



EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA
UNIDADE DE EXECUÇÃO DE PESQUISA DE ÂMBITO
ESTADUAL DE MANAUS

1º SIMPÓSIO BRASILEIRO DO GUARANA

24 a 28 de outubro de 1983

Manaus, AM

ANAIS

Manaus, AM

1983

PROGRAMA PARA SELECIONAR AMOSTRA ALEATÓRIA
ESTRATIFICADA NUMA POPULAÇÃO
DE GUARANAICULTORES

Renato Abeilar R. Gomes¹
Sônia Milagres Teixeira²

APRESENTAÇÃO

O trabalho foi realizado na UEPAE de Manaus, utilizando-se a linguagem BASIC para Microcomputador POLYMAX SS-100, com cinco etapas: leitura dos dados contendo nome dos produtores e área plantada de guaraná, ordenação de áreas plantadas, classificação por estrato de área plantada, cálculo do tamanho das amostras e das subamostras por estrato, seleção randômica e impressão da lista de produtores selecionados. Foi elaborado para servir de suporte a uma pesquisa de caracterização sócio-econômica da guaranaicultura no Estado do Amazonas.

Recolheram-se as informações quanto a nomes dos produtores e área plantada de guaraná, compondo uma listagem de população total de produtores atendidos pela EMATER, mutuários da carteira de crédito dos diversos bancos atuando no Estado e guaranaicultores que compõem o quadro de cooperados da CAMAL (Cooperativa Agrícola de Maués Ltda.)

Essas listagens foram checadas para eliminar duplicidades, tendo sido incluídos 1145 produtores na lista total. Tal listagem é suficientemente abrangente, tendo sido comparada a outras existentes no CODEAMA e FIBGE, órgãos que mantêm estatísticas agrícolas no Estado.

¹ Eng^o Agrícola, estagiário da EMBRAPA/UEPAE de Manaus

² Economista, Ph.D em Economia Rural, pesquisador da EMBRAPA/UEPAE de Manaus

A formulação permitiu selecionar um total de 10% dos produtos, com o cálculo do número de elementos da amostra e sub-amostras baseado na fórmula desenhada por Bowley (1971). Utilizaram-se valores de t com 95% de probabilidades e o erro máximo permitido (d) variando segundo áreas médias dos municípios. Os resultados da amostragem, bem como a distribuição da população de guaranaicultores podem ser analisados no trabalho de Teixeira (1983).

O programa poderá ser utilizado em pesquisas sócio- econômicas ou diagnósticos, onde a seleção aleatória de amostras deva ser realizada.

O procedimento para amostragem randômica estratificada de produtores pode ser visualizado pelo fluxograma da Figura 1. Na ocasião da leitura dos dados, constituídos pelos nomes dos produtores, suas áreas plantadas e as localidades em que se encontram, pode-se optar por uma análise de toda a população ou fazê-lo por sub-regiões, delimitadas pelas localidades distinguidas nas listagens de dados.

O acesso a estas informações pelo programa principal, deverá ser efetuado a partir de um arquivo de dados previamente elaborado, que permita sua fácil manipulação quando se desejar um acréscimo no seu conteúdo.

Para se proceder a ordenação dos dados, empregamos basicamente o método "BUBBLE SORT", através de uma subrotina do programa principal. Para listagens demasiado extensas, aconselha-se buscar um método mais eficiente. O fluxograma da Figura 2 nos dá uma visão do processo.

O conteúdo das variáveis que deverão ser ordenadas são testados dois-a-dois por intermédio do comando iterativo em (4). Se a condição em (5) não se estabelecer, haverá a troca dos valores testados e daqueles que lhe são relacionados. Uma vez percorridos todos os valores, a condição (2) é testada. "K" é uma variável lógica que indica se houve trocas durante as iterações ou se isto não ocorreu (ordenação completa), assumindo os valores "um" ou "zero". Caso a condição em (2) não se estabeleça, haverá novas iterações até que isto ocorra, quando então haverá o retorno ao programa principal. A variável "ULT" é empregada para uma dinamização do processo. Sua finalidade é fazer com que os valores sejam testados somente até onde foi detectada a última troca na iteração anterior, uma vez que, a partir desta os dados deverão estar obrigatoriamente ordenados.

Uma vez lidos todos os dados da população, o programa passa para a classificação dos mesmos por estratos. Em nosso caso particular, os produtores de guaraná foram enquadrados em 5 estratos de

acordo com suas áreas plantadas.

Estrato 1:	menos de 5 ha
Estrato 2:	de 5 a 10* ha
Estrato 3:	de 10 a 20* ha
Estrato 4:	de 20 a 30* ha
Estrato 5:	mais de 49 ha
	(*)exclusive

Para cada um dos estratos serão determinados:

- O número de elementos do estrato (W_i)
- O somatório dos valores de área (SX_i)
- O somatório dos quadrados dos valores de área (SQX_i)
- A variância de cada estrato (V_i)

Onde "i" é o índice que referencia o estrato (i=1 a 5)

O fluxograma da Figura 3 ilustra o procedimento. O vetor "Hj" contendo os valores de área é percorrido até o seu "W-ésimo" termo (W=número total de produtores da população a ser amostrada), através do comando iterativo em (1). A cada iteração, o "j-ésimo" valor de "Hj" é submetido a uma bateria de testes, quando então, o estrato a que este pertence será definido pelo índice "i" e o bloco de operações (6) será executado. Estando todos os dados estratificados, parte-se para a determinação de " W_i ", " SX_i ", " SQX_i " e " V_i ". (SVW é o somatório dos produtos da variância pelo nº de elementos de cada estrato).

OBS: Caso o número de elementos do estrato seja "1" ou "0" a variância é assumida como nula, para se evitar a "divisão por zero" na expressão (10).

A próxima etapa consiste nas determinações do tamanho da amostra para os estratos considerados. Os cálculos foram efetuados com base nas fórmulas Bowley:

$$N = \frac{t^2 \cdot SVW}{(W \cdot d^2) + (t^2 \cdot SVW/W)}$$

$$N_i = W_i \cdot \frac{N}{W}$$

Onde:

N = Tamanho da amostra total

N_i = Tamanho da amostra para o estrato i

W = Número de elementos da população

W_i = Número de elementos do estrato i

t = Fator de correção segundo o grau de risco assumido

d = Erro máximo permitido, representado por uma porcentagem da área média de guaraná no município.

O processo, esquematizado na Figura 4, tem início com a leitura dos valores para " t " e " d ", introduzidos através do console. O computador informará através do "display" ou impressora, qual o tamanho da amostra calculado pela expressão de Bowley em (3). Caso o valor encontrado fique além ou aquém das expectativas, pode-se lançar mão de um artifício, confirmando o resultado ou testando novos valores para " t " e " d " através do teste em (7). Na primeira hipótese, o processo seguirá com o cálculo do tamanho das amostras para cada estrato (9).

A seleção randômica por estratos só é possível quando se tem os dados ordenados. O Algoritmo do processo (fluxograma Figura 5) foi elaborado pelo autor aproveitando as facilidades da linguagem BASIC-80 na geração de números aleatórios.

Quando a subrotina é cessada, cada estrato é analisado sequencialmente pelo comando iterativo em (3), o índice " i " é incrementado até que seu valor ultrapasse o número de estratos previstos, quando então haverá o retorno ao programa principal. Satisfeita a condição (4), é impresso o número do estrato e o tamanho de sua amostra.

O intervalo de amostragem é definido pelas variáveis inteiras LI e LS. Estas conterão os números de ordem dos dados que se encontram nos extremos inferior e superior do estrato " i ".

A sequência randômica, para ser gerada, necessita de um número que a origine. Este número pode ser introduzido externamente pelo operador ou calculado pelo próprio computador se a linguagem dispõe de recursos para isso.

Em (9), tem início o laço que gerará os elementos da amostra que a condição de saída " $j > N_i$ " (o número de randômicos gerados e não repetidos ultrapassou o limite previsto para o estrato i) seja satisfeita. O número randômico é gerado dentro de um intervalo de " \emptyset " a "1" e deve ser integrado em uma função que retorne valores no intervalo desejado. Assim temos:

$$R_j = \text{RAND} * (\text{LS} - \text{LI}) + \text{LI}$$

Onde:

R_j : j -ésimo número não repetido gerado randomicamente no intervalo $[\text{LI}, \text{LS}]$

RAND: Número randômico gerado no intervalo $[0,1]$

LS, LI: Extremos superior e inferior do intervalo desejado

O comando iterativo em (11), testa se o randômico obtido é repetido através da condição em (2). Em caso afirmativo, " j " sofrerá um decremento e provocará uma nova iteração até que ocorra o contrário. A variável " K " é do tipo lógico e indica se houve repetição assumindo os valores " \emptyset " ou "1". No segundo caso ela evitará que sejam impressos os dados relativos a " j -ésima" iteração.

CONCLUSÕES

A formulação desenvolvida para um estudo específico dos aspectos sócio-econômicos da cultura do guaraná no Estado do Amazonas, poderá ser utilizada em situações similares em que a população a ser estudada e quaisquer parâmetros, sejam área total do estabelecimento, área de pastagem, número de animais no rebanho, etc. possam ser coletados, antes de realizar a pesquisa.

O procedimento permite calcular o tamanho da amostra representativa da população a ser pesquisada, subamostras nos diversos estratos e selecionar aleatoriamente, na população, proprietários a serem visitados no levantamento.

REFERÊNCIAS BIBLIOGRÁFICAS

BOWLEY, W. **Mathematical statistics an inferences**, New York, John Willey & Sons. 1971, 345 p.

FEIXEIRA, S. M. **Caracterização da guaranaicultura no Estado do Amazonas**. Manaus, EMBRAPA/UEPAE, 1983.

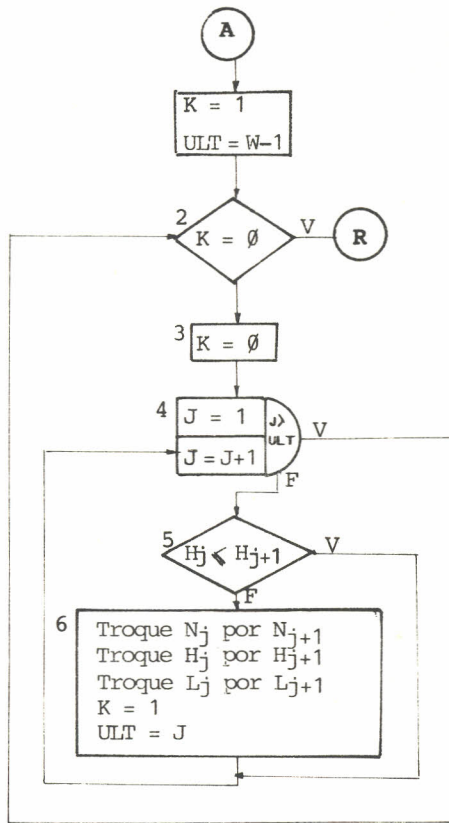


FIGURA 2.

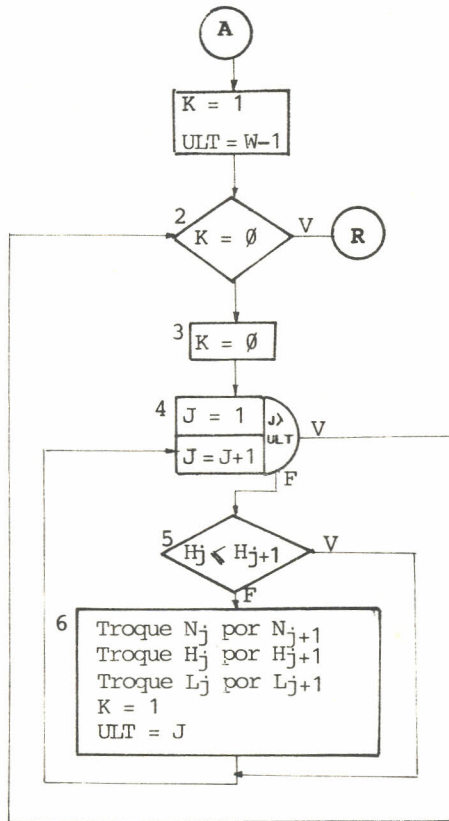


FIGURA 2.

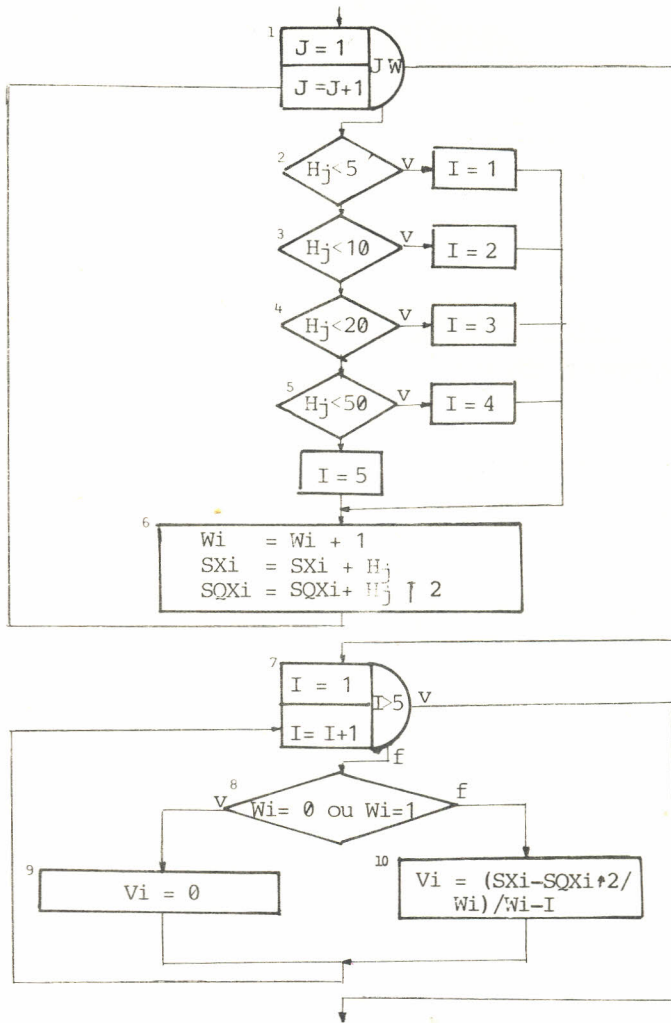


FIGURA 3.

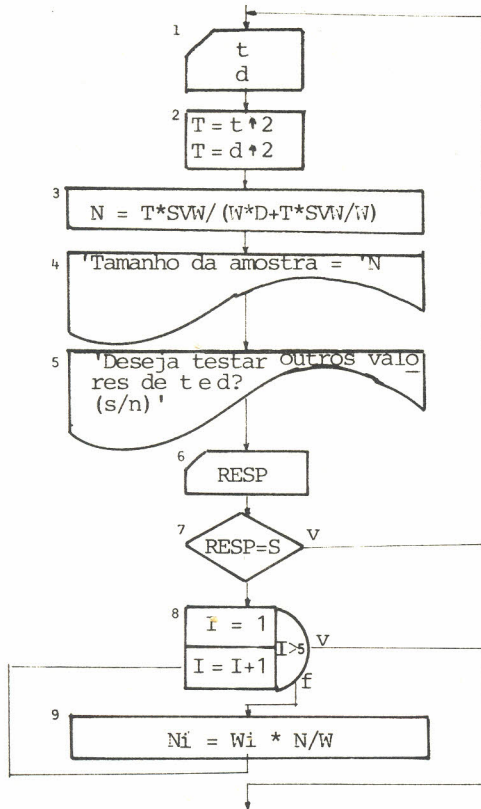


FIGURA 4.

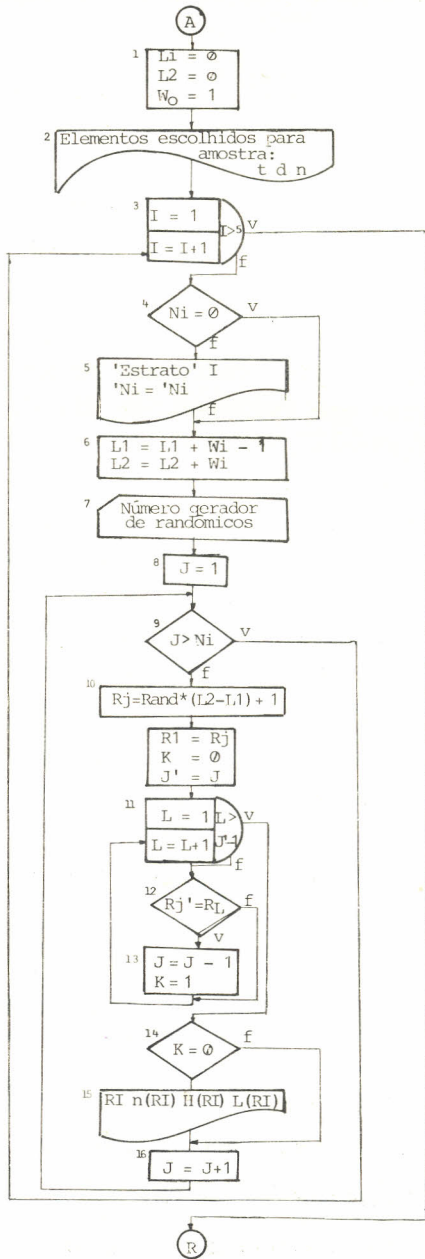


FIGURA 5.