

# COMPARAÇÃO DE DUAS AMOSTRAS PROBABILÍSTICAS PARA A CULTURA DA LARANJA NO ESTADO DE SÃO PAULO

Juliana Di Giorgio Giannotti, Instituto de Economia Agrícola, juliana@iea.sp.gov.br

Vera Lúcia Ferraz dos Santos Francisco, Instituto de Economia Agrícola,

veralfrancisco@iea.sp.gov.br

Denise Viani Caser, Instituto de Economia Agrícola, caser@iea.sp.gov.br

**RESUMO:** Em virtude, principalmente, da importância econômica da cultura da laranja para o Estado de São Paulo, existe a necessidade de obtenção de estatísticas precisas, a fim de que os setores público e privado possam embasar suas políticas agrícolas. Um modo de conseguir tais informações, com menor custo, pode ser por meio de levantamentos probabilísticos. O presente trabalho teve como objetivo obter duas amostras dos imóveis rurais paulistas produtores de laranja, e verificar qual delas apresentava estimativas mais precisas. O número de elementos foi obtido a partir de informações de 31.421 Unidades de Produção Agropecuária que continham a cultura da laranja no Estado de São Paulo, obedecendo um delineamento amostral por alocação proporcional duplamente estratificado, tendo como primeiro estrato grupos homogêneos, obtidos por meio de análise de agrupamento dos 467 municípios paulistas que produzem laranja e, como segundo estrato duas classificações de áreas dos imóveis rurais. Foram geradas duas amostras com tamanhos amostrais de 1330 ( $cv=1,39$ ) e 666 ( $cv=1,28$ ) elementos. De posse dos tamanhos amostrais e erros de amostragem, foram simuladas, segundo estudo de Monte Carlo, 2.500 amostras para cada cenário. Os resultados de simulação auxiliaram na conclusão de que a amostra com menos elementos pode ser mais apropriada em termos de precisão e custo operacional.

**Palavras chave:** amostragem probabilística, análise multivariada, estudo de Monte Carlo.

## 1. INTRODUÇÃO

A importância econômica da citricultura paulista pode ser observada tanto pela expansão da área plantada, que passou de 39.600ha na década de 1930 a 696.100ha entre os anos de 2000 a 2004, quanto por sua participação na produção nacional, correspondendo a 78% do total colhido no Brasil sendo grande parte destinada à produção de suco concentrado para exportação (Caser & Amaro, 2004).

Em virtude da relevância econômica desta cultura, há a necessidade de informações estatísticas que forneçam subsídios à pesquisa científica, à extensão rural, aos produtores e às instituições governamentais, para melhor planejamento de suas atividades. Estas informações podem

ser obtidas por levantamentos de dados estatísticos provenientes de amostras probabilísticas, os quais apresentam menores custos que um censo, dado que a cultura está disseminada em 467 dos 645 municípios que constituem o Estado de São Paulo.

Deste modo, o presente trabalho teve como objetivo obter duas amostras probabilísticas dos citricultores paulistas, estratificada por município e por área das Unidades de Produção Agropecuárias (UPAs, que na maioria dos casos coincide com o imóvel rural), e verificar qual das amostras apresenta estimativas mais precisas.

## **2. METODOLOGIA**

O delineamento amostral utilizado para selecionar as UPAs produtoras de laranja do Estado de São Paulo foi o da amostragem duplamente estratificada. As principais vantagens deste delineamento são o aumento da precisão das estimativas e a redução do tamanho amostral, devido a se trabalhar com uma maior homogeneidade dentro de cada estrato e heterogeneidade entre estratos (Francisco & Pino, 2000).

Para obtenção da primeira estratificação, denominada estrato geográfico, realizou-se análise de agrupamento, utilizando o método hierárquico da ligação completa (Johnson & Wichern, 1998), considerando as variáveis: produção de laranja dada por caixa de 40,8kg; número de pés em produção e; número de pés novos sem produção, de 467 municípios presentes na previsão e estimativa das safras agrícolas do Estado de São Paulo (Banco IEA, 2005), entre os anos de 2000 e 2004. Para a realização desta análise utilizou-se o procedimento *cluster* do SAS (SAS Institute Inc., 2005).

A segunda estratificação, denominada estrato de área, foi definida para a primeira amostra quatro classes: [0-62,5ha); [62,5-125ha); [125-250ha); e acima de 250ha, e para a segunda amostra oito classes: [0-1,5ha); [1,5 a 15,6ha); [15,6 a 31,2ha); [31,2 a 62,5ha); [62,5 a 125ha); [125 a 250ha); [250 a 625ha); e acima de 625ha, de acordo com especialistas da área. Desta maneira, foram estudadas duas amostras, ambas com o mesmo estrato geográfico e com estratos de área contendo quatro e oito classificações, denominadas, respectivamente, de amostra 1 e amostra 2.

O procedimento utilizado para a formação das duas amostras, considerando a estratificação geográfica e as duas estratificações de área, foi o da alocação proporcional, em que uma amostra de tamanho  $n$  é distribuída proporcionalmente ao tamanho dos estratos (Bolferine & Bussab, 2005). Os cálculos necessários para alocação dos elementos nas duas amostras foram realizados no SAS (SAS Institute Inc., 2005), aceitando um erro de amostragem de 1%. Para tanto utilizaram-se as informações de 31.421 UPAs produtoras de laranjas, contidas no censo agropecuário do Levantamento Censitário de Unidades de Produção Agropecuária da Coordenadoria de Assistência Técnica Integral (CATI) da Secretaria de Agricultura e Abastecimento do Estado de S.P., atualizado entre os anos de 1998 e 2003.

A fim de verificar qual das amostras produziram estimativas mais precisas, foi realizado um estudo de Monte Carlo com 2.500 iterações, utilizando uma variação da técnica de *bootstrap*, em que

as duas amostras foram simuladas do cadastro considerando a média populacional de área e o número de elementos calculados para cada amostra. Esta simulação foi realizada no SAS (SAS Institute Inc., 2005).

### 3. RESULTADOS E DISCUSSÃO

A análise de agrupamento dos 467 municípios paulistas produtores de laranja proporcionou a reunião de municípios com características semelhantes, de pés e produção, em um mesmo grupo resultando em 10 estratos geográficos (grupos) compostos por: 418 municípios (grupo 1); 13 municípios (grupo 2); 14 municípios (grupo 3); 10 municípios (grupo 4); 6 municípios (grupo 5); 2 municípios (grupo 6); 1 município (grupo 7); 1 município (grupo 8); 1 município (grupo 9); 1 município (grupo 10).

Considerando os 10 estratos geográficos e, respectivamente os quatro e oito classes de área, calcularam-se duas amostras que resultaram em: 1330 elementos e coeficiente de variação de 1,39 (amostra 1); e 666 elementos e coeficiente de variação de 1,28 (amostra 2). Os tamanhos amostrais representaram, respectivamente para as duas amostras, 4,2% e 2,1% do número total de UPAs presentes no cadastro. Pôde-se verificar que a maior segmentação de áreas, presente na amostra 2, promoveu uma distribuição mais uniforme dos elementos nas classes com relação a amostra 1, onde se observa que mais de 50% das UPAs foram alocadas nas classes contendo as menores e as maiores áreas.

Tabela 1. Tamanho das amostras 1 e 2, por estratos geográficos e de área, Estado de São Paulo

Estrato de Área	Estrato Geográfico										Soma
	1	2	3	4	5	6	7	8	9	10	
AMOSTRA 1											
[0-62,5)	400	88	73	82	66	2	31	6	6	41	795
[62,5-125)	19	6	7	7	8	2	2	2	2	3	58
[125-250)	20	6	5	8	8	2	2	2	2	2	57
> 250	128	42	65	66	70	18	2	12	11	6	420
AMOSTRA 2											
[0-1,5)	4	2	2	2	2	1	2	2	2	2	21
[1,5-15,6)	32	10	7	7	5	2	5	2	2	4	76
[15,6-31,2)	10	4	3	4	3	2	2	2	2	2	34
[31,2-62,5)	10	4	4	4	4	2	2	2	2	3	37
[62,5-125)	11	3	4	4	5	2	2	2	2	2	37
[125-250)	12	4	3	5	5	2	2	2	2	2	39
[250-625)	102	33	44	42	55	13	2	7	9	8	315
> 625	26	9	21	24	15	5	-	3	2	2	107

De posse dos números de elementos de cada uma das amostras, procedeu-se à simulação de 2.500 amostras a fim de conhecer a distribuição relativa das médias e erros de amostragem. Os resultados da simulação indicaram que as estimativas das médias de área apresentaram alta frequência de ocorrência próximas à média verdadeira (26,21ha), fator positivo no que diz respeito à precisão

dessas estimativas, e leve assimetria representada pela cauda mais longa à direita nas duas situações (Figura 1). Os coeficientes de variação estimados apresentaram maiores valores nas simulações da amostra 1, em que o menor valor observado foi de 1,45, o maior de 1,72 e a alta frequência de ocorrência se deu em torno de 1,59. Já as simulações para a amostra 2 apresentaram menores valores de coeficiente de variação, sendo o menor observado de 1,06, o maior de 1,41 e alta frequência de ocorrência se deu em torno de 1,23.

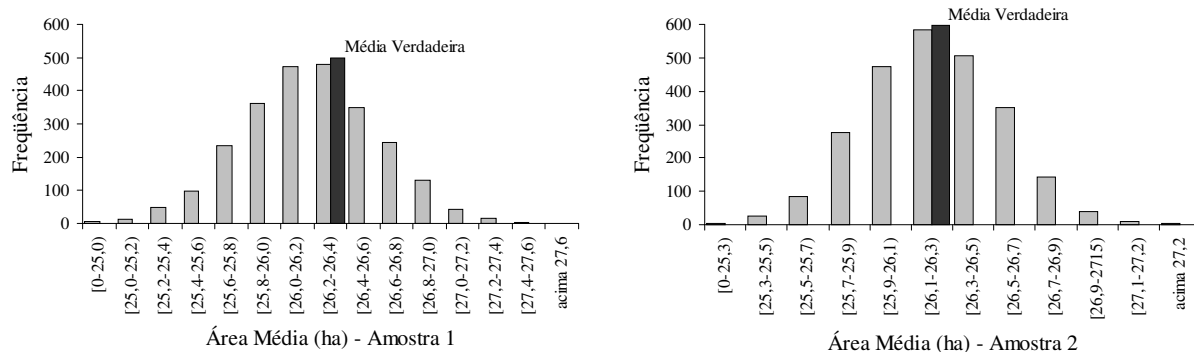


Figura 1. Distribuição de Frequência da Média de Área para as amostras de 1 e 2, Estado de São Paulo

#### 4. CONCLUSÕES

Apesar das duas amostras apresentarem resultados satisfatórios no que diz respeito tanto à precisão das estimativas quanto aos erros de amostragem, a amostra 2 pode ser mais indicada em virtude de apresentar menores coeficientes de variação, quando se observa os resultados das simulações e, principalmente, menos imóveis rurais produtores de laranja a serem visitados, fato que certamente diminuirá, de maneira considerável, o custo de um levantamento.

#### REFERÊNCIAS BIBLIOGRÁFICAS

- Bolfarine, H.; Bussab, W. O. (2005) *Elementos de Amostragem*. São Paulo: Edgard Blücher.
- Banco IEA. (2005) *Previsões e Estimativas das safras agrícolas do estado de São Paulo*. <http://www.iea.sp.gov.br/out/banco/menu.php>
- Caser, D. V. & Amaro, A. A. (2004) Evolução da produtividade na citricultura paulista. *Informações Econômicas*, **34** (10), 7-12.
- Francisco, V.L. dos S.; Pino, F.A. (2000) Estratificação de unidades de produção agrícola para levantamentos por amostragem no Estado de São Paulo. *Agricultura em São Paulo*, **41**(1), 79-110.
- Johnson, R.A.; Wichern, D.W. (1998) *Applied multivariate statistical analysis*. 4.ed. Englewood Cliffs: Prentice Hall.
- SAS Institute Inc., (2005) *SAS OnlineDoc version eight*. <http://v8doc.sas.com/sashtml/>.