

## **Avaliação do tamanho da amostra de segmentos regulares para estimar a área plantada com café na região sul de Minas Gerais**

Marcos Adami  
Mauricio Alves Moreira  
Bernardo Friedrich Theodor Rudorff

<sup>1</sup> Instituto Nacional de Pesquisas Espaciais - INPE  
Caixa Postal 515 - 12245-970 - São José dos Campos - SP, Brasil  
{adami, mauricio,bernardo}@dsr.inpe.br

**Abstract:** This work has the objective to identify the best sample size based on regular segments of 1 x 1 km to estimate coffee crop planted area in 37 municipalities in the South of Minas Gerais State. Samples were randomly selected according to a stratification based on percentage of planted area within each municipality. A comparative evaluation for the estimates from six different sample sizes with three values of  $\alpha/2$  (90, 95 e 99%) and two values of E (5% e 2.5%) was carried out. In order to verify the accuracy of the method 100 random drawing were performed for each sample size. Results indicated that best estimates were obtained for sample size greater than 681 segments.

**Palavras-chave:** sampling; coffee; agricultural statistics; amostragem; café; estatísticas agrícolas.

### **1. Introdução**

A estimativa da área plantada com a cultura do café, nas principais regiões produtoras do Brasil, é obtida de maneira subjetiva, isto é, por meio da coleta de informações junto aos diversos segmentos relacionados ao cultivo do café como: produtores, técnicos da área agrícola, venda de insumos, agentes de financiamento, etc. Essa estimativa subjetiva passa a ser a informação oficial que norteia uma série de ações estratégicas do agronegócio. A área de cultivo do café sofre variações que em grande parte são decorrentes do preço do café. Por sua vez é difícil avaliar e quantificar a magnitude destas variações de tal forma que os métodos subjetivos podem conter erros.

O uso de imagens de satélite de sensoriamento remoto para mapear a cultura do café é uma alternativa que pode melhorar a estimativa da área, pois permite obter informações sobre a distribuição espacial e a área plantada. Além disso, é possível fornecer informações estratégicas para ações governamentais com relação à infra-estrutura e assistência técnica aos produtores. O mapeamento das lavouras de café em todo o Estado de Minas Gerais está sendo feito por Moreira (2006). No entanto, o mapeamento do café por meio de imagens orbitais ainda é um procedimento trabalhoso e demorado, pois o comportamento espectral das lavouras de café é muito variado em consequência de espaçamento, sistema de manejo e cultivo, idade, dentre outros fatores. Moreira et al. (2004) mostraram essa variação ao estudar o comportamento espectral do café em formação e em produção.

Uma metodologia complementar, que pode ser utilizada em conjunto com o mapeamento total das lavouras, é o sistema de amostragem de área, em que a interpretação das áreas cafeeiras é feita apenas numa porcentagem da área total, correspondente aos segmentos amostrados, ou pela visita a campo nos locais sorteados, conforme realizado por Adami et al. (2005). Desta forma, o presente trabalho tem por objetivo avaliar a eficiência da estimativa da área plantada com café, por meio de diferentes tamanhos de amostra de área.

### **2. Materiais e métodos**

Para esta pesquisa o universo amostral correspondeu a uma área de 37 municípios, localizados na região fisiográfica do sul de Minas Gerais. A área está contida na órbita 219, ponto 75, do

sensor Thematic Mapper (TM) do Landsat-5 e foi utilizada uma imagem adquirida em 14 de setembro de 2006.

Para obter a estimativa da área a partir dos diferentes tamanhos da amostra, foi utilizado um mapa temático com todas as lavouras de café da área de estudo (Moreira, 2006). Este mapa serviu de dado de referência uma vez que ele foi obtido por meio de uma cuidadosa fotointerpretação das lavouras de café sobre as imagens multitemporais do satélite Landsat-5, bandas TM3, TM4 e TM5 que foram restauradas para pixel com 10 m, segundo a metodologia proposta por Fonseca (1998). A fotointerpretação foi realizada na tela do computador numa escala de 1:20.000 ou menor utilizando a ferramenta de edição matricial. Imagens Ikonos de alta resolução espacial, disponíveis no site do Google Earth, foram utilizadas para auxiliar na identificação das lavouras de café.

Uma vez obtido o mapa temático a próxima etapa foi estratificar a área em estratos homogêneos de uso do solo em relação à cultura do café. Os limites dos estratos foram estabelecidos em função do percentual da área do município cultivada com café. O tamanho da área de cada estrato e o percentual médio da área ocupada com café é mostrado na **Tabela 1**. O estrato 4 foi composto pela região ocupada pelo lago de Furnas. Este estrato não foi amostrado, devido ao baixo percentual da área ocupada com café.

Tabela 1. Limites inferiores e superiores dos estratos, área do estrato e percentual médio da área cultivada com café.

Estrato	Limite do estrato		Área do estrato (km <sup>2</sup> )	Percentual médio da área utilizado com café
	inferior	superior		
1	0,1	10,0	8.227	4,30
2	10,1	20,0	9.352	12,67
3	20,1	+	7.056	22,50
4	0,0	0,1	245	0,03

Com base nesta estratificação, foi elaborado um painel amostral composto por segmentos regulares, com área de 1 km<sup>2</sup>, conforme descrito por Adami et al, (2005). Para cada segmento obteve-se a área ocupada com café, tomando por base o mapa temático de café. A seguir, foram estabelecidos os tamanhos amostrais, calculados por meio da **Equação 1** (Thompson, 2002):

$$n = \frac{(Z_{\alpha/2})^2 pq}{E^2} \quad (1)$$

em que  $n$  é o número total de segmentos amostrados para estimar a área, para um dado percentual de confiança ( $\alpha/2$ ) e um erro de amostragem ( $E$ );  $p$  é o percentual de área dentro do estrato ocupado com café;  $q$  é a área utilizada com outras ocupações do solo, dentro do estrato, representado por  $(1-p)$ , e  $Z$  é o valor crítico tabelado da função normal padrão, para o intervalo de confiança ( $\alpha/2$ ). Foram calculados 6 tamanhos amostrais que combinaram 3 valores de  $\alpha/2$  (90, 95 e 99%) e 2 valores de  $E$  (5% e 2,5%). Este procedimento permite escolher o conjunto amostral capaz de fornecer estimativas compatíveis com a exatidão requerida, utilizando o menor número de amostras possível. Esse tamanho amostral foi subdividido nos 3 estratos (**Tabela 1**), por meio da **Equação 2**:

$$nh = \frac{Nh \cdot Sh}{\sum_{i=1}^m \frac{Nh \cdot Sh}{\sqrt{Ch}}} \cdot n \quad (2)$$

em que  $Nh$  é o número total de elementos do estrato;  $Sh$  é o desvio padrão do estrato;  $Ch$  é o custo da amostragem, calculado em função de  $p$ , definido por  $Ch=1/p$ ,  $n$  é o número total de segmentos que deve ser amostrado na área e  $i$  é o índice para representar o estrato de uso do solo com  $i = 1, \dots, m$  e  $nh$  representa o número de elementos a ser sorteado por estrato.

Para gerar uma metodologia operacional a qual deve, primordialmente, minimizar o custo do trabalho a campo e fornecer estimativa com exatidão aceitável, foram geradas 100 simulações para cada tamanho amostral, utilizado o Método de Monte Carlo. Essas simulações foram realizadas para verificar a precisão de cada tamanho amostral bem como a variabilidade das estimativas da área plantada com café. Com base nesta simulação, será indicado o melhor tamanho amostral.

Para a estimativa da área ocupada com café foi utilizado o modelo de expansão direta (**Equação 3**):

$$\hat{Z}_c = \sum_{i=1}^m e_i \sum_{k=1}^{n_i} z_{i,k} \quad (3)$$

em que  $\hat{Z}$  é o estimador da área;  $i$  é o índice para representar o estrato de uso do solo com  $i = 1, \dots, m$ ;  $m$  é o número de estratos;  $k$  é o índice para representar o ponto amostrado com  $k = 1, \dots, n_i$ ;  $n_i$  é o número de segmentos amostrados no  $i$ -ésimo estrato de uso do solo;  $e_i$  é o fator de expansão ou inverso da probabilidade que um ponto tem de estar na amostra escolhida no  $i$ -ésimo estrato de uso do solo, sendo definido por  $e_i = (n_i/N_i)^{-1}$ , em que  $N_i$  é o número total de elementos no  $i$ -ésimo estrato de uso do solo; e  $z_{i,k}$  é a área de café, no  $k$ -ésimo ponto do  $i$ -ésimo estrato de uso do solo.

A variância foi estimada pela **Equação 4**:

$$v(\hat{Z}) = \sum_{i=1}^m N_i (N_i - n_i) \frac{s_i^2}{n_i} \quad (4)$$

em que  $s_i^2$  é a variância estimada da cultura no estrato  $i$ , definida pela **Equação 5**:

$$s_i^2 = \frac{\sum_{k=1}^{n_i} (z_{ik} - \bar{z}_i)^2}{(n_i - 1)} \quad (5)$$

e  $\bar{z}_i$  é a área média da cultura no  $i$ -ésimo estrato de uso do solo, definida pela **Equação 6**:

$$\bar{z}_i = \frac{\sum_{k=1}^{n_i} z_{ik}}{n_i} \quad (6)$$

O coeficiente de variação (C.V.) é calculado pela **Equação 7**:

$$C.V.(%) = \frac{\sqrt{\hat{v}(Z)}}{\hat{Z}_c} * 100 \quad (7)$$

As estimativas resultantes da amostragem foram comparadas ao dado (mapa) de referência, para averiguação de desvios e tendências.

### 3. Resultados e Discussão

Os estratos são apresentados na **Figura 1**, na qual se pode observar a variabilidade da área da região ocupada com café. As dimensões da área amostrada, por estrato podem ser observadas na **Tabela 2**.

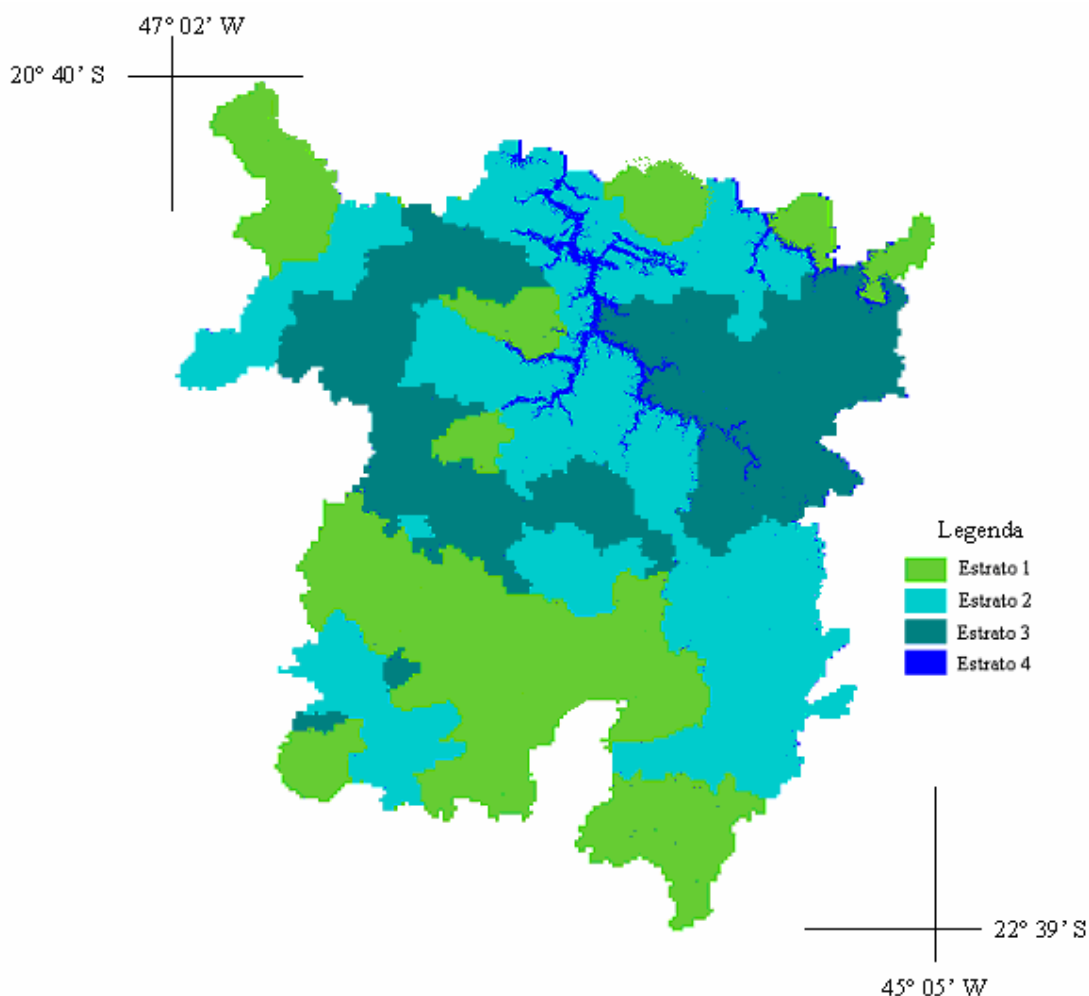


Figura 1. Estratificação da área de estudo em função do percentual da área do município utilizada para o cultivo do café.

A **Tabela 2** apresenta os tamanhos amostrais para estimativa de área cultivada de café. O tamanho amostral variou de 120 a 1.176 segmentos de 1 x 1 km, o que correspondeu a um percentual amostrado de 0,49% a 4,77% em relação à área total.

Tabela 2. Tamanhos amostrais utilizados em função do E e de  $\alpha$  e a subdivisão por estrato.

E		5,0%			2,5%			Área do estrato (km <sup>2</sup> )
		90%	95%	99%	90%	95%	99%	
Estrato	1	14	19	33	54	77	133	8.227
	2	45	65	112	182	258	446	9.352
	3	61	86	149	244	346	597	7.056
Total amostrado		120	170	294	480	681	1.176	24.635
Percentual amostrado		0,49	0,69	1,19	1,95	2,76	4,77	

A **Figura 2** apresenta os resultados das repetições dos sorteios para cada tamanho amostral. Nesta **Figura** pode ser observada a diferença relativa entre o valor estimado pela amostragem e o valor obtido a partir do mapa de referência.

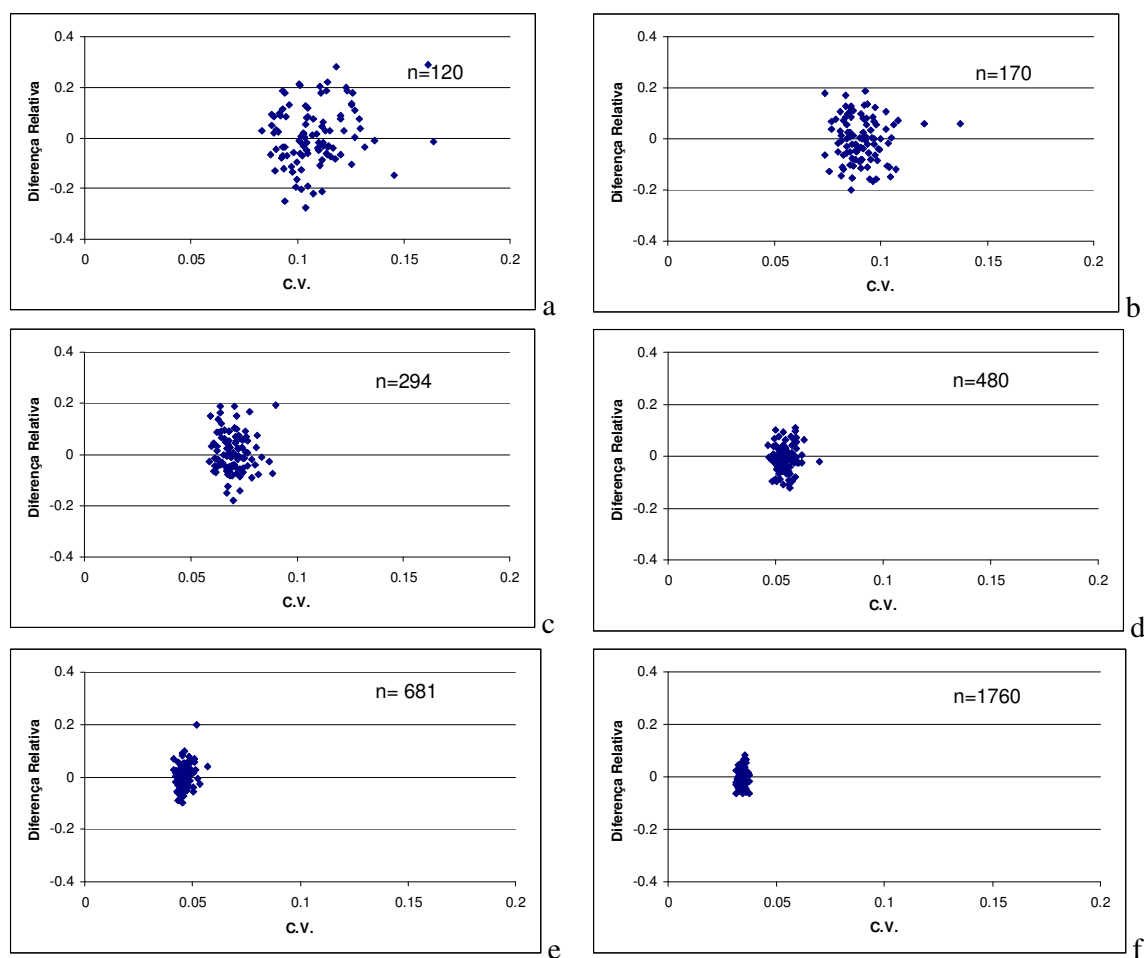


Figura 2. Gráficos de dispersão entre a diferença relativa e o C.V. obtidos para os conjuntos amostrais: a) 120 amostras; b) 170 amostras; c) 294 amostras; d) 480 amostras; e) 681 amostras; e f) 1176 amostras.

De acordo com os resultados nota-se que com o aumento do tamanho amostral, de 120 para 1.176 amostras, ocorre uma redução gradativa do C.V. e da amplitude da diferença relativa (**Tabela 3**). Cabe ressaltar que em todos os tamanhos amostrais a diferença relativa média situou-se em torno de zero (área obtida pelo mapa de referência), o que demonstra que

este método é um estimador não tendencioso para a média (**Figura 2**). Pode-se observar na **Tabela 3** que não há forte correlação entre os valores estimados e o C.V. e estes resultados diferem daqueles encontrados por Rizzi et al. (2006) para a cultura de soja no Rio Grande do Sul, em que foi utilizado o mesmo sistema de amostragem, pois não apresentou correlação entre o C.V. e a diferença relativa.

Tabela 3. Amplitude e média dos C.V. e da diferença relativa e o coeficiente de correlação entre os C.V. e a diferença relativa, por sorteio.

Tamanho amostral	C.V.(%)			Diferença Relativa (%)			Correlação
	mínimo	máximo	média	mínimo	Máximo	média	
120	8,31	16,38	10,78	-27,62	29,02	0,63	0,07
170	7,39	13,73	9,10	-19,89	18,55	-0,37	-0,03
294	5,88	9,02	7,00	-18,09	19,34	0,88	-0,03
480	4,62	7,04	5,48	-12,09	11,09	-0,82	0,08
681	4,11	5,68	4,60	-9,66	19,96	0,20	0,02
1176	3,13	3,77	3,42	-6,22	8,43	-0,24	0,09

As melhores estimativas, tomando por base os valores de C.V. e da diferença relativa, foram obtidas para os tamanhos de amostras de 480 e 1.176 segmentos, o que correspondeu a 1,95 e 4,77% da área respectivamente. Nestes intervalos os coeficientes situam-se entre 3 e 7%.

#### 4. Conclusões

De acordo com os resultados obtidos nesta pesquisa pode-se concluir que a amostragem de área é uma alternativa viável, que pode ser utilizada para uma avaliação rápida da área ou substituir uma interpretação total, no caso da indisponibilidade de imagens livres de cobertura de nuvens.

Com uma área amostrada (1.172), correspondente a 4,77% foi possível estimar a área de café com um C.V. médio de 3,42%.

Em termos de precisão de estimativa de área, os resultados obtidos permitem afirmar que o tamanho ideal de uma amostra de área para estimar o café na região estudada é de 681 segmentos (2,76% da área total), pois apesar de apresentar uma estimativa cuja diferença relativa foi em torno de 20% (**Figura 1**), apresentou C.V. aceitáveis, em torno de 5%,

#### Referências

- Adami, M.; Moreira, M.A.; Rudorff, B.F.T.; Freitas, C.C.; Faria, R.T. Expansão direta na estimativa de culturas agrícolas por meio de segmentos regulares. **Revista Brasileira de Cartografia**, v.1, p. 22 – 27, 2005.
- Moreira, M. A. ; Adami, M. ; Rudorff, B. F. T. . Análise espectral e temporal da cultura do café em imagens Landsat. **Pesquisa Agropecuária Brasileira** , Brasília, v. 39, n. 3, p. 223-231, 2004.
- Moreira, M.A. Mapeamento de café em Minas Gerais. [mensagem pessoal]. Mensagem recebida por: <adami@dsr.inpe.br > em 22 de outubro de 2006.
- Rizzi, R.; Rudorff, B.F.T; Adami, M. Estimativa da área de soja no Estado do Rio Grande do Sul por um método de amostragem, **Ciência Rural**, v.36, p. 30-35, 2006.
- Thompson, S.K. **Sampling** 2 ed. New York: John Wiley & Sons, Inc., 2002. 343 p.