
Notas técnicas

Objetivo

A Pesquisa da Cultura do Café no Estado do Paraná teve como propósito principal fornecer informações de natureza estatística sobre a situação da cafeicultura, através de entrevista direta e do emprego de métodos probabilísticos, que permitissem a associação de um intervalo de confiança aos resultados finais.

Período de coleta

A coleta das informações ocorreu em outubro de 1998, nos meses de abril, julho e novembro de 1999 e 2000, e nos meses de abril e novembro de 2001.

Variáveis pesquisadas

As principais variáveis da pesquisa realizada no mês de outubro de 1998 foram: área plantada (com pés em idade produtiva e com pés novos), produção de café (em coco), efetivo de pés (covas em produção e de pés novos), cultivar, idade (em meses), espaçamento, pragas, doenças e nematóides, por talhão.

Além destas variáveis essenciais à elaboração de estimativas relativas à safra, foram coletadas informações do estabelecimento sobre mão-de-obra empregada na cafeicultura, número de diárias pagas, uso de assistência técnica, práticas agrícolas (conservação do solo, uso de insumos, calagem, defensivos agrícolas, mecanização agrícola, mudas e irrigação), associação a cooperativas, infra-estrutura, tulhas e equipamentos, e viveiros. Também foram investigadas informações sobre a renda média obtida, armazenamento, sistema de colheita (no pano, no chão, a dedo ou mecânica) e comercialização (venda e destino).

Nas pesquisas realizadas no mês de abril dos anos de 1999, 2000 e 2001 foram investigadas as variáveis: área plantada (com pés em idade produtiva e com pés novos), produção de café (em coco), efetivo de pés (em produção e novos) e renda média obtida da safra anterior. Ainda foram coletadas informações sobre armazenamento e comercialização (venda e destino). Em novembro de 1999, 2000 e 2001, além da investigação das variáveis principais - área, produção e efetivo de pés -, inquiriu-se a respeito do sistema de colheita (no pano, no chão, a dedo ou mecânica) e da renda média amostrada, medida ou presumida. Um prognóstico da safra subsequente sobre a área, as covas em produção e a produção total esperada em coco foi feito nas pesquisas realizadas no mês de novembro.

As variáveis investigadas nas pesquisas do mês de julho de 1999 e 2000 foram as mesmas da pesquisa realizada em novembro de 1998, exceto para a forma de levantamento de pragas, doenças e nematóides que, ao invés de ser por talhão, passou a ser pesquisada no estabelecimento na forma de mapeamento (com controle, sem controle ou não ocorreu).

Período de referência das informações

Em todas as pesquisas realizadas, as variáveis área plantada, produção (em coco) e efetivo de pés tiveram como referência o ano da safra investigada - 1998, 1999, 2000 e 2001-, e as informações sobre o encerramento dessas safras foram obtidas nas pesquisas realizadas em outubro de 1998, novembro de 1999, 2000 e 2001. As avaliações do comportamento da safra de 1999, 2000 e 2001 foram feitas nas pesquisas realizadas em abril e julho de cada ano.

As informações sobre cultivares, idade e espaçamento das plantas, pragas, doenças e nematóides, mão-de-obra e número de diárias, também, tiveram como referência o ano da safra investigada.

As informações sobre o uso de assistência técnica, práticas agrícolas, associação a cooperativas, infra-estrutura, tulhas e equipamentos dizem respeito às utilizadas habitualmente pelos produtores.

As informações sobre armazenamento e comercialização e renda média obtida foram referentes à safra anterior. A renda média amostrada medida ou presumida é referente à safra investigada.

Unidade de investigação

A unidade de investigação da pesquisa é o estabelecimento agropecuário, que se define como todo terreno de área contínua, independente de tamanho ou situação (urbano ou rural), onde se processa uma exploração agropecuária com finalidade de comercialização (venda ou troca).

Aspectos da amostragem

Plano amostral

Devido à característica de assimetria das distribuições das principais variáveis a serem investigadas, optou-se por um desenho amostral do tipo amostra estratificada por corte. Nesse tipo de amostragem, um grupo de unidades amostrais, cuja importância em relação ao objetivo da pesquisa é bastante grande, é investigado

censitariamente (estrato certo), enquanto as demais unidades são investigadas efetivamente através de uma amostra aleatória estratificada simples, ou seja, seleção aleatória simples sem reposição dentro dos estratos amostrados.

Para o dimensionamento desta amostra, foi usado um algoritmo, baseado nos estudos de Lavallée e Hidiroglou (1988)¹, satisfazendo o seguinte objetivo: estabelecer o tamanho da amostra o menor possível garantindo o nível de precisão desejado para a variável utilizada na definição do corte, mediante um número determinado de estratos. A alocação da amostra nos estratos foi proporcional a uma medida de tamanho. A medida de tamanho utilizada foi a raiz quadrada da variável *efetivo de pés de café* em cada estrato, que tem a finalidade de suavizar o efeito de valores extremos.

No caso da pesquisa para avaliação da safra de café, verificou-se que a variável mais adequada para a estratificação dos estabelecimentos produtores de café era o *efetivo de pés de café*. Esta variável, ao representar a totalidade dos pés em produção e os novos, revela a potencialidade do estabelecimento em relação à produção, além do que apresenta alta correlação com a grande maioria das demais variáveis objetos da pesquisa.

O cadastro dos estabelecimentos do Censo Agropecuário 1995/1996, constituiu a base da pesquisa.

Em particular, para o Estado do Paraná, decidiu-se por não pesquisar aqueles estabelecimentos com efetivo menor que 1 000 pés, devido ao insignificante peso econômico dos mesmos.

Fixando-se um coeficiente de variação desejável de 2% para estimar a variável *efetivo de pés de café*, e fazendo-se a divisão da população em cinco estratos, sendo um deles o estrato certo, obteve-se o Quadro 1.

Quadro 1 - Estratificação dos estabelecimentos produtores de café - Paraná - 1996

Estrato	Efetivo de pés de café	Total de estabelecimentos	Tamanho da amostra
Total		20 472	327
1	1 000 a 3 798	9 427	34
2	3 799 a 9 932	6 997	48
3	9993 a 29213	3095	51
4	2 9214 a 119 418	807	48
5	Maior que 119 418	146	146

Fonte: IBGE, Diretoria de Pesquisas, Censo Agropecuário 1995/1996.

¹ LAVALLÉE, P. ; HIDIROGLOU, M. On the stratification of skewed populations. *Survey Methodology*, Ottawa: Statistics Canada, v.14, n. 1, p. 33-43, 1988.

Cálculo das estimativas

As estimativas de totais apresentadas no plano tabular para as variáveis investigadas foram calculadas utilizando-se o estimador:

$$\hat{Y} = \sum_{h=1}^L \sum_{i=1}^{n_h} (N_h / n_h) \cdot y_{hi} \quad \text{onde: (1)}$$

L é o número de estratos

N_h é o número de estabelecimentos produtores de café no estrato h , no cadastro utilizado

n_h é o número de estabelecimentos produtores de café no estrato h , na amostra

y_{hi} é o valor da variável no i -ésimo estabelecimento, no h -ésimo estrato, obtido na pesquisa

As estimativas de rendimento médio da lavoura de café foram calculadas através de um estimador de razão da forma:

$$\hat{R} = \hat{P} / \hat{A} \quad \text{onde: (2)}$$

\hat{R} é o estimador da razão das variáveis produção e área da lavoura de café;

\hat{P} é a estimativa do total da produção da lavoura de café ;

\hat{A} é a estimativa do total da área produtiva da lavoura de café; e

\hat{P} e \hat{A} foram obtidos, usando a fórmula (1).

Precisão das estimativas

Numa pesquisa por amostragem, o que se calcula são estimativas de alguns parâmetros das distribuições das variáveis pesquisadas. Mais frequentemente, o interesse recai sobre valores médios ou totais dessas variáveis.

Associados aos valores pontuais das estimativas calculadas, pode-se calcular, também, os erros atinentes às estimativas, uma vez que só foi entrevistada uma parte das unidades que compõem a população objeto da pesquisa. Portanto, os valores das estimativas e os erros a eles associados estão intimamente ligados à amostra selecionada.

Uma medida comumente usada para avaliar o erro de amostragem é o Coeficiente de Variação-CV. Esta medida dá uma idéia do erro relativo cometido, *vis-à-vis* ao valor verdadeiro do parâmetro estimado.

O coeficiente de variação para as estimativas de totais e de rendimento médio da lavoura de café, foram obtidos conforme apresentado a seguir:

$$cv(\hat{Y}) = \sqrt{v(\hat{Y})/\hat{Y}^2} \quad \text{e} \quad cv(\hat{R}) = \sqrt{v(\hat{R})/\hat{R}^2}$$

Onde o estimador da variância de um total² é dada por:

$$v(\hat{Y}) = \sum_{h=1}^L N_h^2 [(1/n_h) - (1/N_h)] \cdot [\sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2 / (n_h - 1)] \quad \text{onde: } \bar{y}_h = (1/n_h) \sum_{i=1}^{n_h} y_{hi} \quad (3)$$

² Cochran, W.G. (1997). Sampling Techniques. Third Edition, John Wiley and Sons.

E a variância da razão é dada por:

$$v(\hat{R}) = (1/\hat{A}^2) \cdot [v(\hat{P}) + \hat{R}^2 v(\hat{A}) - 2\hat{R} cov(\hat{P}, \hat{A})]$$

$$\text{onde: } cov(\hat{P}, \hat{A}) = \sum_{h=1}^L N_h^2 [(1/n_h) - (1/N_h)] \cdot [\sum_{i=1}^{n_h} (p_{hi} - \bar{p}_h) \cdot (a_{hi} - \bar{a}_h) / (n_h - 1)]$$

$$\bar{p}_h = (1/n_h) \sum_{i=1}^{n_h} p_{hi} \quad \text{e} \quad \bar{a}_h = (1/n_h) \sum_{i=1}^{n_h} a_{hi}$$

\hat{A} e \hat{R} são calculados de acordo com a fórmula (1) e (2) respectivamente;

$v(\hat{P})$ e $v(\hat{A})$ são calculadas de acordo com a fórmula (3).

O coeficiente de variação calculado em porcentagem para cada estimativa do plano tabular foi classificado e está apresentado ao lado das estimativas nas tabelas, representado por uma das seguintes letras: Z, A, B, C, D e E, conforme se apresentam no Quadro 2 a seguir.

Quadro 2 - Intervalo de valores dos coeficientes de variação e conceito correspondente para cada indicador apresentado nas tabelas de estimativas

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

Coleta de dados

Os dados relativos à pesquisa da cultura do café foram obtidos mediante a aplicação de questionários na área que compõe o estabelecimento selecionado. Para cada fase da safra - floração, pré-colheita e colheita - um modelo de questionário é utilizado.

Apuração dos dados

Após a etapa da coleta, todo material foi enviado ao DEAGRO para empacotamento. O programa de entrada de dados, validação e expansão da amostra foi desenvolvido por um técnico da Divisão de Análise e Planejamento-ANPLA, do DEAGRO, com supervisão do Departamento de Metodologia-DEMET, utilizan-

do o software IMPS 4.0³ (Integrated Microcomputer Processing System). O dicionário de dados foi elaborado com o módulo DATADICT, enquanto os módulos CENTRY, CONCOR, QUICKTAB foram utilizados para o desenvolvimento da entrada de dados, crítica e apuração dos resultados. O cálculo da precisão das estimativas, expressa em termos do coeficientes de variação, foi realizado com o módulo CENVAR.

Todos os erros apontados pelo programa de crítica foram corrigidos através da análise de relatórios emitidos, com o objetivo de garantir a consistência entre as variáveis investigadas e permitir a expansão dos dados da amostra. Após a eliminação de todos os erros, finalmente foram produzidas as tabelas definidas no plano tabular dando origem aos resultados finais da pesquisa.

³ International Systems Team. U.S. Bureau of the Census, 2000.