

ESTUDO DE AMOSTRAGEM PARA PESQUISA SÓCIO-ECONÔMICA JUNTO A PRODUTORES DE ASPARGO NO RIO GRANDE DO SUL¹

ENEDINO CORRÊA DA SILVA², VERA OSORIO DA FONSECA³ e JOSÉ ALCEU INFELD⁴

SINOPSE.— É proposto um estudo de amostragem para pesquisa sócio-econômica junto a produtores de aspargo.

INTRODUÇÃO

Em amostragem, o fundamental é retirar-se uma amostra mínima e dela obterem-se conclusões válidas extensivas à população, o que pode ser melhor alcançado através da estratificação. Jessen (1942) e Jessen e Houseman (1944) estudaram a efetividade da estratificação com referência a um certo número de setores econômicos mais típicos. Armitage (1947) e Evans (1951) fizeram comparações entre a amostragem acidental simples e a amostragem acidental estratificada com repartição proporcional e ótima ou de Neyman (1934).

A população era constituída pelos fornecedores de aspargo às indústrias de Pelotas, Rio Grande do Sul, com produção inferior a 10.000 kg, em número de 1.060 agricultores⁵. Se considerássemos uma população formada por pequenos e grandes produtores de aspargo, obteríamos uma variância muito alta e a amostra resultaria muito grande, dada a heterogeneidade dos dois estratos. Assim, a estratificação realizada neste estudo buscou estabelecer, com base estatística, qual o número de participantes que representava, de maneira significativa, a população em foco.

MATERIAL E MÉTODOS

O estudo através do qual se determinou a amostra foi feito sobre os fornecedores de aspargo às indústrias de Pelotas. A amostragem foi realizada em função da produção, sendo necessário que houvesse uma estreita correlação entre esta variável e aquelas que seriam objeto de estudo, para que o mesmo tivesse validade.

¹ Aceito para publicação em 8 ago. 1972.

² Eng.º Agrônomo, M.Sc., do Setor de Estatística Experimental e Análise Econômica do Instituto de Pesquisa Agropecuária do Sul (IPEAS), Caixa Postal E, Pelotas, Rio Grande do Sul, e Pesquisador Assistente, bolsista, do Conselho Nacional de Pesquisas (CNPq).

³ Economista do Setor de Estatística Experimental e Análise Econômica do IPEAS.

⁴ Eng.º Agrônomo do Setor de Estatística Experimental e Análise Econômica do IPEAS, e Pesquisador Assistente, bolsista, do CNPq.

⁵ No estudo não foram incluídos os agricultores que produzem acima de 10.000 kg, considerando-se-os somente na estimativa de produção.

A população foi subdividida em sete estratos mais ou menos homogêneos, constituindo cada estrato uma subpopulação, da qual retiramos uma amostra simples ao acaso. Assim, obtivemos os seguintes estratos:

N₁: até 200 kg, com 238 fornecedores;
 N₂: de mais de 200 a 500 kg, com 220 fornecedores;
 N₃: de mais de 500 a 1.000 kg, com 206 fornecedores;
 N₄: de mais de 1.000 a 2.000 kg, com 199 fornecedores;
 N₅: de mais de 2.000 a 3.500 kg, com 116 fornecedores;
 N₆: de mais de 3.500 a 6.000 kg, com 58 fornecedores;
 N₇: de mais de 6.000 a 10.000 kg, com 23 fornecedores,
 sendo N_h o tamanho do estrato h.

O método usado foi o de uma amostragem estratificada proporcional, usando a partilha ótima de Neyman.

Assim, sendo N_h o tamanho do estrato h, a média e a variância do mesmo serão

$$\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} y_{hi} \text{ e } S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2,$$

onde y_{hi} indica o iésimo elemento do estrato h.

Como do estrato h tira-se uma amostra de n_h elementos, as estimativas de \bar{Y}_h e de S_h² serão:

$$\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi} \text{ e } s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2.$$

A estimativa da média da população, com base na amostra estratificada, é dada por

$$\bar{y}_{est} = \frac{1}{N} \sum_{h=1}^H N_h \bar{y}_h, \quad (1)$$

onde N = N₁ + N₂ + ... + N_H é o número total de elementos na população. Essa é uma média ponderada, onde os valores de N_h são os pesos. Aliás, a própria média da população, \bar{Y} , é também uma média ponderada das médias \bar{Y}_h dos estratos.

A variância de \bar{y}_{est} , por definição, é dada por

$$V(\bar{y}_{est}) = E | (\bar{y}_{est} - \bar{Y})^2 |$$

e demonstra-se, para o caso de H estratos, que a variância é

$$V(\bar{y}_{est}) = \frac{1}{N^2} \frac{N-n}{n} \sum_{h=1}^H N_h S_h^2,$$

RESULTADOS E DISCUSSÃO

sendo

$$\hat{V}(\bar{y}_{est}) = \frac{1}{N^2} \frac{N-n}{n} \sum_{h=1}^H N_h s_h^2 \quad (2)$$

a estimativa da variância, onde s_h^2 se obtém por aplicação, à amostra do estrato h , da fórmula

$$s_h^2 = \frac{\sum_y (\Sigma y)^2}{n_h - 1}$$

Por outro lado, demonstra-se que as estimativas de \bar{y}_{est} e $V(\bar{y}_{est})$ são imparciais ("unbiased").

Partilha de Neyman

Sendo $n = n_1 + n_2 + \dots + n_h$ o número total de elementos nas amostras consideradas, a fórmula

$$n_h = \frac{N_h S_h}{\sum N_h S_h} \cdot n$$

nos fornece a partilha ótima, isto é, aquele número de elementos que deve conter a amostra, para que a variância seja mínima.

O intervalo de confiança para a média \bar{Y} tem seus extremos dados por $\bar{y}_{est} \pm t s(\bar{y}_{est})$,

onde

$$s(\bar{y}_{est}) = \sqrt{\hat{V}(\bar{y}_{est})}$$

Para o total $N\bar{Y}$ o intervalo de confiança tem extremos

$$N\bar{y}_{est} \pm N t s(\bar{y}_{est})$$

Se todos os extratos tiverem variâncias semelhantes, o número de graus de liberdade de $s(\bar{y}_{est})$ é aproximadamente a soma dos números de graus de liberdade para os estratos. Mas, se as variâncias forem bastante diferentes, o número de graus de liberdade (n') de $s(\bar{y}_{est})$ é dado aproximadamente pela fórmula Satterthwaite,

$$n' = \frac{(f_1 s_1^2 + f_2 s_2^2 + \dots + f_H s_H^2)^2}{\frac{f_1^2 s_1^4}{n_1 - 1} + \frac{f_2^2 s_2^4}{n_2 - 1} + \dots + \frac{f_H^2 s_H^4}{n_H - 1}}$$

onde

$$f_h = \frac{N_h (N_h - n_h)}{n_h}$$

Para o estudo, foram obtidas amostras para cada estrato equivalentes a 10% do mesmo*, obtendo-se os dados constantes do Quadro 1.

Daí, obtivemos:

$$\begin{aligned} \bar{y}_{est} &= 1.057,88 \text{ kg,} \\ \hat{V}(\bar{y}_{est}) &= 915,99, \text{ e} \\ s(\bar{y}_{est}) &= 30,26. \end{aligned}$$

Aplicando a partilha ótima de Neyman, obtivemos:

$n_1 = 7;$	$s_1^2 = 3.585,32;$	$s_1 = 59,87;$
$n_2 = 8;$	$s_2^2 = 8.402,96;$	$s_2 = 91,12;$
$n_3 = 14;$	$s_3^2 = 20.092,94;$	$s_3 = 141,86;$
$n_4 = 27;$	$s_4^2 = 69.792,38;$	$s_4 = 264,17;$
$n_5 = 22;$	$s_5^2 = 177.563,80;$	$s_5 = 421,38;$
$n_6 = 19;$	$s_6^2 = 453.910,98;$	$s_6 = 673,72;$
$n_7 = 14;$	$s_7^2 = 1.756.922,59;$	$s_7 = 1.325,48;$

e

$$\begin{aligned} \bar{y}_{est} &= 1.179,60 \text{ kg,} \\ \hat{V}(\bar{y}_{est}) &= 822,26, \\ s(\bar{y}_{est}) &= 28,68. \end{aligned}$$

Intervalo de confiança

O valor do IC para média estratificada foi dado por

$$\bar{y}_{est} \pm t s(\bar{y}_{est})$$

Como não há homogeneidade de variância entre os estratos, fomos à tabela de t com n' graus de liberdade. Assim,

$$\begin{aligned} 1.179,60 \pm 2 \times 28,68, \text{ portanto,} \\ 1.179,60 \pm 57,36 \text{ kg,} \end{aligned}$$

havendo, portanto, uma amplitude de 5% da média estimada, o que é considerado bastante aceitável.

* Considera-se que, para estudos dessa natureza, uma amostra com 10% dos elementos seja suficiente.

QUADRO 1. Dados para a estimativa da grandeza da amostra

Estratos	N_h	$n_h(10\%)$	s_h	s_h	$N_h s_h$
1	238	24	3.880,09	62,29	14.825,02
2	220	22	7.176,75	84,71	18.636,20
3	206	21	22.405,85	149,68	30.834,08
4	199	20	100.428,63	316,90	63.063,10
5	116	12	182.219,79	426,87	49.516,92
6	58	6	526.925,95	754,05	42.101,62
7	23	3	1.999.586,12	1.414,06	32.523,38
Total	1.060	108			

O valor do IC para total da população (Y) foi dado por

$$N \bar{y}_{est} \pm N t_{\alpha} s(\bar{y}_{est});$$

assim,

$$1.060 \times 1.179,6 \pm 1.060 \times 2 \times 28,68,$$

portanto,

$$1.250.376,0 \pm 60.801,6.$$

A estimativa da produção total, incluindo dados não considerados na amostragem, por fugirem muito dos demais (dados acima de 10.000 kg), foi:

$$(10.136,5 + 13.884,0 + 19.567,0 + 11.135,0 + 40.300,0 + 13.390,0) + 1.250.376,0 = 1.358.755,5 \text{ kg.}$$

CONCLUSÕES

Sendo Y = 1.261.125,5 kg o total da produção, a estimativa $\hat{Y} = 1.250.376,0$ kg, com amplitude de 60.801,6 kg, foi altamente satisfatória, o mesmo acontecendo com a média estratificada (\bar{y}_{est}), que mostrou a amplitude de 5% da média estimada.

O uso da estratificação mostrou-se altamente vantajoso sobre a amostra simples ao acaso, sendo que o uso da partilha ótima de Neyman melhorou ainda mais as

condições de amostragem, uma vez que a mesma estimou um número n_h de elementos que deve possuir a amostra em cada estrato, tornando mínima a variância.

A partilha ótima de Neyman, determinou

1.º estrato:	$n_1 = 7$	fornecedores;
2.º " :	$n_2 = 8$	" ;
3.º " :	$n_3 = 14$	" ;
4.º " :	$n_4 = 27$	" ;
5.º " :	$n_5 = 22$	" ;
6.º " :	$n_6 = 19$	" ;
7.º " :	$n_7 = 14$	" ;

num total de 111 fornecedores de asparago.

REFERÊNCIAS

Armitage, P. 1947. A comparison of stratified with unrestricted random sampling from a finite population. *Biometrika* 34: 273-280.

Evan, W.D. 1951. On stratification and optimum allocation. *J. Am. Stat. Assoc.* 46:95-104.

Jessen, R.J. 1942. Statistical investigation of a sample survey for obtaining farm facts. *Iowa Agric. Exp. Sta. Res. Bull.* 304.

Jessen, R.J. & Houseman, E.E. 1944. Statistical investigations of farm surveys taken in Iowa, Florida and California. *Iowa Agric. Exp. Sta. Res. Bull.* 329.

Neyman, J. 1934. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Roy. Stat. Soc.* 97:558-606.

ABSTRACT.- Silva, E.C.da; Fonseca, V.O.da; Infeld, J.A. [*Sampling study for a social and economic research with asparagus farmers in Rio Grande do Sul.*]. Estudo de amostragem para pesquisa sócio-econômica junto a produtores de asparago no Rio Grande do Sul. *Pesquisa Agropecuária Brasileira, Série Agronomia* (1973) 8, 193-195 [Pt, en] IPEAS, Caixa Postal E, Pelotas, RS, Brazil.

A sampling method is described for a social and economic research study with asparagus farmers.