



Avaliação do Levantamento Piloto para Estimativa de Safra Cafeeira no Paraná¹

Prover os agentes econômicos de bens públicos é uma das missões intransferíveis dos governos. A oferta periódica de estatísticas confiáveis insere-se dentro do tipo de bem público altamente necessário para o funcionamento razoável das transações econômicas. Em âmbito da agricultura, uma das sentidas carências no Brasil é a de estatísticas de produção. Por essa razão, existe grande empenho dos gestores públicos na mobilização de financiamentos e arregimentação de pessoal técnico capacitado para fazer frente a essa missão pública.

Nos últimos quatro anos tem havido uma intensificação nos esforços de gerar números mais exatos para a safra de café brasileira. Parceria entre o governo federal e os estaduais, para os quais a cafeicultura é atividade socioeconomicamente relevante, vêm permitindo a produção de estatísticas que espelham com objetividade a realidade produtiva desse segmento a cada safra.

O objetivo deste artigo consiste em apresentar avaliação das estimativas obtidas no levantamento piloto cuja finalidade foi a de obter estimativa de safra de café paranaense e ainda extrair inferências sobre aspectos socioeconômicas relativos à atividade cafeeira.

A abrangência do levantamento foi o Estado do Paraná, sendo realizado no período compreendido entre os meses de setembro a novembro de 2011. A unidade de investigação considerada foi o estabelecimento agropecuário, descrita em IBGE (2009).² Utilizou-se o delineamento amostral proposto em Francisco et al. (2010)³ para o sistema de estatísticas cafeeiras por amostragem probabilística, em nível de produtor, constando de 459 estabelecimentos sorteados no cadastro do Censo Agropecuário do IBGE de 2006. Foram aplicados estimadores de expansão direta às estimativas da média (ou do total) e da variância da estimativa da média (ou do total), os quais podem ser obtidos por fórmulas usadas para o delineamento amostral utilizado no estudo e em pós-estratificação ou por fórmulas em que o tamanho do estrato é aleatório.⁴

Para a coleta de dados, empregaram-se questionários estruturados e a realização das entrevistas com os cafeicultores sorteados que, de forma declaratória, informavam quanto possuíam de café (número de pés, área) e quanto esperavam colher. Os questionários foram divididos em dois blocos de questões: geral (contendo dados globais do estabelecimento), e

específico (contendo dados do cafezal), abrangendo questões como:

- a) produção prevista na safra atual - 2011/12 (sc. 60 kg beneficiada);
- b) produção obtida na safra anterior - 2010/11 (sc. 60 kg beneficiada);
- c) área plantada com café em formação;
- d) área plantada com café em produção; entre outras informações.

Depois de preenchidos os questionários em entrevista com os produtores rurais sorteados, os dados foram digitados em um sistema de entrada de dados criado pela CONAB (Sistema Safras Café) e, em seguida, as informações passaram por um conjunto de testes de consistência interna, antes do processamento estatístico.^{5,6}

O controle estatístico de qualidade do levantamento constou de:

- a) exame dos vieses devidos à falta de resposta;
- b) exame dos coeficientes de variação das estimativas;
- c) comparação com dados de outras fontes.

Segundo Pino e Caser (1984)⁷, “do ponto de vista estatístico, as duas principais consequências da falta de resposta são: a) o tamanho da amostra torna-se aleatório; e b) os estimadores podem se tornar viesados.”

A aleatoriedade do tamanho da amostra não chega a causar problemas, segundo esses autores, desde que se adote um ponto de vista bayesiano. Segundo Pino e Caser (1984)⁸, o viés da estimativa é diretamente proporcional ao número absoluto e à proporção de não respondentes, tanto do total quanto da média, respectivamente, e à diferença entre a média dos elementos que não fornecem respostas e a média dos elementos que as fornecem. Portanto, se as médias desses dois grupos forem iguais, a estimativa será não viesada, e o único efeito da falta de resposta consistirá no aumento da variância da estimativa, isto é, a diminuição de sua precisão.

Antes de cada levantamento, os técnicos entrevistadores foram reunidos para padronização de conceitos habitualmente empregados nos trabalhos de previsão e estimativa de safras.

A primeira etapa do levantamento ocorreu no mês de setembro 2011 e veio a se encerrar em meados de outubro do mesmo ano. O manuseio de questionários, a digitação e a depuração dos dados terminaram na primeira quinzena de outubro de 2011.

Dos 459 questionários enviados, 34 apresentaram algum tipo de falta de resposta, restando, portanto, 425 (93%) questionários aproveitáveis. Percebe-se um alto nível de retorno de informações se comparado com o estudo de caso apresentado em Pino e Caser (1984)⁹, que contabilizaram taxa de retorno variando entre 88% e 94% em análise a levantamentos para previsão e estimativa de safras no Estado de São Paulo, realizados no período de 1975 a 1979, com amostras de 6.229 elementos até 1977 e 5.646 elementos após essa data. Para o caso da previ-

são de safra de café em São Paulo, realizado em parceria CONAB e CATI desde 2001, as taxas de falta de respostas variaram entre 1% e 5% sendo, majoritariamente, devido à recusa do cafeicultor em fornecer as informações sobre sua atividade e também produzindo alto nível de retorno.

Os problemas de falta de respostas relacionam-se, geralmente, à má identificação dos estabelecimentos, isto é, a falhas em seu cadastro. Assim, apenas em 1 dos 33 estabelecimentos o proprietário recusou-se a responder. Pode-se desconsiderar o viés causado pela falta de resposta devido ao baixo índice de falta de resposta total. Para suprir essa deficiência, utilizou-se como técnica de imputação de dados a média dos respondentes.

Uma maneira de expressar a variabilidade dos dados, tirando a influência da ordem de grandeza da variável, é pelo coeficiente de variação (CV) interpretado como a variabilidade dos dados em relação à média. A classificação de um coeficiente de variação em imprecisa, pouco precisa, razoável, boa, ótima e exata pode ser difícil. Entretanto, esse padrão varia de acordo com a natureza da variável a ser estimada ou analisada. Assim, quanto menor o CV, mais homogêneo é o conjunto de dados. Nesse levantamento, alguns coeficientes de variação mostram precisão considerada boa dentro da classificação sugerida, embora superior à desejada em Francisco et al. (2010).¹⁰ Aparecem com CV inferior a 15% as principais variáveis desejadas para inferência, a saber: área em produção (13%), pés em produção (11%) e produção da safra corrente (14%) (Tabelas 1 e 2).

Tabela 1 - Estimativa da Safra Cafeeira Arábica, Estado do Paraná, 2010 e 2011

Levantamento		Safra 2011					Safra 2010	
		Área (ha)			Pés (n.)		Produção	Produção
		Formação	Produção	Total	Formação	Produção	sc. 60 kg	sc. 60 kg
Probabilísticos								
Out. 2011	Estimativa	8.872	79.024	87.895	28.824.912	221.270.677	1.731.045	2.001.958
	CV (%)	18	13	12	18	11	14	17
Nov. 2011	Estimativa	8.812	78.535	87.347	28.534.284	219.768.192	1.724.170	1.968.182
	CV (%)	18	13	12	18	11	14	17
Subjetivos								
CONAB	Valor	16.658	74.752	91.410	45.600.000	241.700.000	1.842.000	2.284.000
IBGE ¹	Valor			75.653			1.826.850	2.317.567

¹A unidade para a estimativa de área é de área colhida

Fonte: Dados da pesquisa.

COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB. Acompanhamento da safra brasileira: café - safra 2011, quarta estimativa, dezembro/2011. Brasília: Conab, 2011. 22p.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Levantamento sistemático da produção agrícola (SIDRA). Disponível em: <<http://www.sidra.ibge.gov.br>>. Acesso em: 27 dez. 2012.

Tabela 2 - Faixas de classificação para os coeficientes de variação e conceito

Intervalo de valores de CV	Indicador	Conceito
Zero	Z	Exata
Até 5%	A	Ótima
Mais de 5% a 15%	B	Boa
Mais de 15% a 25%	C	Razoável
Mais de 25% a 50%	D	Pouco precisa
Mais de 50%	E	Imprecisa

Fonte: INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Estatísticas do café no Estado do Paraná 1998-2001. Rio de Janeiro: IBGE, 2003, 54p.

Com o intuito de reduzir os erros amostrais identificados, os enumeradores decidiram retomar o levantamento ainda em novembro de 2011, direcionando-o aos 33 estabelecimentos contabilizados com erros cadastrais. Nessa nova enquete, retornaram sete elementos ainda com problemas cadastrais. Os 26 questionários com as informações coletadas nessa segunda campanha foram agregados aos demais sem causar problemas na estimação por não ter problema de temporalidade dos dados.

De posse dos novos cálculos, verificou-se que os coeficientes de variação praticamente não se alteraram (Tabela 1). O passo seguinte foi analisar as médias intervalares relativas aos estratos de área e observou-se que, em alguns deles, essas médias se distanciavam mais do respectivo ponto médio do intervalo e os coeficientes de variação apresentavam-se altos. Diante desse fato, alguns procedimentos empíricos foram utilizados para baixar os coeficientes de variação.¹¹

Os dados aqui obtidos foram comparados com os de outras fontes de dados com a mesma finalidade para validações de coerência. Utilizaram-se as seguintes: Levantamento para Previsão e Estimativa de Safras SEAB/PR em parceria com CONAB¹² e Levantamento Sistemático da Produção Agrícola (LSPA) do IBGE¹³ (Tabela 1). De forma geral, percebe-se uma convergência nos resultados apresentados. Os valores estimados para as produções da safra 2010 e safra 2011, nas fontes escolhidas para o confronto, estão contidos nos intervalos de confiança das respectivas estimativas produzidas por meio do levantamento probabilístico por amostragem, podendo-se considerar que não diferem. O mesmo ocorre com as estimativas de área em formação e em produção e o número de pés de café para o levantamento proveniente da CONAB.¹⁴ No entanto, a área estimada proveniente do LSPA¹⁵ encontra-se fora do intervalo e, provavelmente, essa diferença deve-se ao conceito utilizado para a definição de área nos levantamentos comparados neste trabalho.

Pelos resultados obtidos e comparações efetuadas, concluiu-se que o levantamento piloto foi satisfatório. Dos itens utilizados para a análise do controle estatístico

do levantamento, não se encontrou evidência de vieses devido à taxa de não respostas total e estimativas consideradas de boa exatidão quando comparadas aos valores provenientes de outras fontes disponíveis. Quanto à precisão, notou-se que o tempo total decorrido entre o período base do cadastro de estabelecimento - Censo Agropecuário do IBGE de 2006 - utilizado no modelo probabilístico e a aplicação da amostra causou variabilidade nas estimativas acima da desejada, devido ao dinamismo da atividade agropecuária; no entanto, é viável de se ajustar desde que se atualize o sistema de referência.

A relevância e os bons resultados atestaram a viabilidade do projeto, colocando-se como alternativa viável de aperfeiçoamento das estatísticas cafeeiras paranaenses por meio da implantação de modelo que possibilitava aferir a safra anual da cultura do café, além de acompanhar o desenvolvimento da mesma durante a safra investigada. É um processo que não é rápido e barato; porém, permite utilizar o potencial da análise estatística dos dados, sendo possível controlar os erros e os vieses. A principal vantagem é a qualidade dos dados obtidos (e, portanto, sua confiabilidade), uma vez que, num levantamento bem delineado e bem executado, os erros são controlados, obtendo-se estimativas precisas, exatas e não viesadas. O acompanhamento sistemático do delineamento amostral faz com que os resultados futuros tenham consistências estatísticas cada vez mais apuradas.

¹Estudo integrante do Projeto BRA/03/034 - CONAB/PNUD.

²INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Censo Agropecuário 2006**. Rio de Janeiro: IBGE, 2009.

³FRANCISCO, V. L. F. dos S. et al. Modelo estatístico e econômico para estimativa da safra brasileira de café. **Informações Econômicas**, São Paulo, v. 40, n. 12, p. 26-36, dez. 2010.

⁴KISH, L. **Survey sampling**. New York: Wiley, 1965. 643p.

⁵PINO, F. A. Análise do viés em alguns procedimentos para falta de resposta e para erros de resposta em levantamentos por amostragem. **Agricultura em São Paulo**, São Paulo, v. 36, n. 2, p. 147-153, 1989.

⁶_____; CASER, D. V. **Análise de erros não amostrais em levantamentos para previsão e estimativa de safras no estado de São Paulo**. São Paulo: IEA, 1984. 25 p. (Relatório de Pesquisa, 10/84).

⁷_____; _____. **Falta de resposta em levantamentos por amostragem: um estudo de caso**. São Paulo: IEA, 1984. 25 p. (Relatório de Pesquisa, 8/84).

⁸Op. cit. nota 7.

⁹Idem nota 7.

¹⁰Op. cit. nota 3.

¹¹Agregaram-se os estratos de 8 ha a 10 ha, 10 ha a 20 ha, e 20 ha a 50 ha, formando um novo grupo de 8 ha a 50 ha. A média amostral do agregado localizou-se mais próxima ao ponto central com menor variabilidade e, conseqüentemente, os coeficientes de variação da estimativa final apresentaram-se menores, conforme o esperado. Entretanto, a pequena margem de diminuição nos coeficientes de variação não justifica adotar o procedimento em unir os estratos. Outras configurações de combinações entre os estratos foram testadas; porém, apresentaram menores margens de mudança em relação à apresentada neste trabalho.

¹²COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB. **Acompanhamento da safra brasileira: café - safra 2011, quarta estimativa, dezembro/2011**. Brasília: Conab, 2011. 22p.

¹³INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Levantamento sistemático da produção agrícola (SIDRA)**. Disponível em: <<http://www.sidra.ibge.gov.br>>. Acesso em: 27 dez. 2012.

¹⁴Op. cit. nota 12.

¹⁵_____. **Estatísticas do café no Estado do Paraná 1998-2001**. Rio de Janeiro: IBGE, 2003, 54p.

Palavras-chave: amostragem probabilística, cafeicultura paranaense, controle de qualidade, estimativa de safra.

Paulo Sérgio Franzini
Especialista em Economia Cafeeira e Agronegócio
DERAL/SEAB
franzini@seab.pr.gov.br

Carlos Hugo Winckler Godinho
Especialista em Agronegócio
DERAL/SEAB
hugowg@seab.pr.gov.br

Celso Luis Rodrigues Vegro
Pesquisador do IEA
celvegro@iea.sp.gov.br

Vera Lúcia Ferraz dos Santos Francisco
Pesquisadora do IEA
veralfrancisco@iea.sp.gov.br

Valéria Maria Rodrigues Fchine
Consultora em Metodologia Estatística
CONAB
valeria.fchine@bra03034.conab.gov.br

Liberado para publicação em: 29/02/2012