

DIMENSIONAMENTO DE AMOSTRAS PARA LEVANTAMENTOS AGRÍCOLAS PELA ESTRATIFICAÇÃO DE EFICIÊNCIA MÁXIMA¹

JOSÉ ROBERTO VICENTE², MILTON NOGUEIRA DE CAMARGO³ e RUBENS FERRACIOLI PERES⁴

RESUMO - Dimensionam-se amostras para levantamento de dados de produtores agrícolas filiados à Cooperativa dos Agricultores da Região de Orlândia Ltda., através do método da estratificação de eficiência máxima. Os resultados foram comparados com os provenientes de amostragem casual simples, e considerados adequados para a finalidade desejada.

Termos para indexação: amostragem, amostra estratificada.

SAMPLES DESIGNS FOR AGRICULTURAL SURVEYS BY THE MAXIMUM EFFICIENCY STRATIFICATION

ABSTRACT - Samples are designed for data surveys of farm owners associated to the Cooperativa dos Agricultores da Região de Orlândia, Brazil, by the maximum efficiency stratification. The results were compared with the method of simple random sample and were considered adequate to the desired finality.

Index terms: sampling, stratified sample.

INTRODUÇÃO

A Cooperativa dos Agricultores da Região de Orlândia Ltda. (CAROL) representa cerca de 1.500 agropecuaristas de 48 municípios do Norte de São Paulo e 44 do Triângulo Mineiro. A esses cooperados, presta assistência técnica nas áreas de Agronomia e Veterinária, além de atuar no fornecimento de insumos, na armazenagem e comercialização de produtos e na industrialização da soja.

Conta hoje com uma Unidade de Esmagamento de Soja, em São Joaquim da Barra, centros de serviços agrícolas em Orlândia, Guafra e Ituverava, somando capacidade estática de armazenamento de grãos de 214 mil toneladas, uma rede de oito lojas e dois supermercados, uma corretora de seguros e uma cooperativa de crédito.

A CAROL necessita de dados e informações atualizadas sobre a sua área de atuação e sobre os seus associados para que possa atendê-los convenientemente.

As informações necessárias, dada a grande área de atuação da Cooperativa e o grande número de atividades agropecuárias desenvolvidas pelos seus cooperados, são inviáveis de obtenção através de levantamentos censitários.

A utilização de amostragem, conforme Cochran (1953), além da factibilidade, propicia outras vantagens, como: custo reduzido, rapidez, maior amplitude e exatidão.

O primeiro levantamento agropecuário foi realizado em 1977, quando a CAROL atendia 540 cooperados com 745 propriedades e tinha como área de atuação apenas os 48 municípios localizados no norte do estado de São Paulo. O método adotado nesse primeiro levantamento, e que se repetiu até 1985, era de escolha, através de sorteio, de 30% das propriedades cadastradas (amostra equiprobabilística), estratificadas de acordo com a sua dimensão: 1º estrato - até 20 alqueires; 2º estrato - de 20 a 50 alqueires; 3º estrato - de 50 a 100 alqueires; 4º estrato - de 100 a 250 alqueires; 5º estrato - de 250 a 500 alqueires; e 6º estrato - propriedades acima de 500 alqueires. A adoção desse método apresentou bons resultados, tendo em vista o alto índice (30%) de propriedades pesquisadas em relação ao universo cadastrado na Cooperativa. Com o crescimento da CAROL, o total de imóveis cadastrados passou de 745 para 2.709, e o número de municípios abrangidos na sua área de atuação elevou-se de 48 para 92, o que tornou inviável manter o mesmo índice e método de amostragem. A redução do número de elementos da amostra não permitiu a obtenção de resultados satisfatórios.

O objetivo deste trabalho é dimensionar amostras que permitam levantar, fidedigna e eficientemente, as informações que a Cooperativa necessita para desempenhar a contento suas funções.

¹ Aceito para publicação em 11 de maio de 1988.

² Eng. - Agr., Instituto de Economia Agrícola (IEA), Caixa Postal 8114, CEP 01051 São Paulo, SP.

³ Eng. - Agr., IEA.

⁴ Eng. - Agr., Dep. Agrop. da Coop. dos Agric. da Região de Orlândia, Rua 6 nº 1.676, CEP 14.620 Orlândia, SP.

MATERIAL E MÉTODOS

Utilizou-se como sistema referencial o cadastro de cooperados, onde constam o número de imóveis, a localização de cada um deles e sua área total classificada em ordem crescente. Como a grande maioria dos itens do levantamento (áreas cultivadas, insumos utilizados, etc.) é função da área total do imóvel, esse cadastro foi considerado adequado para fins de dimensionamento das amostras. Naturalmente, supõe-se que não existam no cadastro omissões ou duplicações de imóveis, que causam problemas para o esquema, conforme pode ser visto, por exemplo, em Schattan (1953).

Como os municípios onde os cooperados possuem imóveis diferem quanto às explorações e quanto ao custo do levantamento (devido às distâncias entre eles e ao número de enumeradores na região), optou-se por estratificá-los, *a priori*, em cinco regiões. A região 1 ficou constituída por onze municípios do estado de São Paulo, onde, em imóveis dos cooperados, predominam grãos, cana-de-açúcar, algodão e pecuária. A região 2 agrupa doze municípios de São Paulo, onde existe principalmente café e pecuária leiteira; a região 3 compreende vinte municípios de São Paulo, onde predominam outras explorações. A região 4 é formada por municípios do estado de Minas Gerais e a região 5 por municípios de fora da área de atuação da cooperativa, em diversos estados. Dentre as diferentes funções da estratificação (Kish 1965), o objetivo desta divisão regional, além da obtenção de dados por região, foi o de diminuir a variância e trabalhar com diferentes níveis de erro de amostragem.

Para o dimensionamento das amostras em cada região, utilizou-se o método da estratificação de eficiência máxima, conforme descrito com detalhes em Camargo et al. (1980). Esse método garante que a fração de amostragem nos estratos seja proporcional ao desvio padrão correspondente e um máximo de eficiência da estratificação, cada estrato com apenas dois elementos na amostra e tão uniformes quanto possível (Schattan 1966). Consiste em determinar estratos com variâncias iguais, a partir de um número fixo pré-estabelecido de elementos na amostra ou de um nível predeterminado de erro de amostragem. A variável a ser determinada é, no caso do erro predeterminado, o número de estratos, cada um contribuindo com dois elementos para a amostra. Então, para estimar-se um total T com variância máxima de $(rT)^2$, onde r é o erro de amostragem desejado, cada um dos h estratos deve contribuir para a variância do total com, no máximo, $(rT)^2/h$. Como a variância do total de um estrato obtido com uma amostra de dois elementos é

$$S_h^2 = N_h^2 \cdot S_{\bar{x}_h}^2 = \frac{N_h \cdot \sum x_h^2 - (\sum x_h)^2}{2}, \text{ onde}$$

N_h é o número de elementos da população (no estrato) e X é a variável em questão, cada estrato deverá respeitar a condição

$$N \sum x_h^2 - (\sum x_h)^2 \leq \frac{2 (rT)^2}{h}$$

O cálculo do número ótimo de estratos (e , conseqüentemente, do número de elementos da amostra) tem que ser feito

por tentativa, estipulando-se um número inicial que é corrigido conforme sobrem ou faltem elementos na população para alocar dentro deles. Normalmente, após três ou quatro tentativas consegue-se a estratificação definitiva. Observe-se que em circunstâncias normais, as propriedades finais de um cadastro classificado em ordem crescente, que possuem as maiores áreas, deverão ser levantadas censitariamente, já que mesmo em pequeno número costumam apresentar variâncias com magnitudes superiores aos limites pré-definidos.

Para demonstrar as vantagens da amostragem e da estratificação, construíram-se dois índices: o rendimento da amostragem (R_a), definido como

$$R_a = \frac{N - (n_a + n_c)}{N} \cdot 100,$$

onde n_a é o número de elementos de amostragem e n_c é o resíduo a levantar censitariamente; e o rendimento da estratificação (R_e), que é igual a

$$R_e = \frac{n_{cs} - n_a}{n_{cs}} \cdot 100,$$

onde n_{cs} é o número de elementos necessários para representar a mesma população da amostra estratificada, obtido por amostragem casual simples considerando-se o mesmo nível de erro; portanto,

$$n_{cs} = \frac{N}{1 + (rT)^2 / N \cdot S_T^2}$$

RESULTADOS E DISCUSSÃO

Pelos cálculos efetuados, observa-se que para obter um erro de amostragem menor do que 1% para a área total dos imóveis é necessário levantar 49 deles na Região 1, 43 na Região 2, 37 na Região 3 e 28 na Região 4, incluindo as parcelas a levantar censitariamente. Na Região 5, para obter-se um erro de 3% necessita-se de 25 imóveis. Quando é aumentado o nível do erro permitido, o número de elementos diminui, possibilitando a concentração de esforços nas regiões mais importantes. As amostras e as estratificações apresentaram rendimentos superiores a 80% (exceto para a Região 3 a 1%), freqüentemente acima de 90% (Tabela 1).

Para a Região 1, os estratos compreendem áreas de 1.349 a 14.025 alqueires (1 alqueire = 24.200 m²), com a amostra de 36 elementos representando 1.790 imóveis e 133.258 alqueires. A parcela residual a levantar censitariamente (13 imóveis) responde por uma área de 13.807 alqueires (Tabela 2).

TABELA 1. Número de estratos e de elementos necessários para a obtenção de diversos níveis de erro de amostragem nas diferentes regiões.

Item	Região 1		Região 2		Região 3			Região 4			Região 5	
	erro de amostragem ¹		erro de amostragem ¹		erro de amostragem ¹			erro de amostragem ¹			erro de amostragem ¹	
	1%		1%	2%	1%	2%	3%	1%	2%	3%	3%	5%
Total de imóveis na população	1.803		293	293	123	123	123	192	192	192	298	298
Área total dos imóveis (mil alqueires)	147,1		23,5	23,5	20,4	20,4	20,4	29,8	29,8	29,8	210,4	210,4
Número de estratos obtido ²	18		14	10	8	8	6	8	8	6	8	6
Imóveis representados pela amostra	1.790		278	284	102	117	115	180	191	191	289	292
Imóveis a levantar censitariamente	13		15	9	21	6	8	12	1	1	9	6
Total de imóveis a levantar	49		43	29	37	22	20	28	17	13	25	18
Número de elementos necessário por ACS ³	1.642		259	254	100	112	103	158	167	136	243	197
Rendimento da amostragem (%)	97,3		85,3	90,1	69,9	82,1	83,7	85,3	91,1	93,2	91,3	93,4
Rendimento da estratificação (%)	97,8		89,6	92,1	84,0	85,7	88,3	89,9	90,4	91,2	93,4	93,9

¹ Erro de amostragem para a área total dos imóveis.

² Cada estrato com dois elementos de amostragem.

³ Amostra casual simples, calculado para a parcela coberta pela amostra estratificada; não inclui os imóveis a levantar censitariamente.

Fonte: Cooperativa dos Agricultores da Região de Orlândia e Instituto de Economia Agrícola.

TABELA 2. Dados do cadastro e esquema de amostragem da região I.

Estrato	Número de imóveis cadastrados	Intervalo das áreas (alq.)	Área total dos imóveis (alq.)	Área média dos imóveis (alq.)	Número de elementos a levantar
1	238	1 a 9	1.349	5,7	2
2	227	9 a 16	2.750	12,1	2
3	235	16 a 24	4.664	19,8	2
4	173	24 a 34	4.904	28,3	2
5	144	34 a 46	5.719	39,7	2
6	126	46 a 59	6.520	51,7	2
7	116	59 a 73	7.574	65,3	2
8	88	73 a 84	6.911	78,5	2
9	88	85 a 102	8.152	92,6	2
10	70	102 a 126	8.043	114,9	2
11	58	127 a 155	8.092	139,5	2
12	51	156 a 188	8.679	170,2	2
13	41	189 a 230	8.523	207,9	2
14	36	235 a 282	9.241	256,7	2
15	34	288 a 339	10.561	310,6	2
16	35	340 a 489	14.025	400,7	2
17	16	500 a 580	8.609	538,1	2
18	14	593 a 700	8.942	638,7	2
Subtotal	1.790	1 a 700	133.258	74,4	36
Censo	13	723 a 2.152	13.807	1.062,0	13
Total	1.803	1 a 2.152	147.065	81,5	49

Fonte: Cooperativa dos Agricultores da Região de Orlândia e Instituto de Economia Agrícola.

Efetuada o levantamento, o cálculo das variâncias e valores expandidos é facilitado pela existência de dois elementos de amostragem em cada estrato. Para a estimativa do total do estrato h , tem-se o produto da média pelo número de elementos do universo:

$$\hat{T}_h = N_h \cdot \frac{x_1 + x_2}{2}$$

fazendo-se $\frac{N_h}{2} = F$, obtém-se $\hat{T}_h = Fx_1 + Fx_2$

A estimativa da variância do total do estrato é obtida multiplicando por N_h^2 a estimativa da variância da média:

$$\hat{S}_{Nh\bar{x}}^2 = N_h^2 \hat{S}_{\bar{x}}^2, \text{ com } \hat{S}_{\bar{x}}^2 = \hat{S}_x^2/n.$$

Considerando-se a correção para população finita

$(1 - \frac{n}{N_h})$, obtém-se

$$\hat{S}_{Nh\bar{x}}^2 = N_h^2 (1 - \frac{n}{N_h}) \cdot \frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1} \cdot \frac{1}{n}$$

como $n = 2$; $(\Sigma x)^2 = x_1^2 + x_2^2 + 2x_1x_2$; e $\Sigma x^2 =$

$x_1^2 + x_2^2$, obtém-se

$$\hat{S}_{Nh\bar{x}}^2 = \frac{N_h(N_h - 2)}{4} \cdot (x_1 - x_2)^2;$$

O erro de amostragem é dado por:

$$EA_h(\%) = \frac{\hat{S}_{Nh\bar{x}}}{\hat{T}_h} \cdot 100$$

O cálculo dos totais regionais é dado, naturalmente, pela soma das variâncias e dos totais dos estratos.

Para algum produto em particular ou para um grupo de produtos em determinada região, existe a possibilidade do aparecimento de erros de amostragem considerados demasiado altos. Nesse caso, pode-se redimensionar a amostra pela seguinte relação (Stevens 1951):

$$n_2 = n_1 \cdot \frac{EA_1^2}{EA_2^2},$$

onde n_1 é o número de elementos na amostra original, EA_1 é o erro de amostragem obtido na amostra original, EA_2 é o erro de amostragem desejado e n_2 é o número de elementos necessário para obter EA_2 .

A qualidade dos resultados deve melhorar se um esquema de depuração de dados for implementado, seguindo em linhas gerais o proposto por Pino (1986). Em amostras estruturadas por pares, a falta de respostas constitui um problema considerável, devendo-se conseguir que todos os imóveis sorteados sejam levantados, uma vez que cada estrato é representado apenas por dois elementos (sobre falta de respostas e métodos de contornar os problemas dela provenientes Cochran 1953, Pino & Caser 1984).

CONCLUSÕES

1. O método da estratificação de eficiência máxima permitiu dimensionar amostras a partir de níveis de erro predeterminados de forma adequada, evitando estratos criados arbitrariamente e caselas vazias;

2. Considerando-se as condições atuais da Cooperativa, optou-se por utilizar no campo uma amostra composta por 156 elementos, com expectativa de erro de amostragem máximo para a área total de 1% nas regiões 1 e 2, de 2% na região 3 e de 3% nas regiões 4 e 5;

3. Os bons resultados obtidos, amostras relativamente pequenas, sugerem que esse método possa ser empregado com sucesso em situações similares, quando se dispuser de cadastros confiáveis para servir de sistema referencial.

AGRADECIMENTOS

Os autores agradecem as sugestões do Pesquisador Científico Francisco Alberto Pino nas fases iniciais da pesquisa.

REFERÊNCIAS

- CAMARGO, M.N.; BARROS, M.S.; JUNQUEIRA, M.E.B.; FIGUEIREDO, V.P.M.; CAMARGO FILHO, W.P. Estrutura do mercado varejista de gêneros alimentícios da Grande São Paulo; metodologia de amostragem e estimação de parâmetros através de estratificação de eficiência máxima. **Agric. S. Paulo**, 27(1):181-253, 1980.
- COCHRAN, W.G. **Sampling techniques**. New York, J. Wiley, 1953.
- KISH, L. **Survey sampling**. New York, J. Wiley, 1965.
- PINO, F.A. Detecção e correção de erros em levantamentos agrícolas. **Pesq. agropec. bras.**, 21(9):979-85, 1986.
- PINO, F.A. & CASER, D.V. **Falta de respostas em levantamentos por amostragem**; um estudo de caso. São Paulo, Instituto de Economia Agrícola, 1984. 25p. (Relatório de pesquisa, 8/84)
- SCHATTAN, S. A amostragem e as estatísticas agrícolas. **R. bras. Estat.**, 14(55):219-26, 1953.
- SCHATTAN, S. Retirada de amostra de municípios para previsão da produção de algodão, amendoim, arroz, feijão e soja no Brasil. **Agric. S. Paulo**, 13(3/4):55-60, 1966.
- STEVENS, W.L. **Estimativa e previsão de safras através de um levantamento por amostragem**. São Paulo, Secretaria da Agricultura, Divisão de Economia Rural, 1951. 13p. (Estudos de economia rural)