

# CONTROLE DE QUALIDADE EM LEVANTAMENTO AGRÍCOLA POR AMOSTRAGEM EM SÃO PAULO<sup>1</sup>

Francisco Alberto Pino<sup>2</sup>  
Vera Lúcia Ferraz dos Santos Francisco<sup>3</sup>

## 1 - INTRODUÇÃO

Ao longo do século XX, a Secretaria de Agricultura e Abastecimento (SAA) do Estado de São Paulo tratou de obter estatísticas agrícolas, em especial as referentes à estimativa de safras agrícolas, utilizando levantamentos de campo, de acordo com diferentes métodos estatísticos e executados de diferentes formas<sup>4</sup>. Nos primeiros quarenta anos daquele século, utilizaram-se censos agropecuários, que são muito bons para conhecer as estruturas agrárias e de produção, mas não para a obtenção de dados conjunturais, que se alteram rapidamente sob influência de mudanças climáticas, de mercado e mesmo políticas. As análises conjunturais exigem que o levantamento e o processamento dos dados, bem como a análise dos resultados, sejam publicados no menor espaço de tempo possível, sob risco de perderem sua utilidade. No início da Segunda Grande Guerra, os censos foram interrompidos e, na falta de melhor opção, iniciou-se em 1942/43 o chamado Levantamento Subjetivo para Previsão e Estimativa de Safras, que existe até hoje. Basicamente, esse levantamento consiste num questionário com perguntas sobre área e produção de muitas dezenas de produtos de origem vegetal e animal, a ser preenchido em cada um dos municípios do Estado pelo Assistente Agropecuário<sup>5</sup> responsá-

vel. Esse levantamento funciona como uma espécie de censo, em que a unidade amostral é o município. Sofre, então, as seguintes críticas:

- a) não tem base estatística, o que impossibilita, portanto, utilizar todo o arsenal dessa ciência para analisar seus dados;
- b) não é possível demonstrar que as estimativas obtidas não contêm vieses nem calcular sua precisão a partir dos próprios dados; estes podem até ser de boa qualidade, mas não há como se verificar;
- c) o responsável pelo preenchimento do questionário responde por uma área muito grande, sobre a qual não teve o menor poder de decisão; isso faz com que a qualidade da resposta dependa intrinsecamente do quanto esse responsável conhece a respeito de cada cultura ou criação naquele município, o que o torna sujeito a pressões, que podem viesar seriamente os dados;
- d) como as medições são grosseiras, não é possível captar variações sutis que ocorrem ao longo do tempo e que, via de regra, são as mais interessantes para o mercado;
- e) essas características tornam esse tipo de levantamento altamente manipulável, podendo-se obter os dados desejáveis, mesmo que distantes da realidade.

Dos anos cinqüentas aos anos noventa, operou-se o chamado Levantamento Objetivo para Previsão e Estimativa de Safras, um levantamento por amostragem probabilística estratificada pela área da unidade rural e região em que se encontra. Com o tempo, outras variáveis tornaram-se importantes nesse levantamento, principalmente as referentes ao trabalho rural. No início da década de 90, quando as estimativas de safras deixaram de levar em conta esse levantamento, valendo-se somente do Subjetivo, ele continuou a ser utilizado para obter as estimativas do trabalho rural.

Em 1968, criou-se uma Seção de Controle de Qualidade das Estatísticas, dentro da Divisão de Levantamentos e Análises Estatísticas,

<sup>1</sup>Trabalho referente aos projetos NIP 3.2000001-68 (Informações Estatísticas de Safras), NIP 3.2000001-58 (Demografia e Trabalho Rural no Estado de São Paulo) e NIP 3.2000001-84 (Previsão e Estimativa de Safras Animais do Estado de São Paulo). Os autores agradecem a todos os participantes do levantamento, em especial às Pesquisadoras Científicas Denise Viani Caser e Maria Carlota Meloni Vicente, ao Engenheiro Agrônomo Ricardo Okano Lourenço, a Maria Cristina Teixeira de Jesus Rowies e a Getúlio B. da Silva.

<sup>2</sup>Engenheiro Agrônomo, Doutor, Pesquisador Científico do Instituto de Economia Agrícola. (e-mail: [pino@iea.sp.gov.br](mailto:pino@iea.sp.gov.br)).

<sup>3</sup>Estatístico, Pesquisadora Científica do Instituto de Economia Agrícola. (e-mail: [veralfrancisco@iea.sp.gov.br](mailto:veralfrancisco@iea.sp.gov.br)).

<sup>4</sup>Para detalhes, ver PINO (1999).

<sup>5</sup>Geralmente Engenheiro Agrônomo ou Médico Veterinário.

do Instituto de Economia Agrícola (IEA), da SAA, que nunca foi instalada. Mesmo assim, na década de 70, alguns estudos foram feitos, como PINO e CASER (1984a, 1984b), PINO; CAMAROGO; VIANI (1979), PINO (1986, 1989).

Em 1995-96, a SAA voltou a realizar um censo agropecuário, chamado Projeto LUPA (PINO et al., 1997). Entre outras finalidades, este censo deveria fornecer a base para novos levantamentos por amostragem. Em 1999, criou-se uma Comissão Técnica de Estatísticas Econômicas (CTEE) na SAA para tratar do assunto, a qual propôs que se instalasse um Sistema Estadual de Estatísticas Agrícolas (SEEA), conforme sugerido em PINO (1999). Razões não-técnicas impediram o prosseguimento dos trabalhos e a Comissão paralisou suas atividades, vindo a ser extinta em 2001. Determinou-se, então, que apenas uma nova amostra para o Levantamento Objetivo fosse apresentada, sem levar em conta as recomendações da CTEE.

O objetivo deste artigo é apresentar o controle de qualidade das estatísticas obtidas em tal levantamento, que foi a campo em novembro de 2000.

## 2 - MATERIAL E MÉTODO

O âmbito do levantamento é o Estado de São Paulo e a base temporal é o mês de novembro de 2000. A unidade amostral é a unidade de produção agrícola (UPA), descrita em PINO et al. (1997), e que na maioria dos casos coincide com o conceito de imóvel rural. Utilizou-se o delineamento amostral proposto em FRANCISCO e PINO (2000) para o SEEA, constando de 3.204 UPAs sorteadas no cadastro do Projeto LUPA, o qual já havia sido testado por PINO; FRANCISCO; LORENA NETO (2001). Embora a amostra não tenha sido delineada para obter estimativas em nível regional, aqui representado pelos Escritórios de Desenvolvimento Rural (EDRs)<sup>6</sup>, elas podem ser obtidas considerando-se uma variável assim definida: igual à variável em estudo, se referir-se a uma UPA dentro de um dado EDR; igual a zero, caso contrário. A estimativa da variância da estimativa da média (ou do total) pode ser obtida por fórmulas usadas em pós-estratifi-

cação ou por fórmulas em que o tamanho do estrato é aleatório, tendo sido utilizada esta última maneira (KISH, 1965, cap. 6).

Depois de preenchidos os questionários em entrevista com os produtores rurais sorteados, os dados foram digitados e passaram por um conjunto de testes de consistência interna, antes do processamento (PINO, 1986; PINO e CASER, 1984b).

O controle 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 (1984a), "*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. O problema mais sério é, pois, o viés introduzido no processo de estimação. Segundo PINO e CASER (1984a), o viés da estimativa, tanto da média quanto do total, é diretamente proporcional à:

- a) proporção de não respondentes (no caso da média) ou ao número de não respondentes (no caso do total);
- b) diferença entre a média dos elementos que não fornecem respostas (não respondentes) e a média dos elementos que fornecem respostas (respondentes). Portanto, se as médias desses dois grupos forem iguais, a estimativa será não viesada e o único efeito da falta de resposta será o aumento da variância da estimativa, isto é, a diminuição de sua precisão.

Existem diversas maneiras de lidar com a falta de resposta. Segundo PINO (1989), a melhor delas consiste em voltar ao campo e levantar os que não haviam respondido. Neste caso, não haverá viés das estimativas. Como isso normalmente não é feito na SAA, resta a técnica de substituição pela média dos respondentes, o que resultará em viés tanto menor quanto menor for a diferença entre os grupos de respondentes e não respondentes (PINO, 1989; PINO e CASER, 1984a). Entretanto, quando não é possível obter a média dos respondentes (se nenhum elemento da amostra responde num estrato, por exemplo), a estimativa acaba sendo igualada a zero. Neste

<sup>6</sup>Para saber quais municípios compõem cada EDR, ver, por exemplo, FRANCISCO et al. (1997).

caso, existe subestimativa tanto do total quanto da média naquele estrato; já no conjunto dos estratos, embora a estimativa da média possa ser não viesada - nas condições da alínea acima (b) - o total continuará subestimado (PINO, 1989).

### 3 - RESULTADOS E DISCUSSÃO

O período base do levantamento era o mês de novembro de 2000. Para tanto, os questionários não preenchidos foram do IEA para a CATI em 18/10/2000. A volta dos questionários preenchidos, da CATI para o IEA, prevista no cronograma para terminar em 30/11/2000, iniciou-se em 29/11/2000 e terminou em 8/1/2001. O manuseio de questionários, a digitação e a depuração dos dados terminaram na segunda quinzena de março de 2001. O processamento dos dados foi feito em três dias, e o controle de qualidade, de 29/3 a 2/4/2001. A análise dos resultados prosseguiu a partir de então. Nota-se que o tempo total decorrido após o período base é muito longo, o que causou o envelhecimento excessivo dos dados: de que adianta saber em abril, quando a colheita já está começando (ou mesmo concluída, para as safras das águas), que até novembro do ano p.p. uma determinada área tinha sido plantada? Nessa data, os agentes econômicos estão interessados em saber quanto será colhido, não quanto foi plantado. Essa defasagem temporal entre a oferta e a demanda de informações estatísticas sobre a agropecuária poderá ser fatal para uma instituição que pretenda atuar no mercado de informações estatísticas do século XXI.

À parte as considerações a respeito de prazos, a qualidade dos dados pode ser avaliada analisando-se a falta de resposta e os erros de amostragem.

#### 3.1 - Falta de Resposta

Dos 3.204 questionários enviados, somente 2.834 (88,5%) foram devolvidos, restando 370 sem qualquer resposta ou explicação. Dos questionários devolvidos, 530 tinham algum outro tipo de falta de resposta, restando, portanto, somente 2.304 (71,9%) questionários aproveitáveis (Tabela 1). Percebe-se um nível altíssimo de falta de resposta, da ordem de 28%, se comparado com o estudo de caso apresentado em PINO e

CASER (1984a). Esses autores analisaram 19 levantamentos para previsão e estimativa de safras no Estado de São Paulo, entre junho de 1975 e junho de 1977, com amostra de 6.229 elementos, e entre setembro de 1977 e fevereiro de 1979, com amostra de 5.646 elementos, tendo a falta de resposta variado entre 6 e 12% ao longo de todo o período.

A distribuição regional da falta de resposta mostra que os EDRs mais afetados, em ordem decrescente, são os seguintes: Itapetininga, Andradina, Sorocaba, Itapeva e São Paulo, todos com mais de 50% de questionários não respondidos. Entre 30% e 50% estão os seguintes EDRs: Franca, Ribeirão Preto, Orlandia, Campinas, Avaré, Registro, Jaú, Presidente Venceslau e Araçatuba. No outro extremo encontra-se o EDR de Moji das Cruzes, totalmente sem falta de resposta, e os EDRs de Botucatu, Marília e Jaboticabal, todos abaixo de 5%.

Expandindo a amostra do levantamento em análise, verifica-se que a falta de resposta corresponde a 62.155 UPAs (22,4%) e a 5.539.019,2ha (27,7%), isto é, cerca de um quarto da área agrícola está deixando de ser levantada apropriadamente (Tabela 2).

A falta de resposta pode ser tipificada ou classificada, quanto às suas origens, em graus de severidade, de tal forma que, mesmo que se resolvam os problemas que determinam a falta de resposta num determinado grau, ela poderá ocorrer nos graus seguintes. Em outras palavras, classifica-se por ordem decrescente de severidade em oito categorias. A classificação aqui apresentada pode ser encontrada em PINO e CASER (1984a), exceto o primeiro e sétimo graus, propostos no presente artigo.

#### 3.1.1 - Primeiro grau: unidade amostral não procurada

Este é o caso mais severo de falta de resposta, no qual a unidade amostral sequer foi procurada. De modo geral, sua ocorrência pode refletir a existência de sérias disfunções organizacionais, como falta de planejamento, falta de responsabilidade (principalmente dos níveis diretivos), falta de envolvimento dos participantes, má administração, falta de conhecimento técnico sobre levantamentos, enfim, falta geral de condi-

TABELA 1 - Falta de Resposta no Levantamento por Amostragem para Previsão e Estimativa de Safras, Estado de São Paulo, Novembro de 2000

EDR	Questionários com Falta de Resposta							Total	%	Número de questionários aproveitados	Total de questionários
	Unidade não procurada	Unidade não localizada	Unidade inacessível	Informante ausente	Recusa do informante	Não entendimento do entrevistador					
Itapetininga	101	0	0	4	1	0	106	86,2	17	123	
Andradina	61	0	0	3	0	0	64	79,0	17	81	
Sorocaba	92	0	0	0	0	1	93	76,2	29	122	
Itapeva	91	2	0	2	0	1	96	73,8	34	130	
São Paulo	20	1	0	0	0	0	21	52,5	19	40	
Franca	23	0	0	8	0	3	34	42,5	46	80	
Ribeirão Preto	8	0	0	23	0	1	32	39,0	50	82	
Orlândia	0	0	0	23	7	5	35	37,6	58	93	
Campinas	22	1	0	1	0	0	24	35,3	44	68	
Avaré	21	1	0	1	2	3	28	35,0	52	80	
Registro	18	3	4	3	0	0	28	35,0	52	80	
Jaú	18	0	0	5	0	1	24	32,4	50	74	
Pres. Venceslau	3	1	0	7	1	6	18	31,6	39	57	
Araçatuba	19	1	0	6	0	1	27	30,3	62	89	
Guaratinguetá	10	0	0	1	1	0	12	29,3	29	41	
Pindamonhangaba	11	1	0	2	1	3	18	28,1	46	64	
Pres. Prudente	9	0	0	14	2	4	29	24,2	91	120	
Barretos	35	0	0	2	1	1	39	23,4	128	167	
Mogimirim	10	0	0	0	0	2	12	21,8	43	55	
Votuporanga	0	0	0	8	0	2	10	19,2	42	52	
Limeira	12	0	0	1	0	0	13	17,6	61	74	
Jales	2	0	0	3	0	3	8	17,4	38	46	
General Salgado	6	0	0	1	0	3	10	16,9	49	59	
Piracicaba	2	0	0	4	0	2	8	16,3	41	49	
Tupã	4	0	0	2	2	0	8	14,0	49	57	
Lins	7	0	0	1	0	0	8	13,8	50	58	
Fernandópolis	0	0	0	3	0	2	5	12,5	35	40	
Araraquara	5	0	0	1	0	3	9	12,2	65	74	
Ourinhos	1	0	0	5	0	0	6	11,8	45	51	
S. José do Rio Preto	2	0	0	6	1	7	16	11,8	120	136	
Dracena	1	0	0	5	0	1	7	11,7	53	60	
Assis	1	0	0	5	0	5	11	9,3	107	118	
S. João da Boa Vista	4	0	0	9	1	1	15	8,8	155	170	
Bragança Paulista	0	0	0	4	1	0	5	8,5	54	59	
Bauru	1	0	0	2	0	3	6	8,0	69	75	
Catanduva	0	0	0	3	2	0	5	5,6	84	89	
Jaboticabal	0	0	0	5	1	0	6	4,8	119	125	
Marília	0	0	0	1	0	2	3	4,5	64	67	
Botucatu	0	0	0	0	1	0	1	2,2	44	45	
Moji das Cruzes	0	0	0	0	0	0	0	0,0	54	54	
Estado	620	11	4	174	25	66	900	28,1	2.304	3.204	

Fonte: Instituto de Economia Agrícola e Coordenadoria de Assistência Técnica Integral.

TABELA 2 - Número e Área de UPAs com Falta de Resposta no Levantamento por Amostragem para Previsão e Estimativa de Safras, Estado de São Paulo, Novembro de 2000

EDR	Número de UPAs			Área das UPAs (ha)		
	Sem resposta	Total	Percentual	Sem resposta	Total	Percentual
Andradina	4.837	5.542	87,3	542.699,4	716.431,7	75,8
Itapetininga	8.944	10.782	83,0	446.132,2	653.979,5	68,2
Franca	3.492	5.553	62,9	329.699,7	503.713,3	65,5
Itapeva	5.676	9.046	62,7	699.981,1	738.330,5	94,8
São Paulo	1.215	2.256	53,9	32.401,0	83.568,3	38,8
Sorocaba	4.468	8.578	52,1	94.337,6	378.516,1	24,9
Araçatuba	2.463	7.784	31,6	196.627,6	572.316,3	34,4
Votuporanga	1.428	4.728	30,2	86.400,1	372.616,2	23,2
Registro	2.372	8.599	27,6	113.796,0	692.962,4	16,4
Barretos	2.299	8.504	27,0	157.289,5	804.673,3	19,5
Ribeirão Preto	1.537	5.995	25,6	270.558,2	642.035,8	42,1
Jaú	1.636	6.402	25,6	166.471,5	500.894,3	33,2
Pindamonhangaba	1.777	6.998	25,4	108.505,5	501.195,0	21,6
Campinas	1.900	7.798	24,4	43.193,9	224.710,5	19,2
Guaratinguetá	1.365	5.741	23,8	255.316,8	446.854,7	57,1
Piracicaba	1.250	7.092	17,6	50.569,4	366.365,3	13,8
Orlândia	824	4.796	17,2	161.526,1	561.242,1	28,8
Limeira	1.361	8.056	16,9	80.942,7	432.192,8	18,7
Presidente Prudente	1.382	8.506	16,2	296.646,4	898.586,1	33,0
Jales	1.278	8.304	15,4	36.302,2	311.733,2	11,6
Lins	699	4.667	15,0	60.719,1	512.359,4	11,9
Mogimirim	817	5.644	14,5	29.735,9	234.438,2	12,7
Avaré	884	6.365	13,9	102.050,7	548.071,4	18,6
Dracena	953	8.664	11,0	97.820,5	452.041,4	21,6
Fernandópolis	427	4.105	10,4	78.328,5	289.197,2	27,1
Presidente Venceslau	554	5.409	10,2	223.930,9	795.539,2	28,1
Bauru	479	5.248	9,1	60.510,1	610.373,2	9,9
São José do Rio Preto	1.028	11.782	8,7	178.119,7	684.777,5	26,0
General Salgado	560	6.725	8,3	42.777,6	578.619,0	7,4
Jaboticabal	691	8.516	8,1	31.517,6	463.416,0	6,8
Araraquara	474	5.906	8,0	79.298,3	624.346,5	12,7
Assis	638	8.096	7,9	40.417,3	627.861,2	6,4
Ourinhos	516	6.789	7,6	26.680,8	469.557,9	5,7
Marília	285	4.251	6,7	76.511,6	452.178,1	16,9
Tupã	363	6.099	6,0	35.659,7	353.025,8	10,1
Bragança Paulista	478	9.756	4,9	72.843,3	273.653,3	26,6
São João da Boa Vista	472	10.164	4,6	62.477,3	556.152,4	11,2
Catanduva	306	7.904	3,9	57.432,7	438.759,6	13,1
Botucatu	25	6.120	0,4	12.790,6	534.593,0	2,4
Mojí das Cruzes	0	3.878	0,0	0,0	98.044,2	0,0
Estado	62.155	277.148	22,4	5.539.019,2	19.999.921,9	27,7

Fonte: Instituto de Economia Agrícola e Coordenadoria de Assistência Técnica Integral, para o número e área de UPAs sem resposta, e Secretaria de Agricultura e Abastecimento do Estado de São Paulo, Projeto LUPA, para número e área de UPAs total.

ções para que o trabalho seja executado.

No levantamento analisado neste artigo este grau de falta de resposta ocorreu de diversas formas. Em 91 questionários (2,8% dos elementos) alegou-se falta de recursos (combustível, veículo, etc.) para executar o levantamento. Este é um caso típico de falha em nível diretivo, pois, uma vez que se estabeleça que um trabalho deve ser feito, é obrigação dos dirigentes prover a equipe dos recursos necessários. Incluem-se, ainda, nesta categoria: casos em que houve atraso na entrega dos questionários para o pessoal de campo; casos em que se alegou não ter havido tempo suficiente para a execução da tarefa; casos em que se alegou que a SAA não dispunha de Casa de Agricultura (e, portanto, de equipe) no município para fazer o levantamento, o que é simplesmente falha de planejamento, uma vez que tais unidades não atendem, necessariamente, apenas um município; casos em que se alegou vagamente que havia dificuldade para obter os dados; e casos em que se utilizaram outras alegações. Finalmente, existem cerca de uma centena de questionários que foram devolvidos em branco, sem qualquer alegação de porque não foram preenchidos, além dos 370 questionários que sequer foram devolvidos. Em princípio, esses casos deveriam ser classificados como falta de resposta sem origem conhecida, mas optou-se por classificá-los nesta primeira categoria, uma vez que fica evidente a causa administrativa do não preenchimento desses questionários. Além disso, a experiência de décadas nesse tipo de levantamento mostra que, quando um bom entrevistador não consegue levantar um questionário, ele costuma escrever o motivo, até como justificativa perante seus superiores. Portanto, o mais provável é que questionários não devolvidos ou devolvidos em branco sequer tenham chegado às mãos dos funcionários que deveriam aplicá-los.

Em resumo, em 620 casos (19,4 % dos questionários) a falta de resposta deveu-se ao fato de a unidade amostral não ter sido procurada (Tabela 1). Essa categoria de falta de resposta mostrou-se crítica nos seguintes EDRs: Itapetininga (82% dos questionários), Sorocaba e Andradina (75% dos questionários), Itapeva (70% dos questionários) e São Paulo (50% dos questionários). Em menor escala, houve problema desse tipo também nos EDRs de Campinas, Franca, Avaré, Guaratinguetá, Jaú, Registro, Aracatuba, Barretos, Mogimirim, Pindamonhangaba,

Limeira, Lins e General Salgado, todos com mais de 10% dos questionários nesta categoria. Por outro lado, convém assinalar que os EDRs de Orândia, Votuporanga, Fernandópolis, Bragança Paulista, Catanduva, Jaboticabal, Marília, Botucatu e Moji das Cruzes não apresentaram falta de resposta devido à unidade amostral não procurada.

Embora PINO e CASER (1984a), analisando o mesmo tipo de levantamento por amostragem para previsão e estimativa de safras, não tenham especificado esta categoria de falta de resposta, pode-se inferir de seus dados que ela está contida em outros motivos ou em motivos não especificados. Portanto, em 1978-80 a participação das unidades amostrais não procuradas em relação ao total de falta de respostas foi igual ou menor do que 31,4%, 29,6% e 38,7%, respectivamente, enquanto no levantamento aqui analisado ela saltou para 68,9% (Tabela 3).

Para evitar casos em que a unidade amostral nem mesmo é procurada, a Comissão Técnica de Estatísticas Econômicas propôs um seminário para dirigentes e responsáveis pelos levantamentos das diversas unidades da SAA, numa estratégia de envolvimento e convencimento, bem como de co-participação e co-responsabilidade, mas o evento não foi aprovado sob a alegação de falta de recursos financeiros.

### 3.1.2 - Segundo grau: unidade amostral não localizada

Neste caso, a unidade amostral foi procurada, mas não foi encontrada. A causa mais comum é o defeito no sistema referencial de amostragem (por exemplo, o cadastro utilizado), mas também pode se dever à falta de conhecimento da região por parte do entrevistador.

No levantamento em análise ocorreram 11 casos, correspondendo a apenas 0,3 % dos questionários (Tabela 1). PINO e CASER (1984a), analisando o mesmo tipo de levantamento por amostragem para previsão e estimativa de safras, nos anos de 1978-80, encontraram os seguintes valores para a participação desta categoria na falta de resposta: 28,2%, 19,9% e 22,1%, respectivamente. No presente levantamento ela caiu para 1,2% (Tabela 3). Essa enorme diferença entre aqueles levantamentos e o atual, quanto à falta de resposta devida à unida-

TABELA 3 - Participação dos Motivos de Falta de Resposta no Levantamento por Amostragem para Previsão e Estimativa de Safras, Estado de São Paulo, 1978-80 e Novembro de 2000

Motivo	1978	1979	1980	2000
Unidade amostral não procurada	...	...	...	68,89
Unidade amostral não localizada	28,21	19,86	22,06	1,22
Unidade amostral inacessível	2,24	2,94	2,3	0,44
Informante ausente	29,49	35,67	26,07	19,33
Recusa do informante	8,65	11,96	10,89	2,78
Não entendimento pelo entrevistador	...	...	...	7,33
Outros motivos	9,94	7,67	10,6	0,00
Motivo não especificado	21,47	21,9	28,08	0,00
Total das faltas de resposta	100,00	100,00	100,00	100,00

Fonte: Instituto de Economia Agrícola e Coordenadoria de Assistência Técnica Integral.

de amostral não localizada, mostra a grande importância de um cadastro bem feito, atualizado e conhecido pelos participantes do levantamento, como é o caso do Projeto LUPA aqui utilizado.

### 3.1.3 - Terceiro grau: unidade amostral inacessível

Nesta categoria, a localização da unidade amostral é conhecida, mas o acesso a ela é impossível, pelo menos temporariamente. O caso mais comum ocorre quando o acesso se dá por estradas não pavimentadas, intransitáveis na época das chuvas.

No levantamento analisado, apenas 4 casos foram registrados (0,1% dos questionários), todos no EDR de Registro (Tabela 1). PINO e CASER (1984a) encontraram para esta categoria os valores 2,2%, 2,9% e 2,3%, respectivamente, nos anos 1978-80, a maior parte na mesma região onde ocorreu no levantamento ora analisado, enquanto em 2000 esse percentual foi de 0,4% (Tabela 3). É provável que o melhoramento das estradas tenha sido decisivo para essa diminuição.

### 3.1.4 - Quarto grau: informante ausente

Aqui, embora a unidade amostral seja encontrada e esteja acessível, a pessoa a ser entrevistada não foi encontrada no endereço. Tanto essa categoria quanto as duas imediatamente

anteriores vêm se tornando menos importantes no Estado de São Paulo, à medida que aumenta o número de proprietários absenteístas, que já atinge 70,2% (PINO et al., 1997). Quando isso acontece, o proprietário passa a poder ser localizado na área urbana, facilitando o levantamento, exceto quando é necessário que o entrevistador faça suas próprias medições e observações na unidade amostral. Por outro lado, quando o proprietário passa de pessoa física para pessoa jurídica, dificulta-se a localização da pessoa autorizada a responder, que às vezes encontra-se na capital do Estado ou mesmo em outros estados.

Neste levantamento, o informante (proprietário da UPA ou seu responsável) não foi encontrado em 170 casos (5,3% dos questionários), sendo o problema mais sério nos EDRs de Ribeirão Preto (28,0% dos questionários), Orlândia (24,7%), Votuporanga (15,4%), Presidente Venceslau (12,3%), Presidente Prudente (11,7%) e Franca (10,0%). Nos EDRs de Sorocaba, São Paulo, Mogimirim, Botucatu e Moji das Cruzes, o problema não existiu (Tabela 1). PINO e CASER (1984a) encontraram para esta categoria os valores 29,5%, 35,7% e 26,1%, respectivamente, nos anos 1978-80, enquanto no levantamento ora analisado a participação desta categoria caiu para 19,3%. Essa diminuição na falta de resposta devida à ausência do informante deve-se, além do aumento de proprietários que residem na cidade, à melhoria no cadastro resultante do Projeto LUPA. Essa categoria poderá ser diminuída ainda mais se houver treinamento dos entrevistadores no sentido de procurar novamente o informante, conforme técnicas descritas em PINO

e CASER (1984a)<sup>7</sup>.

### 3.1.5 - Quinto grau: recusa do informante

Nesta categoria o informante é encontrado, mas recusa-se a responder, seja por desconfiança, falta de tempo, abordagem inadequada do entrevistador ou outro motivo. Este tipo de falta de resposta também aparece quando o proprietário da UPA é uma empresa grande, eventualmente multinacional, cuja política interna restringe ao máximo a saída de informações. Esta é a primeira categoria de falta de resposta que depende estritamente do informante, não havendo muito que se possa fazer, mesmo que todas as questões administrativas e operacionais possam ter sido resolvidas. A solução usual para este problema consiste em enviar outro entrevistador, mais bem preparado ou mais qualificado, para conseguir os dados. O problema também pode ser amenizado mediante:

- a) melhor treinamento do entrevistador, principalmente quanto às técnicas de abordagem;
- b) contratação de entrevistadores de bom nível social e de instrução, com habilidade para comunicação;
- c) envio de correspondência prévia ao informante, assinada por dirigente de alto escalão, explicando os motivos e a utilidade da pesquisa e solicitando que o entrevistador seja atendido;
- d) envolvimento de líderes da comunidade de informantes que possam facilitar o contato entre entrevistador e informante;
- e) publicidade sobre o levantamento em locais e mídias que possam atingir os informantes (jornais, estações de rádio e televisão locais, publicações de cooperativas, associações e sindicatos de produtores, etc.);
- f) envio dos resultados da pesquisa aos informantes (ou publicação em locais e mídias que possam atingi-los), no caso de pesquisas periódicas.

No levantamento em apreço houve somente 32 casos de recusa do informante (1%

<sup>7</sup>A técnica mais comum é a de repasses, que consiste em visitar a unidade tantas vezes quanto for necessário para encontrar o informante. Outra é o método de Politz-Simons, que consiste em estimar a probabilidade de o informante ser encontrado e utilizar essa informação para corrigir o viés. Existe, ainda, o método de Bartholomew, que consiste em duas visitas e algumas alterações nos estimadores.

dos questionários). O problema foi mais sério no EDR de Orlandia (onde a falta de resposta por recusa do informante chegou a 7,5% dos questionários), seguido de Tupã, Avaré, Guaratinguetá, Catanduva e Botucatu, todos acima de 2% dos questionários (Tabela 1). É interessante notar que a maior parte dos EDRs (25 em 40) não apresentou este problema. PINO e CASER (1984a) encontraram para esta categoria os valores 8,7%, 12,0% e 10,9%, respectivamente, nos anos 1978-80, contra apenas 2,78% em 2000 (Tabela 3), mostrando que houve sensível melhoria nesse aspecto.

### 3.1.6 - Sexto grau: ignorância do informante

Neste caso, embora seja encontrado e se disponha a responder, o informante não sabe como responder, seja por não possuir a informação solicitada, seja por não ter grau de instrução suficiente para responder a determinadas perguntas. Questionários mal construídos, com perguntas sobre questões íntimas (no caso de pessoas físicas) ou confidenciais (no caso de pessoas jurídicas), podem contribuir para o aumento de ocorrências nesta categoria. Esta é a segunda categoria que depende mais do informante que do sistema de levantamento. No atual levantamento não se registraram casos de ignorância total do informante, possivelmente porque o nível de escolaridade tem melhorado sensivelmente na zona rural paulista ao longo dos anos.

### 3.1.7 - Sétimo grau: não entendimento por parte do entrevistador

Neste caso, embora o informante seja encontrado, disponha-se a responder e possivelmente saiba a resposta, é o entrevistador que não sabe como agir diante de uma situação inesperada e acaba não preenchendo o questionário. Essa situação é mais comum quando o entrevistador não é qualificado ou não recebeu o treinamento adequado, não sabendo decidir diante de uma situação não prevista nas instruções recebidas.

No levantamento em análise, esse tipo de falta de resposta ocorreu em alguns casos em que a UPA encontrava-se arrendada (e, portanto, não era o proprietário que deveria responder, mas o arrendatário), tinha sido vendida (e o novo



proprietário deveria responder), tinha sido dividida (e cada parte deveria ter sido levantada), tratava-se de assentamento rural, espólio ou situações similares. Foram 66 casos (2,1% dos questionários), sendo o pior no EDR de Presidente Venceslau (10,5% dos questionários), seguido de Jales, Orlandia, São José do Rio Preto, General Salgado e Fernandópolis, todos acima de 5% (Tabela 1). O problema não ocorreu em 15 EDRs.

### 3.1.8 - Oitavo grau: perda de informação

Refere-se à perda de questionários ou de registros após a entrevista, sendo mais comum quando o material é enviado pelo correio ou similar, e não entregue diretamente. Em alguns casos, pode refletir falhas administrativas e operacionais imperdoáveis. Também não se registraram ocorrências nesta categoria no presente levantamento. Em 1978-80 esta categoria também estava incluída em outros motivos e motivos não especificados, no trabalho de PINO e CASER (1984a).

Para evitar casos de falta de resposta que dependam basicamente do entrevistador, a Comissão Técnica de Estatísticas Econômicas propôs que os entrevistadores fossem primeiramente selecionados, depois qualificados e finalmente treinados para esse trabalho. O simples treinamento não basta, porque pressupõe que o funcionário já saiba executar aquele trabalho. Além disso, se um dirigente não estiver convencido da importância do trabalho, tenderá a designar os piores funcionários da unidade para o serviço.

### 3.2 - Variáveis Afetadas pela Falta de Resposta

Como a falta de resposta concentra-se em algumas regiões do Estado, seu efeito deve ser diferente conforme a variável considerada. Assim, o viés poderá ser pequeno em atividades que não ocorram nas regiões onde a falta de resposta foi maior, enquanto em atividades para as quais as regiões afetadas são importantes as estimativas provavelmente não poderão ser utilizadas, ou somente com restrições.

Tomaram-se os dados provenientes do Projeto LUPA somente para os questionários

com falta de resposta no levantamento em análise, e expandiu-se a amostra. A seguir, calculou-se o percentual que esse valor representa no total para cada variável, no Projeto LUPA. Com isso, tem-se uma idéia, ainda que grosseira, do quanto a falta de resposta afeta cada variável, podendo-se decidir se os resultados do levantamento devem ou não ser utilizados (Tabela 4).

Observa-se que em todas as variáveis a perda está acima de 10% do esperado, indo de um mínimo de 11,7% na área plantada com limão até 52,8% na área plantada com tomate. Conclui-se, assim, que a maior parte das variáveis importantes deve ter sido seriamente afetada pela falta de resposta, inviabilizando o uso completo dos dados. Portanto, o máximo que se poderá fazer será apresentar os dados a menos dos EDRs mais afetados, o que apenas vem reforçar a máxima de que as regiões mais atrasadas do mundo, pelo menos de um ponto de vista institucional, sempre têm as piores estatísticas<sup>8</sup>.

### 3.3 - Resultados

Os principais resultados referentes ao controle dos questionários, utilização do solo e estimativa de safras são apresentados a seguir (Tabela 5). Os questionários não respondidos correspondem a 62.124 UPAs, um número muito elevado para que se possa considerar negligível o viés causado pela falta de resposta. A área atual da UPA, bem como a área sobre a qual os questionários foram respondidos, respectivamente 15,7 e 15,4 milhões de hectares, são bem inferiores à área total das UPAS, que é de 20,0 milhões de hectares, segundo PINO et al. (1997). Portanto, a subestimativa dos totais das variáveis estudadas é inevitável.

Comparando-se os resultados do levantamento em análise (Tabela 5) com aqueles obtidos por PINO et al. (1997), verifica-se que a maior parte dessa área que falta concentra-se em pastagens (cerca de 3,4 milhões de hectares), vindo a seguir a área com culturas semi-perenes (1,2 milhões de hectares), a maior parte da qual corresponde à cana-de-açúcar. A soma das áreas

<sup>8</sup>Convencionou-se chamar este fenômeno de *pipe*: piores instituições, piores estatísticas.

TABELA 4 - Levantamento por Amostragem para Previsão e Estimativa de Safras, Efeito da Falta de Resposta sobre Diversas Variáveis, Estado de São Paulo, Novembro de 2000

Variável	UPAs com falta de resposta	Todas as UPAs	Percentual
Limão: área plantada (ha)	4.080,4	34.912,9	11,7
Goiaba: área plantada (ha)	801,3	6.454,3	12,4
Amendoim: área plantada (ha)	7.144,1	51.323,4	13,9
Número de cabeças de bovinos de leite	205.914,0	1.432.512,0	14,4
Manga: área plantada (ha)	3.690,6	25.471,4	14,5
Seringueira: área plantada (ha)	6.153,9	40.649,3	15,1
Laranja: área plantada (ha)	141.477,0	858.761,3	16,5
Tangerina: área plantada (ha)	4.159,5	25.087,4	16,6
Café: área plantada (ha)	38.078,3	228.802,0	16,6
Arroz: área plantada (ha)	9.553,6	44.032,7	21,7
Número de familiares do proprietário	88.091,0	394.411,0	22,3
Número de cabeças de bovinos de uso misto	1.165.029,0	5.177.093,0	22,5
Número de cabeças de bovinos de corte	1.426.655,0	6.056.946,0	23,6
Melancia: área plantada (ha)	2.575,3	10.918,0	23,6
Algodão: área plantada (ha)	34.919,2	136.914,7	25,5
Soja: área plantada (ha)	127.153,1	496.977,6	25,6
Número de trabalhadores permanentes	110.242,0	425.896,0	25,9
Cana-de-açúcar: área plantada (ha)	766.814,1	2.877.172,2	26,7
Milho: área plantada (ha)	359.284,2	1.151.853,3	31,2
Mandioca: área plantada (ha)	9.738,1	30.104,6	32,3
Banana: área plantada (ha)	20.351,1	61.237,5	33,2
Eucalipto: área plantada (ha)	274.052,2	675.505,3	40,6
Feijão: área plantada (ha)	71.443,3	151.770,0	47,1
Cebola: área plantada (ha)	5.574,8	11.688,3	47,7
Tomate: área plantada (ha)	7.956,9	15.078,2	52,8

Fonte: Secretaria de Agricultura e Abastecimento do Estado de São Paulo, Projeto LUPA.

TABELA 5 - Resultados do Levantamento por Amostragem para Previsão e Estimativa de Safras<sup>1</sup>, Estado de São Paulo, Novembro de 2000

Variável	Unidade	Número de UPAs	Total	CV (%)	Média com a característica
Questionários não respondidos	unidade	62.124	62.123,9	5,0	1,0
Área atual da UPA	alq	214.869	6.568.949,5	6,9	30,6
Área atual da UPA	ha	214.869	15.671.906,1	6,8	72,9
Área sobre a qual está sendo respondido este questionário	ha	214.639	15.417.909,4	6,9	71,8
Área de culturas perenes	ha	62.490	972.429,9	5,2	15,6
Área de culturas semi-perenes	ha	41.243	1.720.667,9	6,0	41,7
Área de culturas anuais	ha	72.127	1.260.980,9	5,7	17,5
Área de pastagem cultivada (formada)	ha	124.043	5.833.251,3	4,3	47,0
Área de pastagem não cultivada (incluir natural, invasoras, campo e cerrado)	ha	64.250	1.039.699,7	7,8	16,2
Área de reflorestamento	ha	30.126	694.786,8	13,5	23,1
Área de mata nativa	ha	75.674	2.071.219,7	46,3	27,4
Outras áreas (benfeitorias, inexploradas, etc.)	ha	146.078	502.464,1	9,5	3,4

<sup>1</sup>Dados subestimados devido à falta de resposta em EDRs importantes.

Fonte: Instituto de Economia Agrícola e Coordenadoria de Assistência Técnica Integral.

TABELA 5 - Resultados do Levantamento por Amostragem para Previsão e Estimativa de Safras<sup>1</sup>, Estado de São Paulo, Novembro de 2000

Variável	Unidade	Número de UPAs	Total	(conclusão)	
				CV (%)	Média com a característica
Algodão: área plantada	ha	2.613	56.884,7	20,3	21,8
Algodão: produção esperada	@	2.613	8.988.395,7	20,6	3.439,6
Amendoim das águas em cultura solteira: área plantada	ha	2.373	28.871,6	33,1	12,2
Amendoim das águas em cultura solteira: produção esperada	sc.25kg	2.397	2.419.152,3	29,3	1.009,2
Amendoim das águas em cultura consorciada e/ou intercalada: produção esperada	sc.25kg	144	117.396,2	67,6	817,2
Arroz em cultura solteira de sequeiro: área plantada e por plantar	ha	3.280	14.661,4	68,1	4,5
Arroz em cultura solteira de várzea: área plantada e por plantar	ha	987	5.713,3	22,1	5,8
Arroz em cultura solteira: produção esperada	sc.60kg	3.832	1.028.309,5	40,5	268,3
Arroz em cultura consorciada e/ou intercalada: produção esperada	sc.60kg	291	22.388,6	36,5	76,9
Cana-de-açúcar para indústria: área com cana planta	ha	9.769	2.263.462,5	64,9	231,7
Cana-de-açúcar para indústria: área de primeiro, segundo, terceiro e demais cortes	ha	20.490	1.747.731,1	7,0	85,3
Cana-de-açúcar para indústria: área a ser cortada na próxima safra (2000/01)	ha	18.449	2.537.516,7	34,5	137,5
Cana-de-açúcar para indústria: produção obtida na safra 1999/00	t	20.490	147.856.162,5	7,4	7.215,9
Feijão das águas em cultura solteira: área plantada	ha	8.376	41.320,0	23,1	4,9
Feijão das águas em cultura solteira: produção esperada	sc.60kg	8.376	896.063,3	24,7	107,0
Feijão das águas em cultura consorciada e/ou intercalada: produção esperada	sc.60kg	2.496	44.052,2	67,7	17,6
Milho em grão (não inclui silagem) em cultura solteira: área plantada e por plantar	ha	53.859	681.187,7	8,0	12,6
Milho em grão (não inclui silagem) em cultura solteira: produção esperada	sc.60kg	53.575	55.467.890,5	8,6	1.035,3
Milho em grão (não inclui silagem) em cultura consorciada e/ou intercalada: produção esperada	sc.60kg	3.755	1.290.369,2	40,5	343,6
Soja: área plantada e por plantar	ha	7.005	376.486,1	11,3	53,7
Soja: produção esperada	sc.60kg	7.005	17.316.998,5	11,3	2.472,1
Café: área plantada (novos + adultos)	ha	28.595	244.629,0	9,9	8,6
Café: número de covas (pés novos + pés adultos)	covas	28.370	480.673.780,5	14,4	16.942,9
Café: número de covas (pés novos sem produção)	covas	11.888	92.873.175,0	17,2	7.812,6
Café: produção esperada (coco seco)	sc.40kg	24.412	8.654.480,5	21,6	354,5
Café: Renda esperada no benefício	kg benef./sc.40kg coco	19.438	3.861,9	10,2	0,2
Laranja somente comercial (safra 1999/2000): área plantada (pés novos + pés adultos)	ha	18.326	557.123,9	7,3	30,4
Laranja somente comercial (safra 1999/2000): número de pés (novos + adultos)	pés	18.009	157.365.288,8	8,0	8.738,2
Laranja somente comercial (safra 1999/2000): número de pés (novos sem produção)	pés	4.063	30.991.622,9	26,8	7.627,5
Laranja somente comercial (safra 1999/2000): produção obtida	cx.40,8kg	17.134	271.456.973,6	7,9	15.842,8
Laranja somente comercial (safra 2000/2001): área plantada (pés novos + pés adultos)	ha	18.015	567.562,5	7,1	31,5
Laranja somente comercial (safra 2000/2001): número de pés (novos + adultos)	pés	18.015	159.828.773,6	7,4	8.872,2
Laranja somente comercial (safra 2000/2001): número de pés (novos sem produção)	pés	3.529	25.266.613,4	26,8	7.158,9
Laranja somente comercial (safra 2000/2001): produção esperada	cx.40,8kg	17.329	259.332.444,3	7,5	14.965,3

<sup>1</sup>Dados subestimados devido à falta de resposta em EDRs importantes.

Fonte: Instituto de Economia Agrícola e Coordenadoria de Assistência Técnica Integral.

de reflorestamento e com mata nativa praticamente não sofre alteração, embora individualmente haja subestimativa da primeira e superestimativa da segunda. Em termos percentuais, faltam 41,6% da área de culturas semi-perenes, 33,1% da área de pastagem, 27% da área de culturas perenes e 24,5% da área de culturas anuais.

Deixando de lado o viés da falta de resposta, alguns coeficientes de variação mostram precisão aceitável, embora superior à esperada em FRANCISCO e PINO (2000). Aparecem com CV inferior a 10% as seguintes variáveis: área de pastagem cultivada (4,3%), área de culturas perenes, área de culturas anuais, área de culturas semi-perenes, área atual da UPA, área dos diversos cortes de cana-de-açúcar, área plantada de laranja, número de pés de laranja, produção obtida de cana-de-açúcar, produção esperada de laranja, área de pastagem não cultivada, produção obtida de laranja, área plantada de milho, produção esperada de milho, outras áreas, área plantada com café (9,9%). Isso dá uma indicação de quais variáveis poderiam ter CVs bons caso fosse removido o viés da falta de resposta.

Coeficientes de variação entre 10 e 20% apareceram nas seguintes variáveis: renda esperada no benefício de café (10,2%), produção esperada de soja, área plantada de soja, área de reflorestamento, número total de covas de café, número de covas de café sem produção (17,2%). As demais variáveis tiveram CVs acima de 20%, o que dificulta muito sua utilização.

### 3.4 - Comparação com Outras Fontes

Os dados aqui obtidos podem ser comparados com os de outras fontes, cujos levantamentos, por sua vez, têm também seus problemas específicos. Utilizaram-se as seguintes fontes, todas referentes ao mês de novembro de 2000:

- a) Levantamento Subjetivo para Previsão e Estimativa de Safras IEA/CATI (IEA, 2001);
- b) Levantamento Sistemático da Produção Agrícola (LSPA) do IBGE (IBGE, 2001);
- c) Levantamento por Amostragem para Previsão e Estimativa de Safra de Café da CATI (PINO; FRANCISCO; LORENA NETO, 2001).

Como há subestimativa no levantamento em análise, é natural que seus resultados sejam inferiores aos de outras fontes na maioria dos casos (Tabela 6). Entretanto, a produção de milho resultou maior, mesmo apenas a produção em cultu-

ra solteira, com CV=8,6%, sugerindo subestimativa nos levantamentos das outras fontes. Quanto à produção de café, o resultado mais confiável é o do Levantamento por Amostragem da CATI, com 269,61 mil t, com CV=8,0%, tendo o resultado do LSPA do IBGE sido o que mais se aproximou dele, embora subestimado. Segue-se o do presente levantamento, mesmo subestimado (173,09 mil t, com CV=21,6%), ficando a pior subestimativa com o Levantamento Subjetivo IEA/CATI<sup>9</sup>.

### 3.5 - Ajuste da Amostra para os Respondentes

Sob a suposição de que a falta de resposta ocorre de forma aleatória, o que não costuma acontecer, pode-se tomar a amostra de respondentes como uma amostra probabilística, apenas com tamanho menor<sup>10</sup> (PINO e CASER, 1984a; PINO e FRANCISCO, 2001). Neste caso, podem-se obter estimativas razoáveis, embora as fórmulas de estimação devam ser alteradas para incluir a aleatoriedade do tamanho da amostra. A menos do viés devido à falta de resposta, este tipo de procedimento produz estimativas com variâncias menores do que se a amostra original tivesse tamanho igual ao número de respondentes<sup>11</sup>.

Supondo-se que os elementos faltantes, mesmo que não ocorram de forma aleatória, responderiam, em média, da mesma maneira que os respondentes, o que também não costuma acontecer, poder-se-ia substituir os valores dos não respondentes pela média dos respondentes<sup>12</sup> (PINO, 1989; PINO e FRANCISCO, 2001). Quando a falta de resposta não é muito grave, este procedimento produz estimativas com variâncias menores do que o de tomar a amostra de respondentes como uma amostra probabilística, o contrário podendo ocorrer quando o número de respondentes é muito pequeno em relação à amostra original, como aconteceu em alguns estratos do levantamento em análise<sup>13</sup>.

<sup>9</sup>Ver detalhes em PINO; FRANCISCO; LORENA NETO (2001).

<sup>10</sup>PINO, F. A.; FRANCISCO, V. L. F. S. Some empirical procedures for estimation in sample surveys with non-response. Trabalho em desenvolvimento.

<sup>11</sup>Id. *ibid.*, nota 10.

<sup>12</sup>Id. nota 10.

<sup>13</sup>Op. cit. nota 10.

TABELA 6 - Comparação de Resultados do Levantamento por Amostragem para Previsão e Estimativa de Safras com os de Outras Fontes, Estado de São Paulo, Novembro de 2000

Produto	Variável <sup>1</sup>	Levantamento por amostragem <sup>2</sup>	Levantamento subjetivo <sup>2</sup>	Levantamento sistemático da produção agrícola <sup>3</sup>	Levantamento específico para café <sup>4</sup>
Algodão	Área	56,88	68,36	63,77	...
	Produção	134,83	160,85	138,70	...
Arroz	Área solteira	20,37	45,73	58,80	...
	Produção solteira e consorciada	63,04	124,68	114,66	...
Amendoim	Área solteira	28,87	55,67	64,10	...
	Produção solteira e consorciada	63,41	132,28	126,53	...
Feijão das águas	Área solteira	41,32	65,82	75,95	...
	Produção solteira e consorciada	56,41	86,29	81,51	...
Soja	Área	376,49	502,99	506,10	...
	Produção	1.039,02	1.234,53	1.135,69	...
Milho	Área solteira	681,19	749,50	806,46	...
	Produção solteira e consorciada	3.405,49	2.990,50	3.073,42	...
Laranja (1999/2000)	Área	557,12	671,22	620,00	...
	Produção	11.075,44	14.521,75	14.906,15 <sup>5</sup>	...
Laranja (2000/2001)	Área	567,56	650,36	620,00	...
	Produção	10.580,76	14.186,89	14.906,15 <sup>5</sup>	...
Cana-de-açúcar	Área	2.263,46	2.822,11	2.417,00	...
	Produção	147.856,16	189.040,43	180.621,78	...
Café	Área	244,63	242,32	210,00	241,34
	Número de pés	480,67	...	...	517,47
	Produção	173,09	168,73	210,39 <sup>6</sup>	269,61
Banana	Área	44,70	61,14	52,52	...
Tomate	Área	5,59	11,48	12,29	...
Cebola	Área	3,94	10,62	10,74	...
Mandioca	Área	39,24	59,63	31,07	...
Melancia	Área	6,02	7,64	...	...

<sup>1</sup>Unidades: área em 1.000ha, produção em 1.000t, número de pés em 1.000.000 de pés.

<sup>2</sup>Realizado pelo Instituto de Economia Agrícola e Coordenadoria de Assistência Técnica Integral.

<sup>3</sup>Realizado pelo Instituto Brasileiro de Geografia e Estatística.

<sup>4</sup>Realizado pela Coordenadoria de Assistência Técnica Integral.

<sup>5</sup>Frutos transformados em 1.000t com o fator 163g/fruto.

<sup>6</sup>Saca 40 kg de coco seco de café transformada em 1.000t, beneficiada com o fator 20 kg de café beneficiado por saca 40kg de café coco seco.

Fonte: Instituto de Economia Agrícola; Coordenadoria de Assistência Técnica Integral; Instituto Brasileiro de Geografia e Estatística.

Em ambos os procedimentos há de se considerar ainda que, se houver dois ou mais respondentes em cada estrato, é possível calcular a variância; se houver apenas um, é possível calcular a média e o total, mas não a variância; e, se não houver respondentes, não é possível obter estimativa alguma. Para o caso de apenas um respondente no estrato, tem-se adotado o procedimento empírico de igualar o CV a 100%.

No levantamento aqui analisado, dos

174 estratos, 31 tiveram somente um respondente, não se registrando estratos totalmente sem respondentes.

Utilizando-se esses dois tipos de procedimento, foi possível diminuir sensivelmente os CVs das estimativas para o levantamento ora analisado (Tabela 7). Algumas variáveis, como número de UPAs e área total das UPAs, apresentam valores que se aproximam muito daqueles que aparecem no cadastro original do Projeto

TABELA 7 - Resultados do Levantamento por Amostragem para Previsão e Estimativa de Safras<sup>1</sup>, Estado de São Paulo, Novembro de 2000

(continua)

Variável	Unidade	Número de UPAs	Total	CV <sup>2</sup> (%)	CV <sup>3</sup> (%)	Média com a característica
Área atual da UPA	alq.	277.005	8.707.252,7	1,2	5,2	31,4
Área atual da UPA	ha	277.005	20.806.191,3	1,2	5,2	75,1
Área sobre a qual está sendo respondido este questionário	ha	276.696	20.444.659,5	1,2	5,2	73,9
Área de culturas perenes	ha	79.696	1.233.968,9	1,3	4,2	15,5
Área de culturas semi-perenes	ha	53.846	2.390.685,8	1,7	4,5	44,4
Área de culturas anuais	ha	93.927	1.728.070,2	1,9	4,3	18,4
Área de pastagem cultivada (formada)	ha	160.077	7.745.405,3	1,0	3,4	48,4
Área de pastagem não cultivada (incluir natural, invasoras, campo e cerrado)	ha	83.150	1.369.846,3	1,7	6,0	16,5
Área de reflorestamento	ha	39.105	954.788,2	3,4	10,0	24,4
Área de mata nativa	ha	99.195	2.638.469,2	8,3	36,4	26,6
Outras áreas (benfeitorias, inexploradas, etc.)	ha	187.668	658.391,1	1,8	7,2	3,5
Algodão: área plantada	ha	3.403	82.233,1	12,0	14,9	24,2
Algodão: produção esperada	@	3.403	13.042.654,9	12,4	15,1	3.832,8
Amendoim das águas em cultura solteira: área plantada	ha	3.085	38.490,5	6,0	24,9	12,5
Amendoim das águas em cultura solteira: produção esperada	sc.25kg	3.085	3.203.196,2	6,8	22,2	1.038,4
Amendoim das águas em cultura consorciada e/ou intercalada: produção esperada	sc.25kg	181	160.086,1	9,9	49,6	886,1
Arroz em cultura solteira de sequeiro: área plantada e por plantar	ha	4.361	18.930,3	10,4	52,7	4,3
Arroz em cultura solteira de várzea: área plantada e por plantar	ha	1.260	8.935,4	19,2	16,1	7,1
Arroz em cultura solteira: produção esperada	sc.60kg	5.083	1.428.995,2	10,5	29,4	281,1
Arroz em cultura consorciada e/ou intercalada: produção esperada	sc.60kg	358	31.796,6	5,1	25,7	88,8
Cana-de-açúcar para indústria: área com cana planta	ha	13.067	5.849.875,5	56,8	33,7	447,7
Cana-de-açúcar para indústria: área de primeiro, segundo, terceiro e demais cortes	ha	26.711	2.510.809,3	4,7	5,4	94,0
Cana-de-açúcar para indústria: área a ser cortada na próxima safra (2000/01)	ha	24.094	3.753.545,8	8,4	23,5	155,8
Cana-de-açúcar para indústria: produção obtida na safra 1999/00	t	26.711	214.151.545,4	5,3	5,6	8.017,4
Feijão das águas em cultura solteira: área plantada	ha	11.096	55.628,7	4,3	17,2	5,0
Feijão das águas em cultura solteira: produção esperada	sc.60kg	11.096	1.242.997,5	5,9	18,0	112,0
Feijão das águas em cultura consorciada e/ou intercalada: produção esperada	sc.60kg	3.163	59.105,6	7,2	50,5	18,7
Milho em grão (não inclui silagem) em cultura solteira: área plantada e por plantar	ha	69.782	907.261,3	1,5	6,1	13,0
Milho em grão (não inclui silagem) em cultura solteira: produção esperada	sc.60kg	69.403	74.608.199,3	1,7	6,5	1.075,0
Milho em grão (não inclui silagem) em cultura consorciada e/ou intercalada: produção esperada	sc.60kg	4.871	1.625.217,7	9,0	32,2	333,6
Soja: área plantada e por plantar	ha	9.250	536.283,2	6,3	8,4	58,0
Soja: produção esperada	sc.60kg	9.250	24.460.676,0	5,7	8,3	2.644,4
Café: área plantada (novos + adultos)	ha	36.257	310.289,6	1,5	7,9	8,6
Café: número de covas (pés novos + pés adultos)	covas	35.969	611.014.983,5	2,7	11,4	16.987,0
Café: número de covas (pés novos sem produção)	covas	15.047	117.889.784,5	4,6	13,6	7.834,9
Café: produção esperada (coco seco)	sc.40kg	30.680	10.798.829,6	5,0	17,4	352,0
Café: renda esperada no benefício	kg benef./sc.40kg coco	24.612	4.901,4	1,3	8,1	0,2

<sup>1</sup>Cálculo das estimativas segundo o ajuste da amostra para os respondentes.<sup>2</sup>Utilização do número de respondentes como sendo uma amostra probabilística.<sup>3</sup>Utilização da média dos respondentes como sendo resposta dos não respondentes.

Fonte: Instituto de Economia Agrícola e Coordenadoria de Assistência Técnica Integral.

TABELA 7 - Resultados do Levantamento por Amostragem para Previsão e Estimativa de Safras<sup>1</sup>, Estado de São Paulo, Novembro de 2000

Variável	Unidade	Número de UPAs	Total	(conclusão)		
				CV <sup>2</sup> (%)	CV <sup>3</sup> (%)	Média com a característica
Laranja somente comercial (safra 1999/2000): área plantada (pés novos + pés adultos)	ha	22.927	703.286,3	1,6	5,9	30,7
Laranja somente comercial (safra 1999/2000): número de pés (novos + adultos)	pés	22.533	198.835.197,4	1,9	6,4	8.824,2
Laranja somente comercial (safra 1999/2000): número de pés (novos sem produção)	pés	5.142	40.619.360,3	5,0	20,4	7.900,1
Laranja somente comercial (safra 1999/2000): produção obtida	cx.40,8kg	21.371	343.297.608,0	2,4	6,3	16.063,6
Laranja somente comercial (safra 2000/2001): área plantada (pés novos + pés adultos)	ha	22.469	714.021,3	1,7	5,8	31,8
Laranja somente comercial (safra 2000/2001): número de pés (novos + adultos)	pés	22.469	201.219.991,0	1,9	5,9	8.955,3
Laranja somente comercial (safra 2000/2001): número de pés (novos sem produção)	pés	4.401	32.931.614,6	4,2	20,6	7.483,0
Laranja somente comercial (safra 2000/2001): produção esperada	cx.40,8kg	21.614	327.227.375,9	2,2	6,0	15.139,7

<sup>1</sup>Cálculo das estimativas segundo o ajuste da amostra para os respondentes.

<sup>2</sup>Utilização do número de respondentes como sendo uma amostra probabilística.

<sup>3</sup>Utilização da média dos respondentes como sendo resposta dos não respondentes.

Fonte: Instituto de Economia Agrícola e Coordenadoria de Assistência Técnica Integral.

LUPA e em PINO et al. (1997). Entretanto, há de se perceber que os sérios vieses devidos à falta de resposta permaneceram. Por isso, intervalos de confiança aqui obtidos provavelmente não incluem os respectivos valores verdadeiros. Em outras palavras, conseguem-se estimativas precisas, mas que podem estar distantes da realidade. A comparação com resultados de levantamentos anteriores e de outras fontes mostra-se coerente em alguns casos, evidenciando prováveis falhas também nas outras fontes que servem de comparação.

### 3.6 - Erros no Levantamento

Em todos os levantamentos de dados ocorrem erros que, de alguma forma, devem ser evitados, corrigidos ou, pelo menos, controlados. Existem erros devidos ao esquema amostral, também conhecidos por erros amostrais, e erros não devidos ao esquema amostral, também conhecidos por erros não amostrais. Os erros amostrais compreendem o erro devido à variabilidade da amostra, usualmente medido pelo coeficiente de variação da estimativa, e o viés devido ao estimador, que no levantamento em análise é igual a zero. Os erros não amostrais compreendem diversos erros cometidos durante o

levantamento no campo, como o viés devido à falta de resposta, os erros de preenchimento e outros (para detalhes ver, por exemplo, PINO, 1986). Neste artigo, o interesse principal recaiu sobre os coeficientes de variação de cada estimativa e sobre o viés devido à falta de resposta. O primeiro pode ser melhorado aumentando-se o tamanho da amostra. O segundo não pode ser diminuído dessa maneira, sendo necessário alterações operacionais e administrativas para sua diminuição.

Como foi visto, o viés devido à falta de respostas, um erro do tipo não amostral, supera o erro amostral medido pelos coeficientes de variação das estimativas. Portanto, antes de alterações no esquema amostral ou no tamanho da amostra, deve-se pensar em alterar as condições operacionais e administrativas sob as quais esse tipo de levantamento vem sendo feito.

Do exposto até aqui, conclui-se que os dados levantados apresentam sérios problemas estatísticos; por isso, devem ser utilizados com cautela. É importante notar, entretanto, que, se tais considerações não costumavam ser feitas nos levantamentos anteriores da SAA, nem costumam ser feitas nos levantamentos de outras instituições, isso não significa que tais dados sejam isentos de problemas estatísticos, mas apenas que o controle de qualidade deixou de

ser feito. Portanto, exceto em levantamentos em que o controle de qualidade foi feito e o respectivo relatório foi publicado, nada se pode dizer quanto à qualidade de seus dados.

#### 4 - CONCLUSÕES

Os dados levantados apresentam sérias restrições ao seu uso, principalmente devido ao elevado índice de falta de respostas (28,1% dos questionários, correspondendo a 22,4% em número de UPAs e 27,7% em área), que podem viesar seriamente os resultados, principalmente as estimativas dos totais, que estão subestimados. A seriedade do problema pode ser atestada quando se compara com 19 levantamentos realizados no período 1975-79, nos quais a falta de resposta oscilou entre 6% e 12%. Como a falta de resposta concentra-se em algumas regiões do Estado (notadamente os EDRs de Itapetininga, Andradina, Sorocaba, Itapeva e São Paulo), as variáveis cujos valores não sejam altos nessas regiões poderão ter estimativas razoáveis quanto a esse viés; com variáveis concentradas nessas regiões ocorre o contrário.

O principal motivo para falta de resposta foi a unidade não procurada (19,4% dos questionários). Seguem-se a ausência do informante (5,4%), o não entendimento do entrevistador (2,1%), a recusa do informante (0,8%), a não localização de unidades (0,3%) e a dificuldade de acesso à unidade (0,1%). Conclui-se que os problemas não foram de ordem técnico-científica, mas administrativa e operacional, tendo faltado: a) planejamento dos trabalhos; b) envolvimento dos participantes; c) seleção, qualificação e treinamento dos participantes.

A comparação do levantamento de 2000 com os do período 1978-79 permitiu concluir que o item unidade amostral não procurada praticamente dobrou sua participação como motivo da falta de resposta, provavelmente devido à degeneração institucional. Por outro lado, fatores como melhoramento de estradas, mudança do proprietário para a cidade e, principalmente, o investimento da SAA num cadastro de boa qualidade, levaram à diminuição da participação de motivos como unidade amostral não localizada

(quase 20 vezes menor), unidade amostral inacessível (cinco vezes menor), recusa do informante (quase quatro vezes menor) e informante ausente (quase duas vezes menor).

O período excessivamente longo verificado entre o levantamento no campo e a divulgação dos resultados compromete seriamente o interesse do agronegócio pelas estatísticas obtidas no levantamento analisado.

Para o futuro, sugere-se uma seqüência de medidas, a primeira das quais consiste na decisão política a respeito das estatísticas agrícolas:

- a) ou as estatísticas agrícolas não são importantes e deve-se parar de desperdiçar dinheiro do contribuinte;
- b) ou são importantes, e deve-se investir nelas.

Caso elas sejam consideradas importantes, os dirigentes responsáveis devem:

- a) fornecer recursos humanos, materiais e financeiros necessários e suficientes à realização dos levantamentos;
- b) promover o envolvimento de todos os níveis de direção necessários com a responsabilidade pelo trabalho;
- c) promover a seleção (ou contratação, se inexistirem) de funcionários com a competência necessária e suficiente para executar as diversas etapas do levantamento;
- d) promover a habilitação dos funcionários selecionados (ou contratados) nas funções apropriadas;
- e) promover o treinamento dos funcionários habilitados para o desempenho das tarefas relacionadas ao levantamento;
- f) planejar o levantamento e promover a integração entre as diferentes equipes de execução do levantamento, de tal forma que os resultados de cada etapa fluam perfeita e rapidamente para a etapa seguinte.

Também será imprescindível a atualização do atual cadastro, que já começa a dar sinais de envelhecimento.

Em suma, diante do fenômeno da globalização, talvez o agronegócio não possa esperar mais tempo pelas informações estatísticas agrícolas de boa qualidade sem perder dinheiro, custo esse que recairá tanto sobre o próprio setor como sobre o País.



## LITERATURA CITADA

FRANCISCO, V. L. F. S.; PINO, F. A. Estratificação de unidades de produção agrícola para levantamentos por amostragem no estado de São Paulo. **Agricultura em São Paulo**, São Paulo, v. 47, t. 1, p. 79-110, 2000.

\_\_\_\_\_. et al. Censo agropecuário no estado de São Paulo: resultados regionais. **Informações Econômicas**, São Paulo, v. 27, n. 11, p. 7-140, nov. 1997.

IBGE. Sidra. **Levantamento sistemático da produção agrícola**. Disponível em: <http://www.sidra.ibge.gov.br>. Acesso em: 2 abr. 2001.

IEA. **Previsões e estimativas das safras agrícolas do estado de São Paulo, ano agrícola 1999/2000, levantamento final, novembro de 2000**. Disponível em: <http://www.iea.sp.gov.br>. Acesso em: 2 abr. 2001a.

\_\_\_\_\_. **Previsões e estimativas das safras agrícolas do estado de São Paulo, ano agrícola 2000/2001, 2º levantamento, novembro de 2000**. Disponível em: <http://www.iea.sp.gov.br>. Acesso em: 2 abr. 2001b.

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

PINO, 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, t. 2, p. 147-153, 1989.

\_\_\_\_\_. Detecção e correção de erros em levantamentos agrícolas. **Pesquisa Agropecuária Brasileira**, Brasília, v. 21, n. 9, p. 979-985, set. 1986.

\_\_\_\_\_. Estatísticas agrícolas para o século XXI. **Agricultura em São Paulo**, São Paulo, v. 46, t. 2, p. 71-105, 1999.

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

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

\_\_\_\_\_; CAMARGO, M. L. B.; VIANI, D. N. **Mudanças de posse e tamanho das propriedades agrícolas do estado de São Paulo no período 1972-77**. São Paulo: IEA, 1979. 15 p. (Relatório de Pesquisa, 16/79).

\_\_\_\_\_; FRANCISCO, V. L. F. S.; LORENA NETO, B. Previsão e estimativa de safras cafeeiras no estado de São Paulo. **Agricultura em São Paulo**, São Paulo, v. 48, t. 1, p. 57-68, 2001.

PINO, F. A. et al. (Org.). **Levantamento censitário de unidades de produção agrícola no estado de São Paulo**. São Paulo: IEA/CATI/SAA, 1997. 4 v.

### **CONTROLE DE QUALIDADE EM LEVANTAMENTO AGRÍCOLA POR AMOSTRAGEM EM SÃO PAULO**

**RESUMO:** Apresenta-se o controle de qualidade dos dados estatísticos de um levantamento por amostragem de unidades de produção agrícola usados para estimar safras agrícolas e trabalho rural no Estado de São Paulo. Mostra-se que os resultados são fortemente viesados devido a um problema severo de falta de resposta (28% de não respondentes). Isso é muito maior do que em levantamentos passados e relaciona-se mais a questões administrativas e operacionais que a falhas técnicas ou estatísticas. Alguns procedimentos empíricos são utilizados para melhorar as estimativas da média, do total e

*Informações Econômicas, SP, v.31, n.6, jun. 2001.*

das respectivas variâncias, como a substituição dos valores dos não respondentes pela média dos respondentes ou usando os respondentes como uma amostra de menor tamanho. Apresentam-se algumas sugestões sobre como tratar com o problema no futuro.

**Palavras-chave:** amostra estratificada; estatísticas agrícolas; falta de resposta; vieses; erros amostrais.

#### **QUALITY CONTROL IN AGRICULTURAL SAMPLE SURVEY IN SAO PAULO, BRAZIL**

**ABSTRACT:** *The quality control of statistical data from a sample survey of farms used to estimate agricultural crops and rural labor in the State of Sao Paulo, Brazil, is presented. The results are shown to be strongly biased due to a severe non-response problem (28% of non-respondents). This is much higher than past surveys and is closely related to administrative and operational features rather than to technical or statistical failures. Some empirical devices are used to improve the estimates for the mean, the total and the respective variances, such as substituting the non-respondent values by the respondent average or using the respondents as a lower sized sample. Suggestions for dealing with the problem in the future are given.*

**Key-words:** *stratified sample, agricultural statistics, non-response, biases, sample errors.*

---

Recebido em 24/04/2001. Liberado para publicação em 18/05/2001.