

# ALGUMAS APLICAÇÕES DO SAS<sup>®</sup> EM ECONOMIA AGRÍCOLA<sup>1</sup>

Francisco Alberto Pino<sup>2</sup>  
Vera Lucia F. Santos Francisco<sup>3</sup>  
Sergio Augusto Galvão Cezar<sup>4</sup>  
Ana Maria Pereira Amaral<sup>4</sup>  
Maria de Lourdes Sumiko Sueyoshi<sup>5</sup>

---

<sup>1</sup>Trabalho apresentado ao 2º GUSAS, Congresso Brasileiro de Usuários SAS, Campinas, UNICAMP, 30/11 a 03/12/92.  
Recebido em 18/12/92. Liberado para publicação em 14/01/93.

<sup>2</sup>Engenheiro Agrônomo, Dr., Pesquisador Científico do Instituto de Economia Agrícola.

<sup>3</sup>Estatístico, Pesquisador Científico do Instituto de Economia Agrícola.

<sup>4</sup>Engenheiro Agrônomo, MS, Pesquisador Científico do Instituto de Economia Agrícola.

<sup>5</sup>Matemática, Pesquisador Científico do Instituto de Economia Agrícola.

## 1 - INTRODUÇÃO

O SAS<sup>®</sup> (*Statistical Analysis Software*), em versão para microcomputador, foi introduzido em 1990 pelo Centro de Métodos Quantitativos nos trabalhos científicos e técnicos do Instituto de Economia Agrícola, contando com os módulos básico (SAS/BASE), estatístico (SAS/STAT<sup>™</sup>), econométrico (SAS/ETS<sup>®</sup>), de pesquisa operacional (SAS/OR<sup>®</sup>) e de aplicações (SAS/AF<sup>®</sup>). Para o treinamento dos usuários diversos cursos internos foram ministrados e um trabalho foi publicado (PINO et alii, 1992), contendo um capítulo sobre introdução ao SAS<sup>®</sup>. Algumas das aplicações mais freqüentes, porém, não imediatas, são apresentadas a seguir: amostragem estratificada, números índices e taxas de crescimento. Para as três basta o módulo SAS/BASE (SAS INSTITUTE, 1988a, 1988b).

Usa-se a notação <...> para indicar valores ou nomes a serem supridos pelo usuário. Os arquivos de dados podem ser digitados em qualquer editor de texto, desde que seja em formato ASCII.

## 2 - AMOSTRAGEM

A teoria de amostragem pode ser encontrada em livros texto, como COCHRAN (1953) e KISH (1965). De modo geral, trata-se da obtenção de totais e médias, bem como dos respectivos desvios padrões e erros de amostragem. As sugestões aqui apresentadas servem somente para amostragem probabilística.

### 2.1 - Cálculo do Tamanho da Amostra

Conforme KISH (1965), para uma amostra casual simples, o número de elementos da amostra é dado por:

$$n = \frac{n'}{1 + \frac{N-1}{N} \frac{CV^2}{C^2}}$$

com:

onde  $N$  é o tamanho da população,  $\bar{Y}$  é a média da variável de interesse na população,  $S^2$  é a variância da variável de interesse na

população,  $V^2$  é a variância desejada para a estimativa da média,  $CV^2$  é o coeficiente de variação (erro de amostragem) desejado para a estimativa da média.

Usualmente, quer se calcular  $n$ ; sendo conhecidos  $N$  e  $C^2$  (ou  $S^2$ ), cabe ao pesquisador fixar  $CV^2$  (ou  $V^2$ ). Os valores referentes à população podem, eventualmente, ser obtidos de um levantamento anterior ou de uma pesquisa piloto.

Para o cálculo no SAS<sup>®</sup> apresenta-se a seguinte sugestão:

```
data <nome1>;
input ene s2 media erro;
  np = s2 / (media * media * erro * erro);
  n = np / (1 + np / ene);
cards;
<tamanho da população> <variância da população> <média da população> <erro de amostragem desejado>;
run;
proc print;
run;
```

## 2.2 - Alocação da Amostra em Estratos

Dada uma amostra de tamanho  $n$ , seus elementos podem ser alocados em estratos de forma diretamente proporcional ao tamanho do estrato na população e ao desvio padrão e inversamente proporcional ao custo de levantamento (KISH, 1965):

$$n_h = n \frac{\frac{N_h S_h}{\sqrt{J_h}}}{\sum_{i=1}^n \frac{N_i S_i}{\sqrt{J_i}}}$$

onde  $n_h$  representa o número de elementos da amostra a serem alocados no estrato  $h$ ,  $N_h$  é o número de elementos da população no estrato  $h$ ,  $S_h$  é o desvio padrão da variável em estudo no estrato  $h$  e  $J_h$  é o custo de

levantamento no estrato  $h$ .

Para o cálculo no SAS<sup>®</sup> apresenta-se a seguinte sugestão:

```

data <nome2>;
input amostra $ estrato ene s j;
  v = ene * s / sqrt(j);
cards;
a <código do estrato> <número de elementos da população no estrato> <desvio padrão no estrato> <custo de
levantamento no estrato>
;
proc means noprint;
  by amostra;
  var v;
  output out=saida sum=soma;
run;
data numero;
input amostra $ n;
cards;
a <numero total de elementos na amostra>;
;
data otimo;
  merge saida numero <nome2>;
  by amostra;
run;
data aloca (keep=amostra estrato nh);
  set otimo;
  nh = n * v / soma;
run;
proc print;
  var amostra estrato nh;
run;

```

## 2.3 - Processamento dos Dados Levantados

Considere-se o caso de uma amostra estratificada em que se levantaram dados sobre  $m$  variáveis. A obtenção de resultados pode ser feita em quatro etapas: criação do arquivo de pesos, criação do

arquivo de dados, criação da biblioteca e processamento dos dados.

### 2.3.1 - Criação do arquivo de pesos

O número de elementos na população e na amostra deve ser digitado num editor de textos qualquer (formato ASCII), com a seguinte disposição (um estrato em cada linha):

<código do estrato> <número de elementos da população neste estrato> <número de elementos da amostra neste estrato>

Esses dados são gravados em disquete num arquivo chamado (por exemplo):

pesos.sas

### 2.3.2 - Criação do arquivo de dados

Os dados devem ser digitados num editor de textos qualquer (formato ASCII), com a seguinte disposição (um dado em cada linha):

<código do informante> <código do estrato> <código da variável> <dado ou valor>

com  $n.m$  26 linhas por 4 colunas. O código do informante refere-se a cada elemento da amostra. Alternativamente poder-se-ia digitar:

<código do informante> <código do estrato> <dado da variável1> ... <dado da variável m>

resultando  $n$  27 linhas por  $m + 2$  28 colunas, mas considerar-se-á aqui somente o primeiro caso.

Os dados são gravados em disquete num arquivo chamado:

<nome do arquivo>.<ext>

### 2.3.3 - Criação da biblioteca

Entra-se no SAS<sup>®</sup> com o disquete de dados no *drive* A e cria-se a biblioteca:

```
libname <nome da biblioteca> 'a';
data <nome da biblioteca>.pesos;
infile 'a:pesos.sas';
input estrato ene numero;
run;
data <nome da biblioteca>.<nome do arquivo>;
infile 'a:<nome do arquivo>.<exp>';
input codigo estrato variavel valor;
run;
```

### 2.3.4 - Processamento

O procedimento MEANS é suficiente para os cálculos, embora outros possam ser utilizados. Sugere-se para o título principal o nome do levantamento ou do projeto e para o título secundário a data e/ou o local do levantamento. Para o cálculo no SAS<sup>®</sup> apresenta-se a seguinte sugestão:

```

options linesize=90 pagesize=54;
libname <nome da biblioteca> 'a';
title1 '<título principal>';
title2 '<título secundário>';
proc sort data=<nome da biblioteca>.<nome do arquivo>;
  by variavel estrato;
run;
proc means data=<nome da biblioteca>.<nome do arquivo> noprint;
  by variavel estrato;
  var valor;
  output out=saida n=n mean=media var=var;
run;
proc sort data=saida;
  by estrato;
run;
data calculo;
  merge saida <nome da biblioteca>.pesos;
  by estrato;
run;
data calculo (replace=yes);
  set calculo;
  peso = ene / numero;
  total = ene * media;
  vmedia = (1 - numero / ene) * var / numero;
  vtotal = ene * ene * vmedia;
  erro = vmedia * 100 / ( media * media );
  x = peso * media;
  y = peso * peso * vmedia;
run;
proc sort data=calculo;
  by variavel estrato;
run;
proc print data=calculo;
  var ene numero n media vmedia total vtotal erro x y;
  by variavel estrato;
  pageby variavel;
  sum ene numero n total vtotal x y;
run;

```

## 3 - NÚMEROS ÍNDICES

A caracterização dos números índices pode ser encontrada, por exemplo, em SILVA & CARMO (1986), ou em HOFFMANN (1980). Tomou-se por base os índices de quantidade.

O índice de Laspeyres para quantidade é dado por

$$L = \frac{\sum_{i=1}^n p_{i0} q_{it}}{\sum_{i=1}^n p_{i0} q_{i0}}$$

onde  $p_{it}$  representa o preço do produto  $i$  no ano  $t$  e  $q_{it}$  representa a quantidade do produto  $i$  no ano  $t$ .

dado por:

$$P = \frac{\sum_{i=1}^n p_{it} q_{it}}{\sum_{i=1}^n p_{it} q_{i0}}$$

O índice de Paasche para quantidade é

O índice de Fischer para quantidade é dado por:

dado por:

$$T = \prod_{i=1}^n \left( \frac{q_{it}}{q_{i0}} \right)^{\frac{w_{i0} + w_{it}}{2}}$$

$$w_{it} = \frac{p_{it} q_{it}}{\sum_{i=1}^n p_{it} q_{it}}$$

O índice de Törnqvist para quantidade é

com:

A obtenção de resultados pode ser feita em duas etapas: criação do arquivo de dados e processamento dos dados.

### 3.1 - Criação do Arquivo de Dados

Os dados devem ser digitados num editor de textos qualquer (formato ASCII), com a seguinte disposição em cada linha:

<ano> <produto> <preço> <quantidade>

Eventualmente, o ano pode ser substituído por mês, observação ou código equivalente. Os códigos de ano e produto devem ser numéricos, embora o usuário possa alterar convenientemente o programa para aceitar códigos alfanuméricos. Além disso, para calcular índice de preço, basta inverter no arquivo a coluna de preço com a de quantidade.

Os dados são gravados em disquete num arquivo chamado:

<nome do arquivo>.<ext>

por exemplo, "NUMI31.PGM". Note-se que este arquivo deve ser passível de leitura pelo SAS<sup>®</sup>.

### 3.2 - Processamento dos Dados

Entra-se no SAS<sup>®</sup> com o disquete de dados no *drive* A. Para o cálculo no SAS<sup>®</sup> apresenta-se a seguinte sugestão, que permite até 30 produtos (o usuário poderá alterar facilmente este número):

```

options linesize=90 pagesize=54;
*
* CRIACAO DA BIBLIOTECA E LEITURA DOS DADOS;
*
libname indice 'a';
title1 'TITULO DO TRABALHO';
title2 'Numeros Indices';
data indice.numi31;
infile 'a:numi31.pgm';
input t i p q;
*
* CLASSIFICACAO DOS DADOS E CALCULO DE SOMAS DE PRODUTOS;
*
proc sort data=indice.numi31;
  by t i;
run;
data transf;
  set indice.numi31;
  array pzero{30};
  array qzero{30};
  retain tz pzero qzero;
  if _n_ = 1 then tz=t;
  if t=tz then do;
    pzero{i} = p;
    qzero{i} = q;
  end;
  pzqt = pzero{i} * q;
  pzqz = pzero{i} * qzero{i};
  ptqz = p * qzero{i};
  ptqt = p * q;
run;
proc means data=transf noprint;
  var pzqt pzqz ptqz ptqt;
  by t;
  output out=somas sum= szt szz stz stt;
run;
*
* CALCULO DOS INDICES DE LASPEYRES, PAASCHE E FISCHER;
*
data somas;
  set somas;
  laspeyre = szt * 100 / szz;
  paasche = stt * 100 / stz;
  fischer = sqrt (laspeyre * paasche);
run;

```

```

*
*
* CALCULO DO INDICE DE TORNQVIST;
*
*
data torn;
merge somas transf;
  by t;
run;
data torn;
  set torn;
  array wzero{30};
  array qzero{30};
  retain wzero;
  w = p * q / stt;
  if t = tz then wzero{i} = w;
  x = q / qzero{i};
  y = (wzero{i} + w) / 2;
  u = y * log(x);
run;
proc means data = torn noprint;
  var u;
  by t;
  output out = saida sum = su;
run;
data saida;
  set saida;
  tornqvis = exp(su) * 100;
run;
data tornq;
  merge somas saida;
  by t;
run;
proc print data=tornq noobs;
  var t laspeyre paasche fischer tornqvis;
run;

```

#### 4 - TAXA DE CRESCIMENTO

O cálculo da taxa de crescimento com base em  $n$  36 valores de uma dada variável pode ser obtido através de regressão linear no período correspondente (HOFFMANN et alii, 1976).

Considere-se  $Z_t$  37 como os valores da variável em estudo e  $t$  38o número de períodos ( $t = 0, 1, 2, \dots, n - 1$ ). 39 Seja

$$Z_t = A(1+r)^t$$

onde  $r$  41 é a taxa de crescimento e  $A$  42 corresponde ao valor inicial da série.

Aplicando logaritmos naturais à expressão acima, tem-se:

$$\log Z_t = \log A + t \log(1+r)$$

$$Y = a + bt$$

$$\hat{Y} = \log \hat{Z}_t$$

que corresponde a  
sendo



$$b = \log(I + r)$$

Depois de estimados os parâmetros, tem-

se:

$$\hat{b} = \log(I + \hat{r}) \quad \text{e, portanto,}$$

$$\hat{r} = e^{\hat{b}} - I$$

#### 4.1 - Criação do Arquivo de Dados

Os dados devem ser digitados num editor de texto qualquer (formato ASCII), com a seguinte disposição em cada linha:

<seqüência de ordem da observação> <valor da variável>

Estes dados deverão ser gravados em disquete num arquivo chamado:

<nome do arquivo>.<ext>

#### 4.2 - Criação da Biblioteca

Entra-se no SAS<sup>®</sup> com o disquete de dados no *drive* A e cria-se a biblioteca:

```
libname <nome da biblioteca> `a:';
data <nome da biblioteca>.<nome do arquivo>;
infile `a: <nome do arquivo>.<ext>';
input t z;
  logz=log (z);
  tlogz=t * log (z);
run;
```

#### 4.3 - Processamento

Para o cálculo no SAS<sup>®</sup>, por mínimos quadrados, apresenta-se a seguinte sugestão:

```
options linesize=90 pagesize=54;
libname <nome da biblioteca> `a:';
title 'Taxa de Crescimento (%)';
proc means data=<nome da biblioteca>.<nome do arquivo> noprint;
  var t logz tlogz;
  output out = saida
  sum=st slogz stlogz
  uss=sqrt n=n;
run;
data saida;
  set saida;
  b=(stlogz-st*slogz/n)/(sqrt-st*st/n);
  taxa = 100 * (exp(b)-1);
```

```
run;  
proc print data=saida;  
var taxa;  
run;
```

#### LITERATURA CITADA

COCHRAN, Willian G. **Sampling tech-niques**. New York, John Wiley, 1953. 330p.

HOFFMANN, Rodolfo. **Estatística para economistas**. São Paulo, Pioneira, 1980. 379p.

\_\_\_\_\_. et alii. **Administração da empresa agrícola**. São Paulo, Pioneira, 1976. 323p.

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

PINO, Francisco A. et alii. Cursos introdutórios de informática. **Informações Econômicas**, SP, **22** (supl, 01):1-176, 1992.

SAS INSTITUTE. **SAS<sup>®</sup> language guide for personal computers**: release, 6. 3.ed. Cary, NC, 1988a.

\_\_\_\_\_. **SAS<sup>®</sup> procedures guide**: release, 6. 3.ed. Cary, NC, 1988b.

SILVA, Gabriel L.S.P. & CARMO, Heron C.E. **Como medir a produtividade agrícola**: conceitos, métodos e aplicações no caso de São Paulo. São Paulo, IEA, 1986. 29p. (Relatório de Pesquisa, 3/86).