

PREÇOS MÉDIOS RECEBIDOS
PELOS AGRICULTORES:
METODOLOGIA DE
DIMENSIONAMENTO DE AMOSTRAS⁽¹⁾

Edison Eugênio Peceguini⁽²⁾

O objetivo do trabalho consiste em determinar metodologia de amostragem que reflita o dimensionamento mais adequado das amostras de informantes de preços médios recebidos pelos agricultores paulistas, dadas as características envolvidas na obtenção de tais preços. Discute a metodologia empregada pelo IEA e propõe algumas alterações, no sentido de aumentar a representatividade das médias estimadas, corrigindo possíveis distorções no tamanho da amostra aos níveis de significância desejados.

1 — INTRODUÇÃO

Entre as diversas funções do Instituto de Economia Agrícola, a obtenção de informações fidedignas de preços recebidos pelos agricultores do Estado de São Paulo reveste-se da maior importância uma vez que a situação dos mercados de produtos agrícolas estaria mal caracterizada sem a necessária qualidade dos preços utilizados nas análises desses mercados. A importância do uso de boas informações torna-se ainda mais patente face aos inúmeros empregos ensejados pelas mesmas, salientando-se entre outros o auxílio ao poder público na definição da política do setor e o fornecimento de subsídios aos produtores na orientação da produção às preferências dos consumidores.

(1) Este trabalho é parte integrante do Projeto Análise de Preços.

(2) O autor agradece a valiosa colaboração dos Eng.^{os} Agr.^{os} Maristela Simões do Carmo e Nelson Batista Martin pela orientação, revisão do texto e sugestões apresentadas. Liberado para publicação em 24/05/77.

A representatividade dos preços médios recebidos pelos agricultores paulistas também é importante para os técnicos do Instituto de Economia Agrícola que necessitam dessas informações para a elaboração de estudos e análises.

A fim de se conseguir boas estimativas, deve-se determinar o número adequado de informações de preços. O número de informantes será estabelecido mediante o emprego de metodologia de amostragem apropriada.

2 — OBJETIVO

A finalidade deste estudo é verificar que metodologia de amostragem permitirá dimensionar adequadamente amostras de informantes de preços recebidos pelos produtores paulistas, de modo que os preços médios estimados sejam tão representativos quanto possível.

A importância da análise reside no fato de que se pretende evitar a obtenção de amostras subdimensionadas ou superdimensionadas, para as margens de erro e níveis de significância desejados.

2.1 — Hipóteses de Trabalho

Este trabalho baseia-se nas seguintes hipóteses:

- a) a população de preços recebidos tem distribuição normal de probabilidades;
- b) são desconhecidas as variâncias populacionais dos preços e do preço médio, representadas por V_x^2 e $V_{\bar{x}}^2$, respectivamente;
- c) as amostras preliminares são grandes, ou seja, o número de informações básicas é maior do que 30 (trinta), de sorte que o coeficiente "t" de Student passa a depender apenas do nível de significância adotado, assumindo os mesmos valores da distribuição normal; e
- d) as fórmulas analisadas neste trabalho pressupõem que as amostras sejam sorteadas equiprobabilisticamente. Isto significa que a cada elemento da população é atribuída igual oportunidade de ser sorteado.

Com relação à primeira hipótese, conforme salienta IÓRIO (4), sabe-se que “sendo normal o universo, a distribuição das médias também será normal, e que embora não sendo normal o universo, ainda assim a distribuição das médias das amostras é aproximadamente normal e tanto mais próxima da normalidade quanto maior for o tamanho da amostra”.

Quanto à terceira hipótese, pode-se retirá-la, bastando para isto levar-se em consideração o número de graus de liberdade ao consultar a tábua de valores da distribuição “t”. O número de graus de liberdade é dado pelo número de informações preliminares menos um.

No que diz respeito à última hipótese, convém destacar que, mesmo em se tratando de amostra preliminar dimensionada arbitrariamente, aplicam-se as fórmulas aqui apreciadas na avaliação do erro relativo máximo na estimação da média ou erro de amostragem (8).

3 — METODOLOGIA ATUAL

Estudando as repercussões da regionalização do Estado de São Paulo e os efeitos do número de informações sobre a precisão das estimativas dos preços médios recebidos pelos agricultores paulistas, SENDIN (13) desenvolveu um trabalho onde foi utilizada, pela primeira vez, no Instituto de Economia Agrícola, a seguinte expressão:

$$n = CV^2/P^2 = CV\%^2/P\%^2 \quad (I)$$

onde:

n = número de informantes necessários;

CV = coeficiente de variação ou desvio padrão relativo dos preços recebidos; e

P = desvio padrão relativo da média dos preços recebidos ou erro desejado em torno da média desses preços.

Em outros trabalhos realizados no Instituto de Economia Agrícola (11, 14) também utilizou-se a fórmula acima. Na literatura estatística, a relação (I) é encontrada em MADOW (10), que esclarece tratar-se de uma primeira tentativa para o cálculo do tamanho da amostra. O procedimento seguido na determinação dessa fórmula é indicado a seguir.

Considerando-se:

$$\begin{aligned}\bar{x} &= (1/n) \sum x_i \\ S_x^2 &= \sum (x_i - \bar{x})^2 / (n-1) \\ S_{\bar{x}}^2 &= S_x^2 / n \\ CV\% &= (S_x / \bar{x}) \cdot 100 \\ P\% &= (S_{\bar{x}} / \bar{x}) \cdot 100\end{aligned}$$

onde:

- \bar{x} = estimativa não viesada do preço médio recebido;
 n = número de informações preliminares sobre preços recebidos;
 S_x^2 = estimativa não viesada da variância dos preços recebidos;
 S_x = estimativa não viesada do desvio padrão dos preços recebidos;
 $S_{\bar{x}}^2$ = estimativa não viesada da variância do preço médio;
 $S_{\bar{x}}$ = estimativa não viesada do desvio padrão do preço médio;
 CV = coeficiente de variação estimado dos preços recebidos; e
 P = desvio padrão relativo da média dos preços recebidos.

Desenvolvendo-se:

$$\begin{aligned}S_{\bar{x}}^2 &= S_x^2 / n \\ \therefore n &= S_x^2 / S_{\bar{x}}^2 \\ \therefore n &= (S_x^2 / \bar{x}^2) / (S_{\bar{x}}^2 / \bar{x}^2) \\ \therefore n &= (S_x^2 / \bar{x}^2) \cdot 100^2 / (S_{\bar{x}}^2 / \bar{x}^2) \cdot 100^2\end{aligned}$$

mas: $(S_x^2 / \bar{x}^2) \cdot 100^2 = CV\%^2$ e

$$(S_{\bar{x}}^2 / \bar{x}^2) \cdot 100^2 = P\%^2$$

$$\therefore n = CV^2/P^2 = CV\%^2/P\%^2$$

Cabe aqui um rápido esclarecimento. Em princípio, poder-se-ia supor que a expressão (I) seria inaplicável na determinação do tamanho da amostra, porque ao empregar-se $S_{\bar{x}}^2 = S_x^2 / n$, no cálculo de $P\%$, resultaria sempre $n' = n$, onde:

n' = tamanho da amostra; e

n = número de informações preliminares,

isto é, pela fórmula (I) o tamanho da amostra seria igual ao número de observações preliminarmente levantadas para calcular as estimativas \bar{x} , S_x e $S_{\bar{x}}$.

Este problema ficaria superado caso fosse conhecida a variância populacional do preço médio, $V_{\bar{x}}^2$. Nesse caso, ter-se-ia:

$$P\% = (V_{\bar{x}}^2 / \bar{x}^2) \cdot 100^2$$

e, portanto, $n' \neq n$.

Entretanto, $V_{\bar{x}}^2$ é desconhecida. A fim de sanar a questão, o que se faz é considerar arbitrariamente valores para $P\%^2$, o que significa, implicitamente, atribuir valores desejados para $V_{\bar{x}}^2$.

3.1 — Aplicações da Metodologia Atual

A primeira aplicação da relação (I) é a determinação do tamanho da amostra. Conhecido o coeficiente de variação dos preços recebidos, fixa-se uma margem de erro desejado em torno da média populacional, calculando-se o número mínimo de informantes necessários. Se um produto apresentou um coeficiente de variação estimado dos preços recebidos da ordem de 20%, o número de informantes necessários para um erro desejado em torno da média de 5% será igual a

$$n = (20\%)^2 / (5\%)^2 = 16.$$

Portanto, bastariam 16 informantes para que, com o preço médio recebido estimado \bar{x} , se possa afirmar que a média populacional μ , desconhecida, acha-se compreendida no intervalo: $\bar{x} \pm 0,05 \bar{x}$ (figura 1).

Outro emprego da fórmula (I) diz respeito à estimativa do erro máximo no cálculo da média, ou seja, a obtenção do erro de amostragem de um levantamento já realizado. Assim, a partir de um certo número de observações de preços recebidos, definido por algum critério, ou mesmo arbitrariamente (8), pode-se controlar a qualidade da média desses preços com base nesse número de observações.

Fixando-se, por exemplo, em 36 o número de informantes necessários, com um coeficiente de variação estimado de 30%, o erro relativo máximo no cálculo do preço médio recebido pelos agricultores será igual a 5%. Para obter-se o erro de amostragem (Er) utiliza-se a transformação da fórmula (I), abaixo:

$$Er = P\% = (CV\%) / \sqrt{n} \quad (II)$$

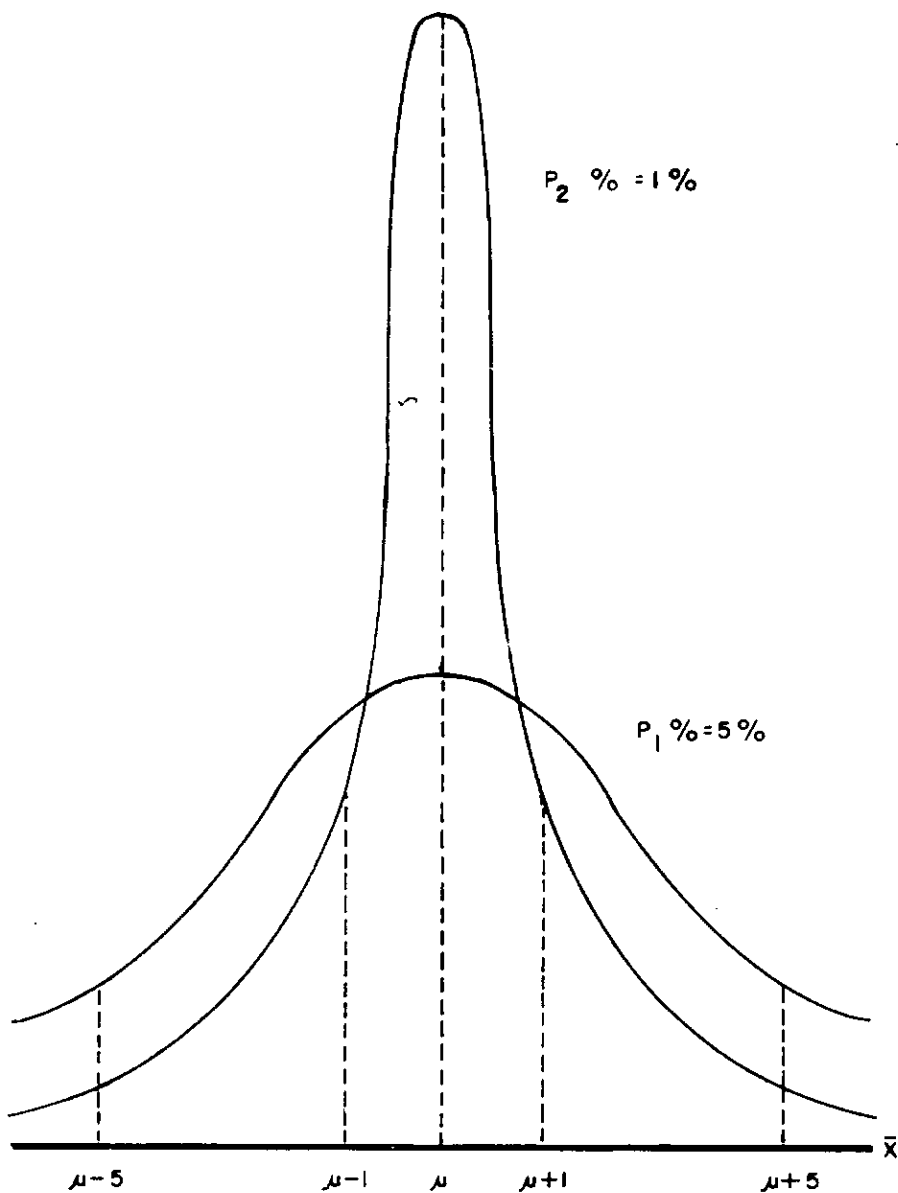


FIGURA 1. — Distribuição de Frequência da Média Amostral, com Diferentes Hipóteses sobre o Verdadeiro Desvio Padrão dessa Média.

Normalmente, o que se faz é combinar as duas aplicações acima apresentadas. Inicialmente dimensiona-se a amostra e, após a coleta e tratamento dos dados, testa-se a qualidade das estimativas obtidas.

3.2 — Comentários sobre a Metodologia Atual

Sem qualquer demérito aos trabalhos já referidos, que empregaram essa metodologia, nenhuma menção é feita quanto ao nível de confiança adotado no cálculo do tamanho da amostra ou na avaliação do erro de amostragem.

Para verificar qual o nível de confiança implicitamente aceito ao empregar-se a expressão (I), demonstra-se que (I) é equivalente à fórmula:

$$n = S_x^2 / L^2 \quad (I)'$$

onde:

S_x^2 = estimativa não viesada da variância dos preços recebidos; e

L = erro permitido para o preço médio, fornecido pela amostra, em torno da média populacional.

Sabe-se que $L = P \cdot \bar{x}$ e, portanto, desenvolvendo tem-se:

$$n = CV^2 / P^2 \quad (I)$$

$$\therefore n = (S_x / \bar{x})^2 / (P \cdot \bar{x} / \bar{x})^2$$

$$\therefore n = S_x^2 / (P \cdot \bar{x})^2$$

$$\therefore n = S_x^2 / L^2 \quad (I)'$$

Os resultados fornecidos por (I) ou (I)', quando aplicados ao dimensionamento de amostras, devem ser entendidos como tamanhos de amostra quando:

- a) o erro tolerado em torno da média diminui à medida que P% diminui; e
- b) o nível de confiança permanece constante.

Em outras palavras, os valores obtidos são tamanhos de amostra ao se fazer variar o erro tolerado ($\pm L$) em torno da média populacional μ , sempre ao mesmo nível de confiança.

Por outro lado, o nível de confiança do erro de amostragem calculado através de (II) também é constante, conforme verificar-se-á adiante.

Demonstra-se que esse nível de confiança constante é de 68,26% (6,8). Quando se atribui um valor para P%, arbitrariamente, fixa-se um tipo de curva de distribuição de frequência para a variável aleatória \bar{x} , tal que a área correspondente a 68,26%, ou seja, um desvio padrão em torno da média populacional, fique compreendida num raio de P% em torno de μ .

Sendo desconhecida a forma verdadeira da distribuição de frequência de \bar{x} , para cada valor de P% tem-se uma forma diferente para o gráfico dessa distribuição, de tal modo que quanto menor o valor atribuído a P% menor será a $V \frac{\sigma}{\bar{x}}$ e, portanto, maior será a concentração de \bar{x} em torno da média populacional μ , desconhecida. Logo, quanto menor P%, mais compacta será a distribuição de frequência de \bar{x} . Conforme a figura 1, se, por exemplo, $\bar{x} = \text{Cr\$ } 100,00$, para P_1 a 5% e P_2 a 1% tem-se, respectivamente, $(V \frac{\sigma}{\bar{x}})_1 = 5,00$ e $(V \frac{\sigma}{\bar{x}})_2 = \text{Cr\$ } 1,00$.

Contudo, qualquer que seja o valor atribuído a P%, o nível de confiança no dimensionamento da amostra ou no cálculo do erro de amostragem é constante e igual a 68,26%, porque o intervalo de confiança é sempre o mesmo e da ordem de um desvio padrão em torno de μ .

Através de um exemplo, pode-se verificar a implicação do emprego da relação (I) no dimensionamento de amostras. Admita-se que: $\bar{x} = 40$ e $S_x^2 = 100$. O nível de significância de 31,74% será sempre o mesmo, enquanto que o erro $\pm L$ em torno da média populacional μ , desconhecida, decresce quando P% decresce. Logo, para diferentes valores de P tem-se:

$$\begin{aligned} P_1\% &= 30\% \text{ sendo } n_1 = 1; \\ P_2\% &= 10\% \text{ sendo } n_2 = 7; \\ P_3\% &= 5\% \text{ sendo } n_3 = 25; \text{ e} \\ P_4\% &= 1\% \text{ sendo } n_4 = 625. \end{aligned}$$

O resultado $n_3 = 25$, por exemplo, significa que de 100 amostras de tamanho 25, extraídas da população, em apenas 68 amostras a média populacional μ , desconhecida, deverá estar contida

no intervalo $\bar{x} \pm P.\bar{x} = 40 \pm (0,05)40$. O resultado $n_1 = 625$ significa que de 100 amostras desse tamanho, extraídas da população, somente 68 amostras deverão conter a média populacional μ no intervalo $\bar{x} \pm P.\bar{x} = 40 \pm (0,01)40$. E assim por diante.

4 — METODOLOGIA PROPOSTA

Para a amostragem aleatória simples, a teoria estatística (1, 3, 6, 15, 16) fornece uma expressão tradicional:

$$n = t_a^2 S_x^2 / L^2 \quad (III)$$

em que:

- t_a = valor do coeficiente "t" ao nível de a%;
- a% = nível de significância;
- (1 - a)% = nível de confiança;
- S_x^2 = estimativa não viesada da variância dos preços recebidos; e
- L = erro permitido para o preço médio fornecido pela amostra, ou, em outras palavras, é o valor máximo em cruzeiros permitido como variação acima ou abaixo da média dos preços.

Portanto, partindo da fórmula (III) tradicional pode-se confirmar algebricamente que o nível de confiança implícito ao usar-se a relação (I) ou (I)' é de 68,26%. De fato, desenvolvendo chega-se a uma expressão (III)' idêntica à fórmula (III):

$$\begin{aligned} n &= t_a^2 S_x^2 / L^2 && (III) \\ \therefore n &= t_a^2 S_x^2 / P^2 \cdot x^2 \\ \therefore n &= t_a^2 (CV^2 / P^2) && (III)' \end{aligned}$$

A fórmula (III)', que é encontrada em BRANDÃO (2), KESSEL (7) e KIEHL (9), permite constatar facilmente que as metodologias de amostragem em discussão fornecerão os mesmos resultados somente quando $t_a^2 = 1,00$, o que ocorre ao nível de confiança de 68,26%.

4.1 — Aplicação da Metodologia Proposta

O emprego da fórmula clássica, proposto, presta-se, do mesmo modo que a metodologia em vigor, ao cálculo do número mínimo

de informantes para o levantamento de preços médios recebidos pelos produtores e também para o controle de qualidade de estimativas.

4.2 — Comentários sobre a Metodologia Proposta

Uma vantagem do emprego da metodologia proposta é tornar explícito o nível de confiança adotado na amostragem e na medida do erro de amostragem. Não se trata de um refinamento teórico, mas de um método bastante empregado na prática, inclusive pela Fundação IBGE, segundo KESSEL (7), e no Instituto de Economia Agrícola, para o cálculo da amostra na obtenção de preços de gêneros alimentícios a nível de varejo, na cidade de São Paulo, conforme JUNQUEIRA et alii (5) e UENO (17).

Por explicitar o grau de confiança, a fórmula tradicional fornece a possibilidade de variação tanto no erro tolerado em torno da média como no grau de confiança envolvido nos cálculos.

Com auxílio do exemplo numérico anterior, deve-se interpretar os resultados fornecidos pela metodologia proposta conforme abaixo:

$$\begin{aligned} a_1\% &= P_1\% = 30\% \text{ sendo } n'_1 = 1; \\ a_2\% &= P_2\% = 10\% \text{ sendo } n'_2 = 17; \\ a_3\% &= P_3\% = 5\% \text{ sendo } n'_3 = 97; \text{ e} \\ a_4\% &= P_4\% = 1\% \text{ sendo } n'_4 = 4.161 \end{aligned}$$

O resultado $n'_3 = 97$, por exemplo, significa que de 100 amostras de tamanho 97, extraídas da população, em 95 amostras a média da população provavelmente deverá estar contida no intervalo $\bar{x} \pm P.\bar{x} = 40 \pm (0,05)40$. O resultado $n'_4 = 4.161$ significa que de 100 amostras desse tamanho, extraídas da população, 99 amostras deverão provavelmente conter a média populacional μ , desconhecida, no intervalo $\bar{x} \pm P.\bar{x} = 40 \pm (0,01)40$. E assim por diante.

Os valores $n_1 = n'_1 = 1$ confirmam a observação de que as duas metodologias em exame apresentam os mesmos resultados apenas para o nível de confiança de 68,26% (por aproximação, $1 - a\% = 70\%$), e que esse é o grau de confiança implícito no emprego da relação (I).

Quando aplicada no controle de qualidade de estimativas, a metodologia sugerida permite que se determine o erro de amostragem, fixando-se de antemão o nível de confiança desejado para o cálculo desse erro. Neste caso, utiliza-se a seguinte transformação da fórmula clássica encontrada em LEME (9):

$$Er = P\% = t_a (CV\% / \sqrt{n}) \quad (IV)$$

Se o nível de confiança desejado é de 68,26%, então $t_a = 1,00$ e recaímos na expressão (II) da metodologia em vigor.

Com os dados $n = 36$ e $CV\% = 30\%$, pode-se determinar qual o erro de amostragem do levantamento para qualquer grau de confiança:

Para $a = 68,26\%$ tem-se: $Er_1 = 1,00 (30/6) = 5,00\%$;

Para $a = 95,00\%$ tem-se: $Er_2 = 1,96 (30/6) = 9,80\%$; e

Para $a = 99,00\%$ tem-se: $Er_3 = 2,58 (30/6) = 12,90\%$.

Os resultados acima devem ser interpretados da seguinte forma: Er_1 indica que a estimativa do preço médio recebido, \bar{x} , apresenta erro relativo máximo inferior a 5,00% em 68 de 100 amostras de tamanho $n = 36$; Er_2 significa que \bar{x} apresenta erro relativo máximo inferior a 9,80% em 95 de 100 amostras com 36 elementos.

Nota-se que sendo desejada uma precisão maior para a estimativa do preço médio recebido, maior deverá ser o tamanho da amostra. Para um $Er = 5,00\%$ com 95,00% de confiança, com base nos dados supra, a amostra deverá possuir $(1,96)^2 \cdot 36 = (3,84)36 = 139$ elementos. Assim, com $n = 139$ afirma-se que, em 95 de 100 amostras desse tamanho, o erro relativo máximo de \bar{x} será menor do que 5,00%.

4.3 — Fator de Ajustamento para Amostras Grandes de Populações Finitas

Se a população da qual será extraída uma amostra for razoavelmente grande, o tamanho desse universo normalmente não afetará a dimensão da amostra. Entretanto, exigindo-se maior rigor estatístico, deve-se corrigir o tamanho da amostra extraída de

universo finito, com o uso da fórmula abaixo, conforme ALLEN (1), KARMEL (6), LEME (9) e SNEDECOR (15):

$$n^* = \bar{n} [1 / (1 + \emptyset)] \quad (V)$$

onde:

- n^* = tamanho da amostra corrigido;
- n = tamanho da amostra fornecido pela relação clássica;
- \emptyset = n/N = fração amostral; e
- N = população.

4.4 — Amostras Extraídas sem Reposição

As fórmulas discutidas até aqui pressupõem que as amostras são extraídas de um universo ilimitado ou que, no caso de universo finito, cada elemento que constitui a amostra é retirado do universo com reposição dos elementos anteriormente extraídos.

Mas, se o universo é finito e os elementos que compõem a amostra forem retirados sem reposição dos anteriormente extraídos, as novas extrações não serão independentes e, nestas circunstâncias, demonstra-se que a fórmula clássica para dimensionar amostras passa a ser:

$$n = t_a^2 S_x^2 N / [L^2 (N-1) + t_a^2 S_x^2] \quad (VI)$$

5 — CONCLUSÕES

Segundo KIEHL (8), “falar em margem de erro sem mencionar o grau de confiança não tem nenhum significado... é preciso saber qual a probabilidade de que o valor levantado pela amostra seja igual ao valor real do universo com a margem de erro mencionada”.

Para isso deve-se distinguir os conceitos de margem de erro e nível de confiança:

- a) margem de erro é a diferença entre a estimativa do preço médio, \bar{x} , obtida através da amostra e o valor real do universo, μ ; sendo μ desconhecido, adota-se uma margem de erro $\pm L$ tolerável; e
- b) nível de confiança pode ser entendido como a probabilidade de que a amostra sorteada seja representativa da população da qual foi retirada.

As duas metodologias analisadas aplicam-se tanto para o cálculo do tamanho de uma amostra de informantes quanto para a avaliação do erro de amostragem de um levantamento já efetuado, cabendo as seguintes considerações quando se adota a metodologia atual:

- a) os valores fornecidos pelas fórmulas (I) ou (I)' crescem à medida que se deseja uma margem de erro cada vez menor, $\pm L$, em torno da média da população; isso se faz atribuindo valores decrescentes a $P^0\%$;
- b) o nível de confiança, não mencionado, é constante e igual a 68,26%; isto significa que ao retirar-se do universo uma amostra ao acaso, a probabilidade dessa amostra ser representativa da população da qual foi extraída será de 68,26%; e
- c) quando se utiliza a expressão (II) para o cálculo da estimativa do erro relativo máximo na obtenção do preço médio, sabe-se qual é a qualidade das estimativas em 68,26% das amostras retiradas, representativas do universo.

Por outro lado, o emprego da metodologia proposta possibilita tecer os seguintes comentários:

- a) a relação (III) ou (III)' é mais geral, porque permite não apenas que a margem de erro $\pm L$ varie, mas que o nível de confiança também varie, de sorte que a representatividade da amostra pode ser modificada. Se o nível de confiança desejado for 68,26%, então $t_a = 1,00$ e recai-se na relação (I) \equiv (I)' da metodologia atual. Se o nível de confiança desejado for maior ou menor do que 68,26%, então $t_a \neq 1,00$ e a fórmula clássica deverá ser a utilizada; e
- b) para o controle de qualidade das estimativas obtidas em um levantamento, a fórmula (IV) permite avaliar o erro de amostragem para qualquer nível de confiança desejado. Observe-se que só quando $(1 - a)\% = 68,26\%$, $t_a = 1,00$ é que se recai na fórmula (II) da metodologia em vigor.

Apontadas as implicações que ocorrem quando do emprego das metodologias em questão, discute-se a seguir algumas razões que recomendam a adoção da metodologia proposta.

Para que a margem de erro tolerado em torno da média venha apresentar significado estatístico é necessário frisar o nível de

confiança a que ela está ligada, o que se obtém atribuindo ao nível desejado o valor de "t" correspondente.

Elimina-se assim a possibilidade do erro tolerado em torno da média vir a ser confundido com nível de significância. Uma identificação equivocada desses conceitos poderia levar a admitir como válido o emprego das fórmulas (I) \equiv (I)' no cálculo do tamanho de amostras quando o nível de confiança desejado diferir de 68,26%. Ocorreria então que para níveis de significância inferiores a 31,74% o emprego dessas fórmulas subestimaria o número mínimo de informações necessárias em $t_n^2 > 1$ vezes. A 5% e 1%, por exemplo, a amostra estaria subestimada em $(1,96)^2 = 3,84$ e $(2,58)^2 = 6,66$ vezes, respectivamente. Por outro lado, haveria superestimação do número de observações necessárias a níveis de significância superiores a 31,74%, onde essas fórmulas forneceriam amostras aumentadas de $t_n^2 < 1$ vezes.

Na prática adotam-se para análises de variância, construção de intervalos de confiança, testes "t", verificação da significância das estimativas dos parâmetros de uma regressão pelo método dos mínimos quadrados, etc., os níveis de significância de 5% ou 1%. No tocante, à amostragem, BRANDÃO (2) propõe o emprego da fórmula (III)', com t a 1% para determinar o número ideal de informantes agrícolas na amostra, enquanto que noutro extremo SCHATAN (12) adota a fórmula (I), da metodologia de amostragem para obtenção de preços médios recebidos, em vigor no Instituto de Economia Agrícola, cujo nível de significância é sempre o mesmo e de 31,74%.

Sabe-se que o tamanho de uma amostra depende fundamentalmente de três fatores: a variabilidade da população (ou de seus extratos); o grau de precisão almejado; e a disponibilidade de recursos.

Quanto maior a exatidão desejada, maior o tamanho da amostra e, portanto, maior deverá ser a disponibilidade de recursos. Tomando como base uma amostra cujo tamanho foi calculado para o nível de significância de 31,74%, mantendo a mesma margem de erro tolerado em torno da média, verifica-se facilmente através de (III) ou (III)' quantas vezes maior deverá ser o número de elementos da amostra para diferentes níveis de significância, conforme discutido anteriormente.

Respeitada a disponibilidade de recursos para o levantamento de informações e conhecida a variabilidade da população de preços recebidos, por meio das estimativas S_x^2 ou CV%, deve-se empregar o mais elevado nível de confiança possível.

Para o controle de qualidade das estimativas de um levantamento já concretizado, a fórmula (IV), da metodologia proposta, possibilita medir o erro de amostragem qualquer que seja o nível de confiança desejado, ao passo que a relação (II) fornece o erro de amostragem apenas para o nível de confiança de 68,26%.

Como observação final, em se tratando de amostra grande extraída de população finita, o ajustamento do tamanho dessa amostra deverá ser feito utilizando-se a expressão (V) e, no caso de admitir-se que as amostras são retiradas sem reposição de seus elementos, deverá ser empregada a fórmula (VI).

SUMMARY

This report aims at discussing the appropriate sampling methodology to obtain meaningful estimates of prices paid to farmers in the State of São Paulo. It suggests some changes in the methodology used by IEA, in order to obtain greater reliability and to correct distortions in sample size at desired significance levels.

LITERATURA CITADA

1. ALLEN, R. G. D. *Estatística para economistas*. Trad. Austregésilo Gomes Spinola. Rio de Janeiro, Fundo de Cultura, 1964. 214 p. cap 9.
2. BRANDÃO, E. D. *Levantamentos para administração rural*. Viçosa, MG, UREMG, Escola de Especialização, 1961. (mimeo.) p. 38-50.
3. HADLEY, G. *Elementary statistics*. San Francisco, Holden-Day, 1969. cap. 3.
4. IÓRIO, O. Nota técnica sobre amostragem. *R. Bras. Estat.*, IBGE, 36(143):485-501, jul./set. 1975.
5. JUNQUEIRA, P. C.; LAZZARINI, M. I.; CANTO, W. L. Análise comparativa de preços do varejo dos gêneros alimentícios na capital de São Paulo. *Agric. em São Paulo*, 19(II):113-165, 1972.
6. KARMEL, P. H. & POLASEK, M. *Estatística geral e aplicada para economistas*. São Paulo, Atlas, 1972. cap. 7 e 8.
7. KESSEL, M. Aplicação da técnica de amostragem aos levantamentos de fluxos internacionais e interregionais de mercadorias. *R. Bras. Estat.*, IBGE, 29(115):347-360, jul./set. 1968.
8. KIEHL, L. F. O tamanho da amostra na pesquisa de mercado. *R. Adm. Emp.*, FGV, 10(4):205-216, dez. 1970.
9. LEME, Ruy A. S. *Curso de estatística: elementos*. 6.^a ed. Rio de Janeiro, Livros Técnicos e Científicos, 1974. p. 202-204.
10. MADOW, W. G. *Teoria dos levantamentos por amostragem*. Rio de Janeiro Conselho Nacional de Estatística, IBGE, 1951. p. 207-208.
11. PADULA, A. F. S. Preços médios de pescado, recebidos pelos armadores da pesca, no Est. de São Paulo. *Inf. Econ.*, IEA, 5(10):9-15, out. 1975.
12. SCHATAN, S. A amostragem na previsão da produção de café no Estado do Paraná. *R. Bras. Estat.*, IBGE, 16(62):105-123, abr./jun. 1955.
13. SENDIN, P. V. Preços médios recebidos pelos lavradores: efeito do número de informações e a regionalização sobre a precisão das estimativas. *Agric. em São Paulo*, 15(9/10):19-25, out. 1968.
14. ——— & CARMO, M. S. Análise da qualidade das informações dos preços médios recebidos pelos produtores de milho no Estado de São Paulo, 1969. *Agric. em São Paulo*, 17(7/8):1-17, ago. 1970.
15. SNEDECOR, G. W. *Métodos estadísticos*. 5.^a ed. México, Companhia Editorial Continental, 1966. p. 586-589.
16. SPURR, W. A. & BONINI, C. P. *Statistical analysis for business decisions*. Chicago, Ill., Richard D. Irwin, 1973. p. 242-248.
17. UENO, L. H. Cesta de mercado: redimensionamento da amostra de equipamentos comerciais. *Inf. Econ.*, IEA, 6(2):i-q, fev. 1976.



IMPrensa OFICIAL DO ESTADO S/A - IMESP
SÃO PAULO - BRASIL

1978