratificação pelo tamanho das propriedades.

Neste caso considerou-se:

$$d = 0,10 \bar{y}_{est}$$

onde:

$$\bar{y}_{est} = \frac{\sum_{h=2}^{11} N_h \bar{y}_h}{N'}$$

assim, tem-se:

$$V = \frac{d^2}{t^2} = 0,0025 (\bar{y}_{est})^2$$

e, portanto:

$$n' = \left(\frac{\sum_{h=2}^{11} W_h s_h}{V} \right)^2$$

Observe-se que foi considerado $t = 2$, que corresponde aproximadamente ao valor da tabela ($\alpha = 0,05$) com infinitos graus de liberdade.

Considerou-se um mínimo de duas propriedades por estrato. Este mínimo foi expandido, tomando-se como fator de expansão:

$$f = \frac{\text{dimensão total calculada, da amostra expandida}}{\text{dimensão total calculada, da amostra não expandida}}$$

Critério 2 – Admitindo-se cada estratificada pelo tamanho das propriedades, como uma população distinta, priedades:

$$d_h = 0,10 \bar{y}_h$$

e então,

$$V_h = 0,0025 (\bar{y}_h)^2$$

e,

$$n'_h = \frac{\left(\sum_{j=3}^{14} W_{hj} s_{hj} \right)^2}{V_h}$$

onde

$$W_{hj} = \frac{N'_{hj}}{N'_h}$$

Observe-se que o dimensionamento neste caso, meramente obtido por: da amostra para todo o Estado é,

$$n' = \sum_{h=2}^{11} n'_h$$

Critério 3 – Admitindo-se, como uma população distinta, porém considerando-se: no caso anterior, cada DIRA como

$$d_h = 0,20 \bar{y}_h .$$

donde:

$$V_h = 0,01 (\bar{y}_h)^2$$

e,

$$n'_h = \frac{\left(\sum_{j=3}^{14} W_{hj} s_{hj} \right)^2}{V_h}$$

$$n' = \sum_{h=2}^{11} n'_h$$

Para os critérios 2 e 3 foi também considerado um mínimo de duas propriedades por estrato. O fator de expansão foi obtido tomando-se em cada DIRA h:

$$f_h = \frac{\text{dimensão total calculada, da amostra expandida}}{\text{dimensão total calculada, da amostra não expandida}}$$

Em casos onde a dimensão total calculada da amostra era zero tomou-se também na amostra expandida o mínimo de duas propriedades por estrato.

2.2.4 – Saída dos dados

De conformidade com a estrutura dada ao dimensionamento da amostra, foi possível, para cada cultura considerada, programar também a saída dos dados para posterior divulgação, como se segue:

a) número de produtores no estrato j da DIRA h:

– Conforme já visto,

$$N'_{hj} = f_{hj} \cdot n'_{hj}$$

e conseqüentemente:

$$N'_h = \sum_{j=3}^{14} N'_{hj}$$

$$N' = \sum_{h=2}^{11} N'_h \quad ;$$

b) estimativa da média do estrato j da DIRA h:

$$\bar{y}_{hj} = \frac{\hat{y}_{hj}}{n'_{hj}} \quad ;$$

c) estimativa do total da população do estrato j da DIRA h:

$$\hat{Y}_{hj} = N'_{hj} \bar{y}_{hj} = N'_{hj} \cdot \frac{\hat{y}_{hj}}{n'_{hj}} = f_{hj} \cdot \hat{y}_{hj} \quad ;$$

d) estimativa da média da DIRA h:

$$\bar{y}_h = \frac{\sum_j N'_{hj} \bar{y}_{hj}}{N'_h} = \frac{\sum_j \hat{Y}_{hj}}{N'_h} \quad ;$$

e) estimativa do total da população da DIRA h:

$$\hat{Y}_h = N'_h \bar{y}_h = N'_h \cdot \frac{\sum_j \hat{Y}_{hj}}{N'_h} = \sum_j \hat{Y}_{hj} \quad ;$$

f) estimativa da média do estrato j:

$$\bar{y}_j = \frac{\sum N'_{hj} \bar{y}_{hj}}{N'_j} = \frac{\sum \hat{Y}_{hj}}{N'_j} ;$$

g) estimativa do total da população do estrato j:

$$\hat{Y}_j = N'_j \bar{y}_j = N'_j \frac{\sum \hat{Y}_{hj}}{N'_j} = \sum \hat{Y}_{hj} ;$$

h) estimativa da média do Estado:

$$\bar{y} = \frac{\sum N'_h \bar{y}_h}{N'} = \frac{\sum \hat{Y}_h}{N'} = \frac{\hat{Y}}{N'} ;$$

i) estimativa do total da população do Estado:

$$\hat{Y} = N' \bar{y} = N' \frac{\sum \hat{Y}_h}{N'} = \sum \hat{Y}_h ;$$

j) estimativa da variância do estrato j da DIRA h:

$$s^2_{\bar{y}_{hj}} = \frac{\sum_i y^2_{hji} - \frac{\left(\sum_i y_{hji}\right)^2}{n_{hj}}}{n'_{hj} - 1} ;$$

k) estimativa da variância da estimativa da média do estrato j da DIRA h:

$$\hat{V}(\bar{y}_{hj}) = \frac{s^2_{\bar{y}_{hj}}}{n'_{hj}} ;$$

l) estimativa da variância da estimativa do total da população do estrato j da DIRA h:

$$\hat{V}(\hat{Y}_{hj}) = \hat{V}(N'_{hj} \bar{y}_{hj}) = (N'_{hj})^2 \hat{V}(\bar{y}_{hj}) ;$$

m) estimativa da variância da estimativa da média da DIRA h:

$$\hat{V}(\bar{y}_h) = \hat{V}\left(\frac{\sum_j N'_{hj} \bar{y}_{hj}}{N'_h}\right) = \sum_j \left(\frac{N'_{hj}}{N'_h}\right)^2 \hat{V}(\bar{y}_{hj}) ;$$

n) estimativa da variância da estimativa do total da população da DIRA h:

$$\begin{aligned} \hat{V}(\hat{Y}_h) &= \hat{V}(N'_h \bar{y}_h) = \hat{V}\left(N'_h \frac{\sum_j N'_{hj} \bar{y}_{hj}}{N'_h}\right) \\ &= \hat{V}\left(\sum_j \hat{Y}_{hj}\right) = \sum_j \hat{V}(\hat{Y}_{hj}) \end{aligned}$$

o) estimativa da variância da estimativa da média do estrato j:

$$\hat{V}(\bar{y}_j) = \hat{V}\left(\frac{\sum_h N'_{hj} \bar{y}_{hj}}{N'_j}\right) = \sum_h \left(\frac{N'_{hj}}{N'_j}\right)^2 \hat{V}(\bar{y}_{hj}) ;$$

p) estimativa da variância da estimativa do total da população do estrato j:

$$\hat{V}(\hat{Y}_j) = \hat{V}(N'_j \bar{y}_j) = \hat{V}\left(N'_j \frac{\sum_h N'_{hj} \bar{y}_{hj}}{N'_j}\right)$$

$$= \hat{V}\left(\sum_h \hat{Y}_{hj}\right) = \sum_h \hat{V}(\hat{Y}_{hj}) \quad ;$$

q) estimativa da variância da estimativa da média do Estado:

$$\hat{V}(\bar{y}) = \hat{V}\left(\frac{\sum_h N'_h \bar{y}_h}{N'}\right) = \hat{V}\left(\frac{\sum_h \hat{Y}_h}{N'}\right) = \frac{1}{(N')^2} \hat{V}\left(\sum_h \hat{Y}_h\right)$$

$$= \frac{1}{(N')^2} \sum_h \hat{V}(\hat{Y}_h) \quad ;$$

r) estimativa da variância da estimativa do total da população do estado:

$$\hat{V}(\hat{Y}) = \hat{V}(N' \bar{y}) = (N')^2 \hat{V}(\bar{y}) = \sum_h \hat{V}(\hat{Y}_h) \quad ; \text{ e}$$

s) intervalos de confiança (95%)

Da média do estrato j da DIRA h:

$$\bar{y}_{hj} \pm t \ s(\bar{y}_{hj})$$

onde t é o valor da tabela para $\alpha = 0,05$, com $n'_{hj} - 2$ graus de liberdade,

Do total da população do estrato j da DIRA h:

$$\hat{Y}_{hj} \pm 2 \ s(\hat{Y}_{hj}) \quad ,$$

tomando-se $t = 2$,

- Da média da DIRA h:

$$\bar{y}_h \pm 2 s(\bar{y}_h) ;$$

- Do total da população da DIRA h:

$$\hat{Y}_h \pm 2 s(\hat{Y}_h) ;$$

- Da média do estrato j:

$$\bar{y}_j \pm 2 s(\bar{y}_j) ;$$

- Do total da população do estrato j:

$$\hat{Y}_j \pm 2 s(\hat{Y}_j) ;$$

- Da média do Estado:

$$\bar{y} \pm 2 s(\bar{y}) ; e$$

- Do total da população do Estado:

$$\hat{Y} \pm 2 s(\hat{Y})$$

Com essas medidas pode-se organizar modelos de saída de dados.

3 - RESULTADOS E DISCUSSÃO

cultura, a redução da população (cadastro geral) a nível de produtores, conforme item a de 2.2.2.

3.1 - População de Produtores

De acordo com o número de respostas obtidas em cada estrato e para cada DIRA, procedeu-se para cada

Em termos de totais, os resultados obtidos foram os que se observam no quadro 5.

QUADRO 5. – Propriedades Produtoras no Estado de São Paulo, para Diferentes Culturas, Estimadas Através de Levantamentos por Amostragem

| Cultura | Levantamento | |
|-----------------|--------------|---------|
| | Jan./74 | Mar./74 |
| Café | 61.877 | 65.934 |
| Arroz | 76.882 | 78.910 |
| Algodão | 24.700 | 25.358 |
| Milho | 121.107 | 121.763 |
| Cana-forrageira | 37.569 | 36.527 |
| Amendoim | 19.413 | 20.075 |
| Feijão | 44.139 | – |
| Soja | | 6.898 |

As pequenas variações observadas entre os resultados de janeiro e de março se devem ao fato de que são estimativas independentes e obtidas com um número de respostas de questionários, que variou de um levantamento para outro.

Conforme se observa no quadro 5, o milho é a cultura mais difundida no Estado, sendo produzido em aproximadamente 50% das 257.955 propriedades rurais consideradas neste estudo.

A soja, embora ainda a menos difundida (aproximadamente 3%), foi incluída neste trabalho, devido a sua grande e crescente importância atual.

3.2 – Amostras Específicas para cada Cultura

Considerando-se as populações estimadas de produtores, para cada cultura foram dimensionadas as amostras específicas a cada uma delas. Admitiu-se, pois, a existência de cadastros específicos, numericamente inferidos do cadastro geral (comum para todas as culturas).

Os tamanhos das amostras específicas, obtidos pelos três critérios e nos levantamentos de janeiro e março de 1974, constam de quadros disponíveis no IEA.

Evidentemente, na hipótese de se dispor de um cadastro específico para cada cultura e condições materiais e humanas para se proceder a todos os levantamentos periódicos, aqueles seriam os tamanhos ideais das amostras.

Dar-se-ia, como é natural, preferência ao critério 2 ($d = 0,10 \bar{y}_h$) por ser, dentre os três, o que apresenta maior precisão nas estimativas obtidas. Observe-se que, por este critério, em nível de DIRA, tanto as estimativas de médias como as de totais são esperadas com um erro padrão da ordem de 5% do valor considerado.

Os critérios 1 ($d = 0,10 \bar{y}_{est}$) e 3 ($d = 0,20 \bar{y}_h$) praticamente se equivalem, conforme se pode constatar pelos próprios resultados obtidos nos dimensionamentos. Pelo critério 1 é de se esperar um erro-padrão para a média (total) do Estado, em tomo de 5% da mesma. Pelo critério 3, espera-se, em nível de DIRA, um

erro-padrão da média (total), em torno de 10% da mesma, devendo no entanto, este valor sofrer uma redução, quando referido à média (total) do Estado. Através dos dados estimados pôde-se organizar o quadro 6:

QUADRO 6. - Número de Propriedades nas Amostras Específicas, 1974

| Cultura | Critério | | |
|-----------------|----------|-------|-------|
| | 1 | 2 | 3 |
| Café | 1.224 | 3.900 | 1.060 |
| Arroz | 1.850 | 5.357 | 1.405 |
| Algodão | 1.553 | 2.793 | 793 |
| Milho | 1.636 | 4.897 | 1.289 |
| Cana-forrageira | 2.042 | 5.289 | 1.394 |
| Amendoim | 1.320 | 2.396 | .656 |
| Feijão | 1.103 | 3.819 | 1.034 |
| Soja | 721 | 1.229 | 452 |

Observa-se por esse quadro que o menor tamanho de amostra ocorreu, pelos três critérios, para a cultura de soja (721; 1.229 e 452 propriedades, respectivamente pelos critérios 1, 2 e 3).

O maior tamanho ocorreu para a cana de forragem no critério 1, com 2.042 propriedades, e para o arroz, nos critérios 2 e 3, com respectivamente 5.357 e 1.405 propriedades.

Observa-se ainda que, com exceção da cana, os resultados obtidos através dos levantamentos de janeiro e de março foram muito concordantes. No caso da cana, houve uma queda bem acentuada no tamanho da amostra no levantamento de março. Talvez isso se deva à maior precisão das informações obtidas, através daquele levantamento.

3.3 - Amostras Expandidas

Uma vez que o cadastro de propriedades utilizados nos levantamentos foi único para todas as culturas, fez-se a expansão das amostras específicas, conforme ítem d de 2.2.2, cujos resultados constam de 42 quadros à disposição, no I E A.

Obviamente, tratando-se meramente de uma expansão numérica das amostras específicas, sem alterar as suas estruturas, os comentários feitos em 3.2 se estendem a este caso.

Observe-se ainda, no caso do critério 2, que é o mais rigoroso, a necessidade, na maioria dos casos, de um tamanho relativamente grande das amostras, o que o torna, nas condições usuais de trabalho, quase impraticável.

3.4 - Amostra Conciliatória

Conforme o exposto em e) de 2.2.2, com base nos resultados obtidos em 3.3, foram determinadas as amostras

QUADRO 7. – Tamanho da Amostra Conciliatória, Referente ao Dimensionamento Obtido Através do Levantamento de Janeiro e Março de 1974

| DIRA | E s t r a t o | | | | | | | | | | | | Total |
|-----------------------|---------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-------|
| | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | |
| São Paulo | 10 | 16 | 17 | 14 | 36 | 39 | 31 | 21 | 18 | 26 | 10 | 10 | 248 |
| Vale do Paraíba | 13 | 16 | 40 | 48 | 49 | 76 | 151 | 59 | 51 | 16 | 13 | 11 | 543 |
| Sorocaba | 14 | 18 | 41 | 32 | 71 | 109 | 83 | 49 | 73 | 49 | 41 | 31 | 611 |
| Campinas | 18 | 164 | 45 | 33 | 95 | 58 | 65 | 60 | 79 | 65 | 37 | 17 | 736 |
| Ribeirão Preto | 10 | 11 | 16 | 32 | 69 | 93 | 73 | 51 | 44 | 84 | 36 | 16 | 535 |
| Bauru | 14 | 14 | 19 | 65 | 31 | 75 | 64 | 26 | 45 | 60 | 54 | 16 | 483 |
| São José do Rio Preto | 19 | 19 | 40 | 44 | 124 | 118 | 77 | 62 | 35 | 40 | 42 | 12 | 632 |
| Araçatuba | 12 | 13 | 18 | 27 | 41 | 53 | 39 | 64 | 19 | 31 | 40 | 21 | 378 |
| Presidente Prudente | 12 | 22 | 55 | 38 | 42 | 60 | 40 | 20 | 18 | 25 | 16 | 21 | 369 |
| Marília | 11 | 11 | 31 | 43 | 65 | 82 | 75 | 38 | 50 | 41 | 48 | 12 | 507 |
| Total | 133 | 304 | 322 | 376 | 623 | 763 | 698 | 450 | 432 | 437 | 337 | 167 | 5042 |

conciliatórias, tomando-se em cada estrato j e para cada DIRA h , a média dos n_{hj} relativos a cada cultura considerada (quadro 7).

Foram calculados os tamanhos das amostras conciliatórias para cada critério separadamente e, também considerando isolada e conjuntamente os levantamentos de janeiro e de março.

Como nos casos anteriores, os quadros não publicados estão a disposição para consulta no I E A.

Observa-se pelo simples exame dos dados que houve, para os três critérios, uma boa concordância dos resultados obtidos através dos dois levantamentos, considerados isoladamente e em conjunto.

3.5 - Considerações Finais

As condições ideais para se proceder aos levantamentos por amostragem seriam alcançadas se se dispusesse de um cadastro específico para cada cultura. Isso resultaria numa redução drástica do tamanho das amostras e um aumento na precisão das estimati-

vas obtidas. Mas, por outro lado, implicaria também na necessidade de se proceder a levantamentos individuais para cada cultura, uma vez que os cadastros e conseqüentemente a composição das amostras seriam distintos.

Os atuais recursos materiais e humanos disponíveis para essa natureza de trabalho constituem ainda um grande obstáculo à sua exequibilidade.

Lançando-se mão das amostras conciliatórias e dentro dos recursos atuais disponíveis é recomendável a adoção do critério 3 ($d = 0,20 \bar{y}_h$) quando se consideram conjuntamente os levantamentos de janeiro e de março, ou seja, a composição da amostra apresentada no quadro 7, constituída de 5.042 propriedades rurais.

Entretanto, em condições mais favoráveis, poder-se-ia adotar o critério 2 ($d = 0,10 \bar{y}_h$) que apresenta um erro-padrão esperado, das estimativas, de 5% de seus valores. Para este critério, o número de propriedades na amostra é de 16.587.

Acrescente-se também que, como o critério 1 praticamente não difere do 3, poder-se-ia também adotar o primeiro, com 5.487 propriedades.

THE DETERMINATION OF THE SAMPLE SIZE FOR ESTIMATING HARVESTS IN THE STATE OF SÃO PAULO

SUMMARY

The objective of this paper is to determine the size of a sample to be used in forecasting crops in the State of São Paulo. As previous calculations indicated a sample size smaller than had previously been budgeted, no consideration was given to sampling costs.

The frame used was a list of farm owners used by Federal Government (INCRA) for taxes purposes. This list contains 257,955 farm owners distributed in 10 Regional Divisions (DIRAs). In each DIRA 12 strata were considered.

The data for the determination of the sample size came from two prior surveys done in January and March, 1974. The sample size for those surveys was of 6,996 farms. The crops covered were coffee, rice, cotton, corn, sugar cane, peanut, bean and soybeans. The Neyman partition was used for each individual crop, according to the following three criteria:

$$\text{Criterion 1: } d = 0,10 \bar{y}_{est}$$

$$\text{Criterion 2: } d = 0,10 \bar{y}_h$$

$$\text{Criterion 3: } d = 0,20 \bar{y}_h$$

where

d = half width of the confidence interval of the mean, at the 95% level of probability.

\bar{y}_{est} = the State mean.

\bar{y}_h = the mean for the DIRA h .

For coffee, the data considered were the number of bearing trees and for other crops the cultivated area.

Since only a single sample will be used for all of the crops, alternative allocations were obtained by taking the average of the sample sizes for each one of the 8 crops in each of the 12 strata and considering the two surveys (January and March) separately and together.

Criterion 3 was chosen, even though it was not the ideal one, but it was the most easily executable given the existing manpower situation. It consists of a sample of 5,042 farms.

LITERATURA CITADA

1. COCHRAN, William G. Sampling techniques. New York, John Willey & Sons, 1953.
2. SCHATAN, Salomão. Obtenção de estatísticas agrícolas pelo método de amostragem. São Paulo, Secretaria da Agricultura, Departamento da Produção Vegetal, 1953. (Estudos de Economia Rural, 7).
3. STEVENS, Wilfred L. Levantamento por amostragem da safra de café. Rio de Janeiro, Instituto Brasileiro de Café, 1954.