

FALTA DE RESPOSTA EM LEVANTAMENTOS POR AMOSTRAGEM:
UM ESTUDO DE CASO

Francisco Alberto Pino e Denise Viani Caser

Governo do Estado de São Paulo
Secretaria de Agricultura e Abastecimento
Coordenadoria Sócio-Econômica

Instituto de Economia Agrícola



**FALTA DE RESPOSTA EM LEVANTAMENTOS POR AMOSTRAGEM:
UM ESTUDO DE CASO**

Francisco Alberto Pino
Denise Viani Caser

Í N D I C E

1 - INTRODUÇÃO	i
1.1 - Classificação da Falta de Resposta	2
1.2 - Métodos para Tratar a Falta de Resposta	3
1.2.1 - Cuidados no planejamento do levantamento	3
1.2.2 - Repasses	3
1.2.3 - Método de Folitz - Simmons	4
1.2.4 - Método proposto por Bartholomew	4
1.2.5 - Estudo especial do grupo que não responde	5
1.2.6 - Substituição dos não respondentes	6
1.2.7 - Outros estudos	6
1.3 - Objetivos	6
2 - ASPECTOS ESTATÍSTICOS DA FALTA DE RESPOSTA	7
2.1 - Aleatoriedade do Tamanho da Amostra	8
2.2 - Viés do Estimador da Média	9
3 - ESTUDO DE CASO	10
3.1 - Metodologia	10
3.2 - Resultados e Discussão	11
3.2.1 - Diferenças regionais	12
3.2.2 - Diferenças por tamanho do imóvel	12
3.2.3 - Diferenças ao longo do tempo	17
3.2.4 - Motivos da falta de resposta	17
3.3 - Conclusões	21
3.4 - Recomendações	21
4 - CONSIDERAÇÕES FINAIS	22
LITERATURA CITADA	22
RESUMO	24
SUMMARY	25

Francisco Alberto Pino
Denise Viani Caser

1 - INTRODUÇÃO

Os métodos de amostragem tornaram-se comuns na obtenção de dados estatísticos. Na economia agrícola tais métodos são especialmente importantes para levantar dados de preços e de produção, uma vez que a atividade agropecuária é realizada num grande número de empresas espalhadas por todo o país.

No Estado de São Paulo, desde 1953, estatísticas agrícolas são obtidas em amostras de imóveis rurais, através do preenchimento de questionários em entrevistas com os proprietários ou administradores. Entretanto, quando não se conseguem respostas para todos os elementos da amostra por alguma razão, surge um problema comum em amostragem: o da falta de resposta. Usa-se a expressão "falta de resposta" para designar a não obtenção de algumas ou de todas as medidas desejadas a respeito de alguma unidade amostral sorteada. Não há, a priori, nenhuma garantia de que o conjunto dos elementos que responderam forneça estimativas não viesadas. Logo, a falta de resposta é uma fonte de erros que pode invalidar os resultados do levantamento.

Evidentemente, o problema é menos preocupante quando se realiza apenas um levantamento numa área restrita e com um número pequeno de quesitos, porque, neste caso, um esforço adicional dos pesquisadores e entrevistadores pode minimizá-lo. Entretanto, quando, como no caso dos levantamentos para previsão e estimativa de safras agrícolas do Estado de São Paulo, os levantamentos são feitos cinco vezes por ano, durante muitos anos, para algumas centenas de itens, o problema não pode ser desprezado. As causas precisam ser descobertas e analisadas para que mecanismos de controle sejam acionados em cada levantamento. Mesmo porque, o custo do levantamento aumenta devido à falta de respostas, pois, obriga o uso de procedimentos mais complicados (5).

1.1 - Classificação da Falta de Resposta

Quanto às suas causas, a falta de resposta pode ser assim classificada:

a) Unidade amostral não localizada

A unidade amostral sorteada não é encontrada. Neste caso, se-quer o endereço do informante é localizado. Isto pode ser devido a imperfeições no sistema referencial de amostragem (cadastro, rol, relação, etc.), a partir do qual a amostra foi sorteada, ou a mudanças na unidade amostral (por exemplo, morte de indivíduos, extinção de empresas, etc.), ou à falta de conhecimento da região por parte do entrevistador, etc.

b) Unidade amostral inacessível

É impossível o acesso à unidade amostral sorteada, embora se conheça sua localização. Essa impossibilidade pode ser temporária ou permanente. As principais causas são as condições meteorológicas adversas e as deficiências dos meios de transporte. Por exemplo, no caso de a unidade amostral ser um imóvel rural, é comum que algumas estradas não pavimentadas fiquem intransitáveis nas épocas chuvosas.

c) Informante ausente

O elemento a ser entrevistado não é encontrado em seu endereço. Neste caso, a unidade amostral sorteada é encontrada, mas, o indivíduo a ser entrevistado encontra-se temporariamente ausente.

d) Ignorância do informante

O entrevistado não sabe responder. Às vezes o entrevistado não dispõe da informação solicitada ou não tem suficiente grau de instrução para responder a determinadas perguntas.

e) Recusa do informante

O entrevistado recusa-se a responder. Razões como costumes, desconfiança e falta de tempo podem levar o entrevistado a negar sua colaboração com o levantamento.

f) Perda de informação

Perda de questionário após a entrevista. É mais freqüente quando os questionários são remetidos através dos correios ou meio similar, ao invés de serem entregues diretamente.

1.2 - Métodos para Tratar a Falta de Resposta

Cada tipo de falta de resposta requer um tipo de tratamento específico para que seu efeito possa ser minimizado. Encontram-se na literatura vários métodos para tratar do problema da falta de resposta, sendo os principais apresentados a seguir (14).

1.2.1 - Cuidados no planejamento do levantamento

Um levantamento bem planejado pode evitar muitos dos problemas que geram falta de resposta. Assim, a questão da unidade amostral não localizada pode ser resolvida usando-se um sistema referencial de amostragem de boa qualidade e atualizado, bem como entrevistadores qualificados. A adequação da época do levantamento pode reduzir os casos de unidades amostrais inacessíveis. Entrevistadores persuasivos podem diminuir os casos de recusa. Um questionário bem elaborado pode, eventualmente, contornar o problema de ignorância do informante, bem como pode ajudar a diminuir os casos de recusa, pois, pode se dever a um questionário excessivamente longo. Mudança periódica das unidades amostrais sorteadas é recomendável, pois, a recusa também pode ser devida ao desgaste do informante.

1.2.2 - Repasses

Uma prática comum no caso de informante ausente consiste em visitar a unidade amostrada diversas vezes até encontrar o elemento a ser entrevistado, ou pelo menos, um número mínimo de vezes estabelecido a priori (normalmente três vezes) antes de desistir. DURBIN e STUART (9) mostram que tal procedimento não leva a vieses sérios. Experiências citadas por COCHRAN (5) sugerem que o custo adicional das

novas visitas é bem menor do que seria de se esperar.

1.2.3 - Método de Politz-Simmons

A causa básica do viés que surge quando o elemento a ser entrevistado não é encontrado é que a probabilidade de o elemento ser achado está correlacionada com as variáveis do levantamento. Com base nisso, POLITZ & SIMMONS (15) propuseram que o questionário deveria conter perguntas que permitissem estimar tal probabilidade (ver também BARTHOLOMEW (1) e RAJ (16)). Pergunta-se ao entrevistado se ele estava em casa à hora da entrevista nos últimos cinco ou seis dias. Esta informação pode ser usada para estratificar os elementos da amostra de acordo com sua chance de ser encontrado em casa, o que permite obter estimativas das quais foi removida a maior parte do viés.

Algumas desvantagens deste método são as seguintes:

a) o método exige um esforço de memória dos entrevistados, o que pode levar a imprecisão nas respostas;

b) a variância é aumentada. Entretanto, se se utilizarem duas ou três visitas ao invés de somente uma, o aumento da variância é bem menor; e

c) o custo pode ser bastante aumentado por se tornarem a execução e a análise mais complicadas.

DURBIN e STUART (9) usaram este método, mas, sem conseguir resultados muito bons, enquanto RAJ (16) conseguiu resultados interessantes.

1.2.4 - Método proposto por Bartholomew

Suponhamos que numa amostra de tamanho n uma dada característica da população toma valores X_1, \dots, X_n . Queremos estimar a média μ da população utilizando duas visitas. Seja n_1 o número de respostas da primeira visita e n_2 o número de respostas na segunda visita. Então,

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \bar{X}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i \quad \text{e} \quad \lambda = \frac{1}{n - n_1} \sum_{i=n_1+1}^n X_i,$$

onde λ é a média dos que não responderam na primeira visita. A média da amostra fica sendo:

$$\bar{X} = \frac{n_1}{n} \bar{X}_1 + \left(1 - \frac{n_1}{n}\right) \lambda$$

Suponhamos que a segunda visita é efetivamente aleatória e produz as observações $X_{n_1+1}, \dots, X_{n_1+n_2}$. Se o viés aparecer somente na primeira tentativa, então, pode-se estimar λ por,

$$\hat{\lambda} = \frac{1}{n_2} \sum_{i=n_1+1}^{n_1+n_2} X_i,$$

obtendo-se

$$\hat{\bar{X}} = \frac{n_1}{n} \bar{X}_1 + \left(1 - \frac{n_1}{n}\right) \hat{\lambda}$$

Usando este método, BARTHOLOMEW (1) conseguiu bons resultados, tendo sido pequeno o aumento da variância. Entretanto, é bom atentar para as duas suposições básicas:

a) a segunda visita é aleatória, i.e., os elementos que respondem na segunda tentativa constituem uma amostra aleatória da população; e

b) a maior parte do viés deve-se à falta de resposta na primeira tentativa.

1.2.5 - Estudo especial do grupo que não responde

Um estudo preliminar consiste em anotar as causas da falta de resposta em cada um dos questionários não respondidos, estudando, depois, soluções que podem ser diferentes para cada causa.

Num levantamento periódico, que se estende por longo período de tempo, pode ser compensador o esforço maior de um estudo especial dos que não respondem, por exemplo, tomando-se uma subamostra dos que não responderam e tentando-se nova entrevista com entrevistadores mais qualificados. O questionário pode, inclusive, ser reduzido a um número de itens essenciais que permitam testar se há diferença significativa entre as respostas dos dois grupos, o de respondentes e o de não respondentes.

Se não houver diferença significativa pode-se, por exemplo, substituir os elementos que não responderam por outros escolhidos aleatoriamente, ou mesmo, sortear uma amostra de tamanho maior que o necessário, de modo a compensar as faltas de respostas.

1.2.6 - Substituição dos não respondentes

A substituição dos elementos que não respondem por outros pode, às vezes, ser feita, conforme KISH (12), possibilitando o controle exato do tamanho da amostra e facilitando o trabalho de campo caso se use uma amostra maior que o necessário para compensar a falta de respostas. Para fazer a substituição divide-se a amostra em subclasses internamente homogêneas segundo as variáveis do levantamento, caso contrário, cria-se um viés.

BIRNBAUM & SIRKEN (3) apresentaram um processo para calcular o tamanho da amostra quando há falta de respostas, no caso em que a proporção dos que não respondem é conhecida ou pode ser estimada a partir de levantamentos anteriores (ver também COCHRAN (5)).

Entretanto, as desvantagens da substituição normalmente superam as vantagens, podendo ser usada apenas em casos específicos.

1.2.7 - Outros estudos

Outros estudos sobre o assunto são os de DEMING (7), DURBIN (8), HANSEN & HURWITZ (10), HENDRICHS (11), KISH & HESS (13) e SIMMONS (17).

1.3 - Objetivos

São objetivos deste trabalho:

a) discutir algumas implicações estatísticas da falta de resposta em levantamentos por amostragem; e

b) apresentar um estudo de caso, o dos levantamentos para previsão e estimativa de safras do Estado de São Paulo.

Na apresentação do estudo de caso pretende-se encontrar as principais razões da falta de resposta, testar diferenças regionais e diferenças relativas ao tamanho do imóvel, bem como propor soluções.

2 - ASPECTOS ESTATÍSTICOS DA FALTA DE RESPOSTA

Do ponto de vista estatístico, as duas principais conseqüências da falta de resposta são: a) o tamanho da amostra torna-se aleatório; e b) os estimadores podem se tornar viesados. Vamos introduzir uma notação para, a seguir, estudar essas duas conseqüências.

A falta de resposta induz uma partição no espaço amostral, dividindo-o em dois conjuntos: a) o conjunto dos elementos que fornecem respostas; e, b) o conjunto dos elementos que não fornecem respostas. Consideremos uma população finita na qual:

N é o número total de elementos,
 N_1 é o número de elementos que fornecem respostas,
 N_2 é o número de elementos que não fornecem respostas,
 p é a proporção de elementos que fornecem respostas e,
 q é a proporção de elementos que não fornecem respostas,
isto é,

$$N = N_1 + N_2$$

$$p = 1 - q = N_1/N$$

Suponhamos que uma amostra casual simples é retirada desta população, onde:

n é o número total de elementos da amostra,
 n_1 é o número de elementos da amostra que fornecem respostas e,
 n_2 é o número de elementos da amostra que não fornecem respostas,
isto é,

$$n = n_1 + n_2$$

Seja X a variável aleatória que representa uma dada característica da população e $p_\theta(x)$ a sua distribuição (em geral, admite-se que X tem distribuição normal com média μ e variância σ^2). Supomos, ainda, que:

μ é a média populacional de X ,
 μ_1 é a média de X no conjunto dos elementos que fornecem respostas,
 μ_2 é a média de X no conjunto dos elementos que não fornecem respostas, e

\bar{X} , \bar{X}_1 e \bar{X}_2 são as correspondentes médias amostrais.

2.1 - Aleatoriedade do Tamanho da Amostra

Com jã foi dito, uma das conseqüências da falta de resposta é que o tamanho da amostra torna-se aleatório. A questão é saber se este fato causa algum problema do ponto de vista estatístico.

Seja Y a variável aleatória que representa o tamanho atual da amostra, isto é, o número de elementos da amostra que forneceram respostas. Então, Y tem distribuição hipergeométrica com função densidade da da por

$$f_Y(y; N, N_1, n) = \frac{\binom{N_1}{y} \binom{N_2}{n-y}}{\binom{N}{n}}, \text{ para } y = 0, 1, \dots, n$$

e igual a zero em caso contrário. Tem-se que

$$E(Y) = n \cdot \frac{N_1}{N} = np$$

$$V(Y) = n \cdot \frac{N_1}{N} \cdot \frac{N_2}{N} \cdot \frac{N-n}{N-1} = n \cdot p \cdot q \cdot \frac{N-n}{N-1},$$

isto é, a média de Y coincide com a de uma distribuição binomial, de parâmetros n e p e sua variância é igual a $\frac{N-n}{N-1}$ vezes a variância da

sa distribuição binomial.

Podemos estimar p e N_1 , respectivamente, por

$$\bar{p} = n_1/n \text{ e,}$$

$$\hat{N}_1 = \bar{p} \cdot N$$

Como conseqüência, $T = (Y, \bar{X}_1, \sum_{i=1}^Y X_i^2)$, é uma estatística su-

ficiente, porém, não é completa, uma vez que Y é uma estatística ancilar e a dimensão do espaço paramétrico é menor que a dimensão da estatística. O parâmetro de maior interesse é a média μ da população, e a variância de seu estimador fica sendo

$$V(\bar{X}_1) = V \left(\frac{1}{Y} \sum_{i=1}^Y X_i \right)$$

$$\begin{aligned}
&= E \left[V \left(\left. \frac{1}{Y} \sum_{i=1}^Y X_i \right| Y \right) \right] + V \left[E \left(\left. \frac{1}{Y} \sum_{i=1}^Y X_i \right| Y \right) \right] \\
&= E \left[\frac{1}{Y^2} \sum_{i=1}^Y V(X_i) \right] + V \left[\frac{1}{Y} \sum_{i=1}^Y E(X_i) \right] \\
&= E \left[\frac{\sigma^2}{Y} \right] + V(\mu) = \sigma^2 \cdot E \left[\frac{1}{Y} \right]
\end{aligned}$$

Entretanto, se é aceito o princípio da condicionalidade (2 e 5), o qual diz que a informação contida no experimento original sobre o parâmetro de interesse permanece a mesma quando se condiciona o experimento ao valor obtido para a estatística ancilar, então, toma-se a distribuição condicional $p_{\theta}(x|y)$ e, agora,

$$V(\bar{X}_1 | Y = y) = \sigma^2/y$$

e $T^* = (\bar{X}_1, \sum_{i=1}^Y X_i^2)$ é uma estatística condicionalmente suficiente e completa, \bar{X}_1 é o estimador de μ e o impasse é resolvido

2.2 - Viés do Estimador da Média

O viés do estimador \bar{X}_1 é dado por

$$\begin{aligned}
\text{viés}(\bar{X}_1) &= \mu - E(\bar{X}_1) \\
&= \frac{N_1 \mu_1 + N_2 \mu_2}{N} - \mu_1 \\
&= \frac{N_1 \mu_1 + N_2 \mu_2 - N\mu_1}{N} \\
&= \frac{N_2 \mu_2 - N_2 \mu_1}{N}
\end{aligned}$$

$$= \frac{N_2}{N} (\mu_2 - \mu_1)$$

$$= q (\mu_2 - \mu_1)$$

Logo, se $\mu_1 = \mu_2$, a estimativa da média, utilizando-se apenas as respostas fornecidas, não é viesada e, como

$$V(\bar{X}) = \frac{\sigma^2}{n} \leq \frac{\sigma^2}{n_1} = V(\bar{X}_1 \mid Y = n_1),$$

o único efeito da falta de respostas será o de aumentar a variância da estimativa, uma vez que o tamanho da amostra é diminuído. Por outro lado, se o viés ocorrer, ele será diretamente proporcional à proporção da falta de respostas.

Na prática, para saber se o problema de viés existe, é necessário dispender um esforço adicional, com entrevistadores de melhor qualificação do que os usuais, para obter uma estimativa entre as unidades da amostra para as quais houve falta de resposta. Aplica-se, então, o teste t de Student para diferença de médias entre \bar{X}_1 e \bar{X}_2 .

3 - ESTUDO DE CASO

Apresenta-se a seguir o caso da amostra para previsão e estimativa de safras agrícolas do Estado de São Paulo.

3.1 - Metodologia

As previsões e estimativas de safras do Estado de São Paulo são efetuadas pela Secretaria da Agricultura e Abastecimento deste Estado e baseiam-se numa amostra de imóveis rurais, duplamente estratificada, por região (Divisão Regional Agrícola, abreviadamente DIRA) e por tamanho do imóvel (estrato de área), conforme descrito em CAMPOS & PIVA (4), nos meses de setembro, novembro, fevereiro, abril e junho.

No presente trabalho utilizaram-se 19 levantamentos, de junho de 1975 a fevereiro de 1979. De junho de 1975 a junho de 1977 a amostra constava de 6.229 elementos, enquanto no restante do período ela constava de 5.646 elementos.

A variável em estudo é a percentagem de unidades da amostra que forneceram resposta em relação ao tamanho da amostra, que, para estabilizar a variância, foi transformada (18) através de:

$$y = \text{arc sen } (x/100)^{1/2},$$

onde x é o dado original em percentagem e y é o dado transformado.

Fizeram-se comparações entre DIRAs (D) e entre estratos de área (E), tendo os sucessivos levantamentos como bloco (B), utilizando análise de variância e teste de Duncan. Utilizou-se o seguinte modelo hierárquico:

$$Y_{ijk} = \mu + D_i + E_{j(i)} + B_k + e_{ijk}$$

com $i = 1, 2, \dots, 10$

$j(i) = 1, 2, \dots, 11$ para cada i ,

$k = 1, 2, \dots, 19$

Testaram-se as seguintes hipóteses nulas:

a) não existe efeito de DIRA sobre a percentagem de questionários respondidos:

$$D_1 = \dots = D_{10} = 0$$

b) não existe efeito de estrato dentro da DIRA sobre a percentagem de questionários respondidos:

$$E_{1(1)} = \dots = E_{11(10)} = 0;$$

c) não existe efeito de bloco (levantamento):

$$B_1 = \dots = B_{19} = 0.$$

Os motivos da falta de resposta foram analisados somente para três levantamentos: junho de 1978, 79 e 80. Tomaram-se três anos para captar variações no tempo, utilizando-se os anos mais recentes para os quais se dispunha de informação. O levantamento de junho foi o escolhido por ser o mais importante do ano agrícola.

3.2 - Resultados e Discussão

3.2.1 - Diferenças regionais

O quadro 1 traz a situação média no período.

Dos quadros 2 e 3 conclui-se que a falta de respostas é um problema mais sério em algumas DIRAs do que em outras. O índice de falta de respostas é maior no leste e no sul do Estado, isto é, nas DIRAs São Paulo, Vale do Paraíba e Sorocaba.

3.2.2 - Diferenças por tamanho do imóvel

Os resultados dos quadros 2 e 4 mostram que a falta de respostas varia com o tamanho do imóvel dentro de cada DIRA. Em média o índice de falta de respostas cresce com o tamanho do imóvel, podendo a grosso modo, classificar-se os estratos em ordem crescente da seguinte maneira:

- a) imóveis de 5 a 20ha;
- b) imóveis de 20 a 100ha;
- c) imóveis de 3 a 5ha;
- d) imóveis de 100 a 500ha;
- e) imóveis de mais de 1000ha.

Entretanto, a influência do tamanho do imóvel sobre a falta de respostas varia entre regiões, uma vez que variam as características do imóvel e do processo produtivo conforme a DIRA.

QUADRO 1. - Percentagem Média de Questionários Respondidos, Levantamento para Previsão e Estimativa de Safras, Estado de São Paulo, junho de 1975 a fevereiro de 1979

DIRA	Percentagem
Bauru	94,8
Ribeirão Preto	94,6
Marília	94,3
Presidente Prudente	93,7
Campinas	92,5
Araçatuba	92,3
São José do Rio Preto	91,0
Sorocaba	89,8
Vale do Paraíba	88,0
São Paulo	82,2

Fonte: Instituto de Economia Agrícola (IEA).

QUADRO 2. - Análise de Variância da Percentagem Transformada do Número de Respostas

Fontes de variação	G.L.	S.Q.	Q.M.	F
DIRA	9	9,1937	1,0215	114,798***
Estrato	100	8,0253	0,0803	9,019***
Bloco	18	2,1891	0,1216	13,667***
Resíduo	1962	17,4588	0,0089	
Total	2089	36,8668		

***: significativo ao nível de 0,5%.

QUADRO 3. - Teste de Duncan para Diferenças entre DIRAs

DIRA	Percentagem média transformada
Ribeirão Preto	1,372
Bauru	1,370
Marília	1,369
Presidente Prudente	1,336
Araçatuba	1,313
Campinas	1,306
Sao José do Rio Preto	1,278
Sorocaba	1,258
Vale do Paraíba	1,245
São Paulo	1,151

As médias unidas por barras não diferem significativamente entre si ao nível de 5%.

QUADRO 4. - Teste de Duncan para Diferenças entre Estratos dentro de DIRAs

(continua)

Ribeirão Preto		Bauru		Marília		Presidente Prudente	
Estrato (ha)	Média	Estrato (ha)	Média	Estrato (ha)	Média	Estrato (ha)	Média
5 - 10	1,513	10 - 20	1,482	5 - 10	1,543	30 - 50	1,431
10 - 20	1,491	30 - 50	1,450	3 - 5	1,467	500 - 1000	1,415
20 - 30	1,419	3 - 5	1,415	20 - 30	1,419	5 - 10	1,350
3 - 5	1,392	20 - 30	1,412	30 - 50	1,399	300 - 500	1,334
50 - 100	1,375	50 - 100	1,395	50 - 100	1,398	20 - 30	1,325
30 - 50	1,366	500 - 1000	1,354	200 - 300	1,359	50 - 100	1,325
200 - 300	1,353	200 - 300	1,339	100 - 200	1,349	100 - 200	1,325
300 - 500	1,319	100 - 200	1,326	10 - 20	1,345	10 - 20	1,320
100 - 200	1,317	5 - 10	1,309	300 - 500	1,328	3 - 5	1,319
500 - 1000	1,313	+ de 1000	1,296	+ de 1000	1,263	200 - 300	1,301
+ de 1000	1,232	300 - 500	1,292	500 - 1000	1,194	+ de 1000	1,255

As médias unidas por barras não diferem significativamente entre si ao nível de 5%.

QUADRO 4. - Teste de Duncan para Diferenças entre Estratos dentro de DIRAs

(continua)

Araçatuba		Campinas		São José do Rio Preto	
Estrato (ha)	Média	Estrato (ha)	Média	Estrato (ha)	Média
10 - 20	1,398	10 - 20	1,361	30 - 50	1,342
100 - 200	1,378	3 - 5	1,350	20 - 30	1,334
50 - 100	1,366	5 - 10	1,340	50 - 100	1,332
3 - 5	1,359	50 - 100	1,336	100 - 200	1,309
20 - 30	1,298	30 - 50	1,326	10 - 20	1,284
300 - 500	1,298	20 - 30	1,325	3 - 5	1,284
30 - 50	1,291	200 - 300	1,305	5 - 10	1,246
5 - 10	1,285	100 - 200	1,280	+ de 1000	1,235
500 - 1000	1,276	300 - 500	1,268	500 - 1000	1,234
200 - 300	1,252	500 - 1000	1,261	200 - 300	1,228
+ de 1000	1,239	+ de 1000	1,219	300 - 500	1,226

As médias unidas por barras não diferem significativamente entre si ao nível de 5%.

QUADRO 4. - Teste de Duncan para Diferenças entre Estratos dentro de DIRAs

(conclusão)

Sorocaba		Vale do Paraíba		São Paulo	
Estrato (ha)	Média	Estrato (ha)	Média	Estrato (ha)	Média
50 - 100	1,319	5 - 10	1,339	300 - 500	1,246
30 - 50	1,300	500 - 1000	1,336	50 - 100	1,219
100 - 200	1,282	10 - 20	1,329	30 - 50	1,198
200 - 300	1,279	100 - 200	1,307	5 - 10	1,183
300 - 500	1,274	300 - 500	1,267	20 - 30	1,176
20 - 30	1,260	30 - 50	1,236	3 - 5	1,174
10 - 20	1,239	50 - 100	1,235	10 - 20	1,146
3 - 5	1,235	3 - 5	1,232	100 - 200	1,118
500 - 1000	1,230	200 - 300	1,192	+ de 1000	1,102
5 - 10	1,224	20 - 30	1,174	500 - 1000	1,061
+ de 1000	1,193	+ de 1000	1,049	200 - 300	1,035

As médias unidas por barras não diferem significativamente entre si ao nível de 5%.

3.2.3 - Diferenças ao longo do tempo

Pelos resultados do quadro 2 percebe-se que há efeito significativo do bloco, isto é, a falta de resposta varia ao longo do tempo. Pela figura 1 percebe-se que no período estudado essa variação ao longo do tempo pode se dever a tendência e também sazonalidade, uma vez que a falta de respostas sempre decai em junho, quando se faz o levantamento final da colheita para a maioria das culturas, e sobe em setembro, que marca o início do ano agrícola.

3.2.4 - Motivos da falta de resposta

A análise do quadro 5 mostra que, no Estado como um todo, a ausência do informante tem sido a principal causa de falta de resposta, respondendo por cerca de um terço do problema. Isto pode ser solucionado com relativa facilidade através de repasses, que normalmente não são feitos. Os maiores índices de ausência aconteceram na DIRA Vale do Paraíba.

A segunda causa tem sido a não localização de unidades amostrais sorteadas, o que pode ser explicado pelas falhas do cadastro utilizado, que não foi construído para tal finalidade. Esta causa tem sido mais importante nas DIRAs São Paulo e Sorocaba.

Os motivos não especificados também têm pesado muito, o que poderia ser evitado melhorando-se a qualificação dos entrevistadores e com certas medidas administrativas, uma vez que os entrevistadores são funcionários responsáveis também por inúmeras outras tarefas. Além disso, em algumas regiões o número de funcionários é insuficiente.

A recusa em responder tem sido responsável por cerca de 10% dos casos, o que pode ser considerado razoável, devido ao desgaste, uma vez que os informantes permaneceram os mesmos em oito levantamentos. Os maiores índices de recusa aconteceram na DIRA Campinas.

As dificuldades de acesso às unidades sorteadas mostrou-se importante apenas na DIRA São Paulo.

Entre os outros motivos incluem-se desmembramentos, partilhas, anexações, vendas, inundação por barragens, propriedades abandonadas, etc.

Os números de junho de 1979 do quadro 5 estão especificados no quadro 6 por estrato de área do imóvel, verificando-se que o problema da falta de resposta foi mais sério nos imóveis entre 50 e 1000ha, prin

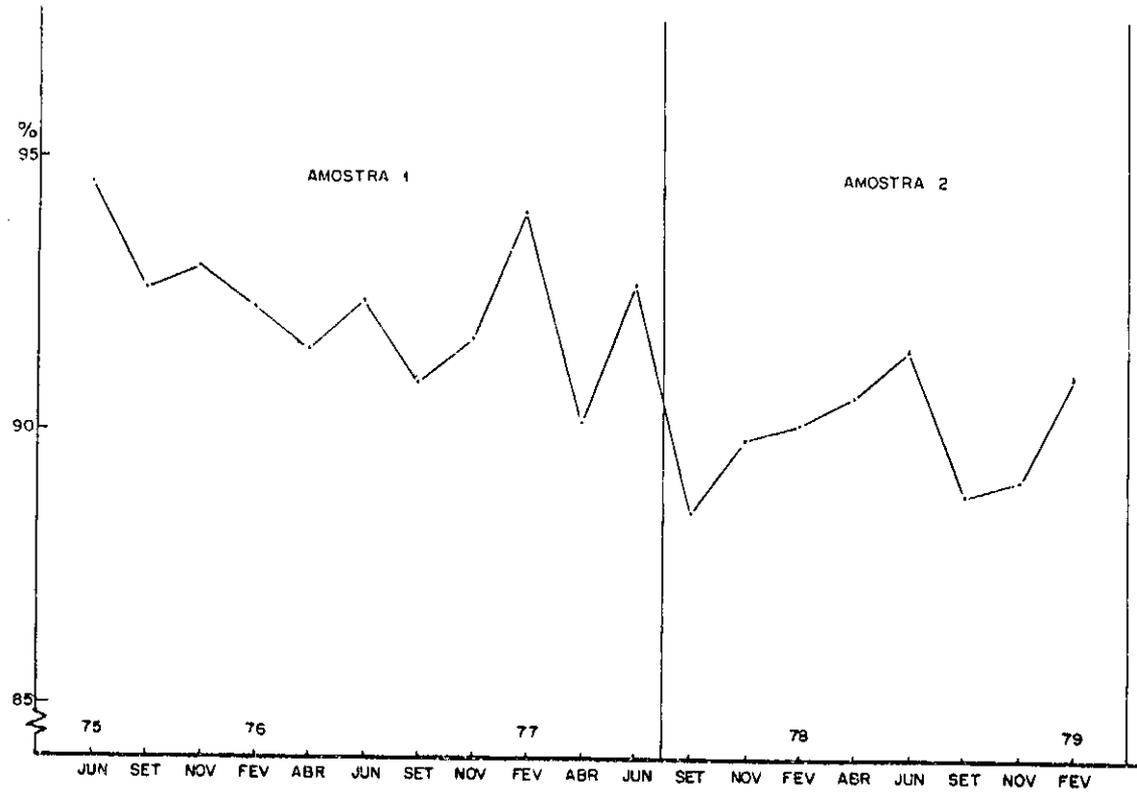


FIGURA 1 - Evolução da percentagem média de questionários respondidos por levantamento, junho de 1975 a fevereiro de 1979.

QUADRO 5. - Motivos de Falta de Resposta por DIRA, Junho de 1978, 1979 e 1980
(em percentagem)

DIRA	Ano	Unidade não localizada	Unidade inacessível	Informante ausente	Recusa	Outros motivos	Motivo não especificado	Total
São Paulo	1978	34,09	85,71	17,39	7,41	6,45	10,45	20,19
	1979	35,23	69,23	15,20	3,77	-	-	14,90
	1980	36,36	75,00	25,27	5,26	2,70	6,12	18,91
Vale do Paraíba	1978	10,22	-	23,91	7,41	9,68	7,46	13,14
	1979	7,95	-	20,25	7,55	8,82	2,06	10,84
	1980	10,39	-	20,88	5,26	8,11	13,27	12,90
Sorocaba	1978	25,00	-	21,74	7,41	12,90	13,43	18,27
	1979	25,00	-	14,56	13,21	5,88	12,37	14,90
	1980	22,08	-	5,49	-	10,81	27,55	15,19
São José do Rio Preto	1978	9,09	-	9,78	11,11	12,90	8,96	9,62
	1979	7,95	23,08	13,92	5,66	5,88	9,28	10,38
	1980	7,79	12,50	13,20	7,88	10,81	12,25	10,89
Campinas	1978	3,41	-	7,61	18,51	16,13	16,42	9,94
	1979	5,68	-	13,92	18,87	14,71	16,50	13,09
	1980	3,90	-	8,79	26,32	10,81	7,15	9,17
Araçatuba	1978	4,55	-	9,78	7,41	12,90	5,97	7,37
	1979	5,68	-	4,43	7,55	17,65	14,43	8,13
	1980	7,79	-	5,49	18,42	18,92	6,12	8,88
Presidente Prudente	1978	3,41	14,29	1,09	7,41	-	2,98	2,88
	1979	2,28	7,69	1,27	5,66	-	14,43	4,96
	1980	2,60	12,50	4,40	10,54	2,70	5,10	4,87
Marília	1978	6,82	-	1,09	14,81	3,23	14,93	7,05
	1979	5,68	-	1,90	11,32	2,94	19,59	7,67
	1980	5,19	-	5,49	7,89	5,41	11,22	7,16
Bauru	1978	3,41	-	1,09	7,41	9,68	7,46	4,49
	1979	3,41	-	5,06	5,66	26,47	6,19	6,55
	1980	2,60	-	3,30	-	16,22	6,12	4,87
Ribeirão Preto	1978	-	-	6,52	11,11	16,13	11,94	7,05
	1979	1,14	-	9,49	20,75	17,65	5,15	8,58
	1980	1,30	-	7,69	18,42	13,51	5,10	7,16
Estado	1978	28,21	2,24	29,49	8,65	9,94	21,47	100,00
	1979	19,86	2,94	35,67	11,96	7,67	21,90	100,00
	1980	22,06	2,30	26,07	10,89	10,60	28,08	100,00

Fonte: Instituto de Economia Agrícola (IEA).

QUADRO 6. - Motivos de Falta de Resposta por Estrato de Área do Imóvel, junho de 1979

(em percentagem)

Estrato de área (ha)			Unidade não localizada	Unidade inacessível	Informante ausente	Recusa	Outros motivos	Motivo não especificado	Total
3,1	a	5,0	3,41	7,69	3,16	3,77	8,82	2,06	3,61
5,1	a	10,0	15,91	-	5,70	3,77	-	2,06	6,10
10,1	a	20,0	17,05	7,69	3,16	-	11,76	7,22	6,32
20,1	a	30,0	11,36	-	9,49	1,89	8,82	10,31	8,80
30,1	a	50,0	9,09	7,69	10,76	7,55	2,94	7,22	8,58
50,1	a	100,0	13,64	38,46	15,19	15,09	20,59	16,49	16,25
100,1	a	200,0	11,36	7,69	13,92	9,43	14,71	12,37	12,42
200,1	a	300,0	9,09	7,69	9,49	16,98	11,76	9,28	10,38
300,1	a	500,0	3,41	-	12,66	13,21	8,82	8,25	9,26
500,1	a	1000,0	4,55	23,08	11,39	20,75	8,82	10,31	11,06
acima	de	1000,0	5,68	-	5,06	7,55	2,94	14,43	7,22
Total			19,87	2,93	35,67	11,96	7,67	21,90	100,0

principalmente no que diz respeito à ausência do informante e à recusa. Já a não localização da unidade sorteada foi mais frequente em imóveis entre 5 e 300ha.

3.3 - Conclusões

Do estudo do caso apresentado conclui-se:

- a) o índice de falta de resposta é maior nas regiões leste e sul do Estado, especialmente no litoral e no Vale do Ribeira;
- b) em geral o índice de falta de resposta cresce com o tamanho do imóvel;
- c) parece haver tendência no tempo e sazonalidade no índice de falta de resposta, caindo no final da safra e subindo no início;
- d) as principais causas de falta de resposta têm sido a ausência do informante e a não localização da unidade amostral sorteada;
- e) a maior incidência de ausência do informante tem se dado na DIRA do Vale do Paraíba, o de unidade não localizada nas DIRAs São Paulo e Sorocaba, e o de recusa na DIRA Campinas, e o de unidade amostral inacessível na DIRA São Paulo.

3.4 - Recomendações

Do exposto, ficam as seguintes recomendações para minimizar o problema da falta de resposta nos levantamentos para previsão e estimativa de safras agrícolas do Estado de São Paulo:

- a) alocar pessoal em número suficiente e com a qualificação adequada para as entrevistas com os informantes;
- b) fornecer meios aos entrevistadores para que possam ser feitos repasses;
- c) melhorar o contato entre o pessoal envolvido no levantamento: os que planejam a amostra e o levantamento, os entrevistadores e os que analisam os resultados;
- d) elaborar um sistema referencial de amostragem melhor que o atual. Vale a pena investir neste ponto, pois, o custo das operações seguintes do levantamento poderá ser diminuído;

e) aumentar a frequência com que as unidades amostrais são trocadas. Por exemplo, trocar metade da amostra todo ano, de tal modo que cada informante permaneça na amostra no máximo 2 anos;

f) fazer chegar ao informante o resultado final para o qual ele contribuiu; e

g) reformular os questionários, procurando eliminar ou modificar questões que o informante terá dificuldade para responder ou para as quais não se venham obtendo bons resultados estatísticos.

4 - CONSIDERAÇÕES FINAIS

Em resumo, a falta de respostas pode ser fonte séria de vieses que comprometem ou invalidam os resultados de um levantamento. Alguns cuidados no planejamento e na execução do levantamento podem restringir bastante o problema. Vários procedimentos como repasses, substituições e outros artifícios podem reduzir o viés até um nível aceitável. Entretanto, todos eles levam a um aumento maior ou menor na variância do estimador e nos custos. Não existem soluções gerais, sendo a minimização desses três fatores - viés, variância e custo - o que deve ser levado em conta para se conseguirem bons resultados num levantamento em que há falta de respostas.

LITERATURA CITADA

1. BARTHOLOMEW, D.J. A method of allowing for 'not-at-home' bias in sample surveys. Applied Statistics, 10:52-59, 1961.
2. BASU, D. Statistical information and likelihood. Sankhya, India, 37: 1-71, 1975. Série A.
3. BIRNBAUM, Z.W. & SIRKEN, M.G. Bias due to nonavailability in sampling surveys. Journal of the American Statistical Association, Washington, 45:98-111, 1950.

4. CAMPOS, Humberto & PIVA, Luiz H.O. Dimensionamento de amostra para estimativa e previsão de safra no Estado de São Paulo. Agricultura em São Paulo, SP, 21(3):65-68, 1974.
5. COCHRAN, W.G. Técnicas de amostragem. Rio de Janeiro, Fundo de Cultura, 1965. 555p.
6. COX, D.R. & HINKLEY, D.V. Theoretical statistics. London, Chapman and Hall, 1973. 511p.
7. DEMING, W.E. On a probability mechanism to attain an economic balance between the resultant error of non-response. Journal of the American Statistical Association, Washington, 48:743-772, 1953.
8. DURBIN, J. Non-response and callbacks in surveys. Bulletin of the International Statistical Institute, Holanda, 34(2):72-86, 1954.
9. DURBIN, J. & STUART, A. Callbacks and clustering in sample surveys: an experimental study. Journal of the Royal Statistical Society, London, 117:387-428, 1954. Série A.
10. HANSEN, M.H. & HURWITZ, W.N. The problem of non-response in sample surveys. Journal of the American Statistical Association, 41:517-529, 1946.
11. HENDRICKS, W.A. Adjustment for bias by non-response in mailed surveys. Agricultural Economics Research, Washington, 1:52-56, 1949.
12. KISH, Leslie. Survey sampling. 2.ed. New York, John Wiley, 1967. 643p.
13. KISH, L. & HESS, I. A'replacement' procedure for reducing the bias of non-response. American Statistician, Washington, 13(4):17-19, 1959.
14. PINO, F.A. & SEVER, F.A.A. Falta de respostas em levantamentos por amostragem. S.L., s.ed., s.d. 19p.

15. POLITZ, A.N. & SIMMONS, W.R. An attempt to get the "not-at-homes" into the sample without callbacks. Journal of the American Statistical Association, Washington, 44:9-31, 1949.
16. RAJ, D. Sampling theory. 3.ed. New York, Mc Graw-Hill, 1968. 302p.
17. SIMMONS, W.R. A plan to account for "not-at-homes" by combining weighting and callbacks. Journal of Marketing, Chicago, 11: 42-53, 1954.
18. SIQUEIRA, A.L. Uso de transformação em análise de variância e análise de regressão. São Paulo, Instituto de Matemática e Estatística, Universidade de São Paulo, 1983. 154p. (Tese de Mestrado).

RESUMO

A questão da falta de resposta em levantamentos por amostragem é apresentada com um estudo de caso. Mostram-se suas causas, suas conseqüências (tamanho de amostra aleatório, ancilaridade e viés) e os métodos para evitar ou contornar o problema.

SUMMARY

The problem of non-response is presented in this paper with a case study: the sample survey for crops estimation in the State of Sao Paulo, Brazil. There are several kinds of non-response: a) the sample unit cannot be found, due to frame imperfections or because interviewer doesn't know the area very well; b) the sample unit cannot be accessed, usually due to the wheather conditions or the inexistence of good roads (specially during rainy seasons); c) the person to be interviewed isn't able to answer the questions; d) the person to be interviewed refuse to colaborate; and e) loss of material after interview. Two statistical questions are discussed: a) the random sample size arises an ancillary statistic, but it is possible to obtain a conditionally sufficient and complete statistic in order to obtain a uniformly minimum variance unbiased estimator; and b) estimator may be biased since non-response induces a partition of the sample space and the means of the two parts may differ from each other.

**SECRETARIA DE AGRICULTURA E ABASTECIMENTO
INSTITUTO DE ECONOMIA AGRÍCOLA**

Comissão Editorial:

Coordenador: José Roberto Viana de Camargo

Membros: Antonio Augusto Botelho Junqueira

Celuta Moreira Cesar Machado

Elcio Umberto Gatti

Flavio Condé de Carvalho

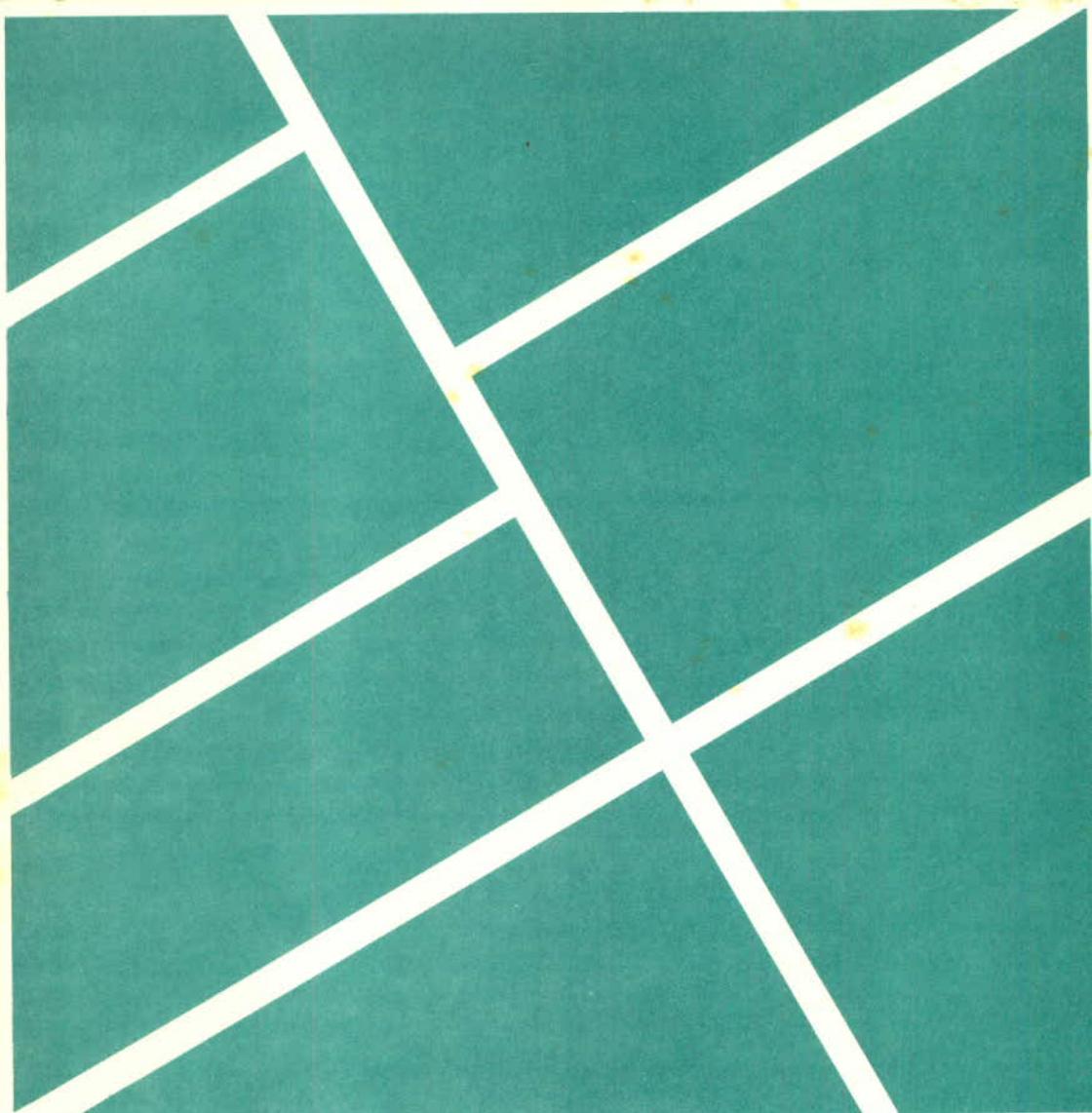
José Luis Teixeira Marques Vieira

Rosa Maria Pescarin Pellegrini

Bibliografia: Fátima Maria Martins Saldanha Faria

Relatores especiais: Fernando Antonio de Almeida Séver

Maria de Fátima Packer



Governo do Estado de São Paulo
Secretaria de Agricultura e Abastecimento
Coordenadoria Sócio-Econômica

Instituto de Economia Agrícola

Relatório de Pesquisa
Nº 8/84