

# UTILIZAÇÃO DO SAS® EM ESTUDOS DE SAZONALIDADE<sup>1</sup>

Vera Lucia Ferraz dos Santos Francisco<sup>2</sup>

Francisco Alberto Pino<sup>3</sup>

Maria de Lourdes Sumiko Sueyoshi<sup>4</sup>

Sérgio Augusto Galvão César<sup>5</sup>

Ana Maria Pereira Amaral<sup>5</sup>

## 1 - INTRODUÇÃO

O Centro de Métodos Quantitativos vem introduzindo nos trabalhos científicos e técnicos do Instituto de Economia Agrícola, em sazonalidade e ajustamento sazonal, a utilização do SAS® (*Statistical Analysis Software*), em versão para microcomputador, contando com dois módulos: SAS/STAT™ e SAS/ETS®. Neste trabalho apresentam-se instruções básicas sobre os seguintes métodos:

a) versão X-11 do método II do Bureau do Censo, Estados Unidos, a ser resolvido com o procedimento X11 do módulo SAS/ETS®;

b) análise espectral, a ser resolvida com o procedimento SPECTRA do módulo SAS/ETS;

c) método baseado em modelos de regressão, a ser resolvido com o procedimento GLM do módulo SAS/STAT™; e

d) método baseado em modelos ARIMA, a ser resolvido com o procedimento ARIMA do módulo SAS/ETS®.

A descrição dos métodos encontra-se em PINO et alii (1994). Para iniciar-se no uso do SAS® recomendam-se PINO et alii (1992) ou SAS INSTITUTE (1988a). Para melhor entendimento dos procedimentos recomendam-se os manuais do SAS

INSTITUTE (1988b e 1988c).

### 1.1 - Arquivos de Dados

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. Neste trabalho adotou-se a seguinte disposição de registros para dados mensais:

<ano> <mês> <dado ou valor>

Cada linha corresponde a um mês, sendo este representado por um número: 1 para janeiro, 2 para fevereiro, etc. Para dados trimestrais a disposição é análoga. Se existir mais de uma variável (série), seus dados serão digitados em seqüência na mesma linha, isto é, cada coluna representa uma variável. Grava-se em disquete num arquivo chamado, por exemplo:

<nome do arquivo>.<ext>

Se for necessário importar dados, por

<sup>1</sup>Trabalho apresentado no III Congresso Brasileiro do Grupo de Usuários do SAS (GUSAS), Londrina, PR, 15 a 18 de março de 1994. Recebido em 25/03/94. Liberado para publicação em 06/06/94.

<sup>2</sup>Estatístico, Pesquisador Científico do Instituto de Economia Agrícola.

<sup>3</sup>Engenheiro Agrônomo, Dr., Pesquisador Científico do Instituto de Economia Agrícola.

<sup>4</sup>Matemático, Pesquisador Científico do Instituto de Economia Agrícola.

<sup>5</sup>Engenheiro Agrônomo, MS, Pesquisador Científico do Instituto de Economia Agrícola.

exemplo do LOTUS™, ver PINO et alii (1992). A seguir, acessa-se o SAS®. Com o disquete de dados no *drive* A, cria-se a biblioteca, digitando-se na janela PGM:

```
libname <nome da biblioteca> 'a:';
data <nome da biblioteca>.<nome do arquivo>;
infile 'a:<nome do arquivo>.<ext>';
input ano mes valor;
date = mdy(mes,1,ano);
format date monyy5.;
run;
```

e submetem-se essas instruções à execução (teclar F10).

Apresentam-se também outras instruções alternativas de elaboração de arquivo de dados:

a) as palavras *ano*, *mês* e *valor* usadas acima podem ser alteradas pelo usuário, desde que estas sejam feitas ao longo de todo o trabalho;

b) se houver mais de uma variável (série), por exemplo cinco, a quarta linha do conjunto de instruções acima pode ser modificada para:

```
input mes ano valor1-valor5;
```

c) caso o arquivo contenha uma só variável, mas cada linha do registro corresponda a um ano, com a seguinte disposição dos dados:

```
<ano> <valor de janeiro> <valor de fevereiro>
<valor de março> <valor de abril> <valor de maio>
<valor de junho> <valor de julho> <valor de agosto>
<valor de setembro> <valor de outubro> <valor de novembro>
<valor de dezembro> ,
```

o conjunto de instruções poderá ser modificado para:

```
libname <nome da biblioteca> 'a:';
data <nome da biblioteca>.<nome do arquivo>;
drop i val1-val12;
array val[12] val1-val12;
infile 'a:<nome do arquivo>.<ext>';
input ano val1-val12;
do i = 1 to 12;
mes = i;
```

```
valor = val[i];
date = mdy(mes,1,ano);
format date monyy5.;
output;
end;
run;
```

## 2 - MÉTODO DO BUREAU DO CENSO (VERSÃO X-11)

Este método consiste na decomposição da série original em componentes sazonal, de ciclo-tendência, ajustamento ao calendário e irregular:

$$O = S \cdot C \cdot TD \cdot I$$

no modelo multiplicativo

$$O = S + C + TD + I$$

no modelo aditivo onde, *O* é a série original,

*S* é a componente sazonal, *C* é a componente de ciclo-tendência, *TD* é a componente de influência do calendário e *I* é a componente irregular, residual ou aleatória. O método de decomposição foi desenvolvido inicialmente por Frederick R. Macaulay no National Bureau of Economic Research, Estados Unidos (NBER), em 1954. O Bureau do Censo utiliza-o em larga escala através do programa de computador desenvolvido por Julius Shiskin. Desde então o Bureau do Censo vem desenvolvendo o programa com versões mais refinadas, culminando em 1965 com a versão X-11.

Apresenta-se a seguinte sugestão para dados mensais com uma variável *x*. Digitar e submeter com o disquete de dados no *drive* A:

```
libname <nome da biblioteca> 'a:';
options linesize=90 pagesize=60;
proc x11 data = <nome da biblioteca>.<nome do arquivo>;
monthly date = date charts = all;
var valor;
tables b1 d10 d11 d12 d13 e1 e2 e3 e4 e5 e6 f2;
output out = saida d10 = tabd10 d11 = tabd11;
```

```
title1 '<título do trabalho>';
title2 'Sazonalidade';
run;
```

Sugerem-se algumas observações importantes:

a) o procedimento X11 não pode ser utilizado em séries que contenham *missing values*. No caso do arquivo original de dados conter *missing value* deve-se indicar nas instruções o ponto a partir do qual a série está livre do *missing value*, alterando a terceira linha da sugestão acima para:

```
proc x11 data=<nome da biblioteca>.<nome do
arquivo> (firstobs=<n>);
```

onde *n* indica a primeira informação após a ocorrência do *missing value*; se esta for a informação de número *y* então, *n* será *y+1*;

b) as instruções acima servem para modelos multiplicativos. Para modelos aditivos a quarta linha das instruções acima deve ser alterada para:

```
monthly additive date=date charts=all;
```

c) se somente uma parte da série for utilizada no ajustamento sazonal, a quarta linha das instruções acima deve ser alterada para:

```
monthly date=date charts=all start=<mmmy>
end=<mmmy>;
```

onde *start* refere-se ao primeiro mês e *end* refere-se ao último mês do período a ser considerado no ajustamento sazonal. As letras *mmmy* devem ser substituídas pela abreviatura do mês em inglês (jan, feb, mar, apr, may, jun, jul, aug, sep, oct, nov, dec) e o ano; por exemplo, *dec70* significa dezembro de 1970;

d) se o usuário quiser especificar o número de decimais nas tabelas de resultados, basta alterar a quarta linha das instruções acima para:

```
monthly date=date charts=all ndec=<número de
decimais>;
```

e) a quinta linha das instruções acima especifica a variável do arquivo de dados que será utilizada no processamento, lembrando que o modelo é univariado;

f) a instrução *output* na sétima linha da receita acima permite guardar os fatores sazonais e a série ajustada sazonalmente num arquivo para uso posterior, por exemplo, para fazer gráficos no SAS® ou em outro *software*; e

g) para dados trimestrais basta substituir a palavra *monthly* por *quarterly* na quarta linha das instruções acima, valendo as demais considerações sobre aquela linha.

## 2.1 - Tabelas de Resultados e Gráficos

O procedimento X11 permite a impressão de até 63 tabelas, com as diversas etapas dos cálculos e de até 4 gráficos. Sugere-se a utilização das seguintes tabelas:

- a) tabela B1: série original;
- b) tabela D10: fatores sazonais finais;
- c) tabela D11: série sazonalmente ajustada final;
- d) tabela D12: ciclo-tendência final;
- e) tabela D13: série irregular final;
- f) tabela E1: série original sem valores estranhos (*outliers*);
- g) tabela E2: série sazonalmente ajustada modificada;
- h) tabela E3: série irregular modificada;
- i) tabela E4: razões entre totais anuais;
- j) tabela E5: mudanças percentuais na série original;
- l) tabela E6: mudanças percentuais na série sazonalmente ajustada final; e
- m) tabela F2: medidas de resumo.

Já para os gráficos sugerem-se os quatro:

- a) gráfico G1: série sazonalmente ajustada final e estimativa da componente de ciclo-tendência (D11 e D12).
- b) gráfico G2: razões S-I e fator sazonal final mês a mês;
- c) gráfico G3: razões S-I em ordem normal do

calendário diário ;

d) gráfico G4: estimativa final da série irregular e da série irregular modificada.

### 3 - ANÁLISE ESPECTRAL

A análise espectral, também conhecida por análise harmônica, estuda a série no domínio da frequência transformando a série na forma de combinação linear de senos e co-senos com representação dada por:

$$z_t = z(t) = \sum_{\lambda} A_{\lambda} \sin(\lambda t + \phi_{\lambda})$$

onde,  $t$  é o período, isto é, o tempo necessário para completar o ciclo,  $\lambda$  é a frequência angular (ou simplesmente frequência), medida em radianos por unidade de tempo, e dada por  $\lambda = 2\pi f$ ,  $A$  é a amplitude, medida na mesma unidade que as observações da série,  $\phi$  é a fase, sem dimensão, com  $|\phi| \leq \pi$  (Figura 1).

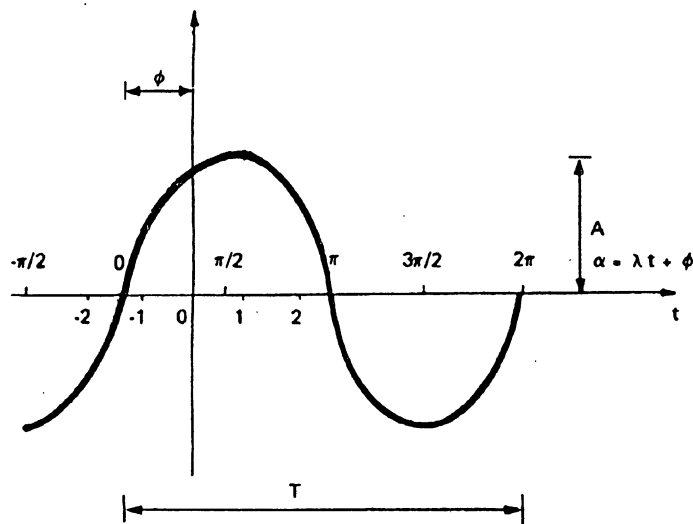


FIGURA 1 - Gráfico de  $A \sin(\lambda t + \phi)$ .

Fonte: PINO; NOGUEIRA JUNIOR; TOLOI (1983).

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

```
libname <nome da biblioteca> 'a:';
proc spectra data= <nome da biblioteca>. < nome
do arquivo> p s;
var valor;
weights 1 2 3 4 5 4 3 2 1;
run;
proc print;
run;
```

Se for necessário exportar os dados, por exemplo para o LOTUS<sup>™</sup>, a fim de fazer o gráfico do periodograma, ver PINO et alii (1992).

### 4 - MODELOS DE REGRESSÃO

O método baseado em modelos de regressão considera uma decomposição semelhante ao descrito em 2, onde a sazonalidade pode ser escrita

$$S_t = \sum_{j=1}^s \alpha_j d_{jt}$$

onde  $s$  é o período da sazonalidade (em geral, igual a 12 para dados mensais) e  $d_{jt}$  são variáveis binárias (*dummy*) sazonais dadas por

$$d_{jt} = \begin{cases} 1, & \text{se } t \text{ corresponde ao mês } j \\ 0, & \text{caso contrário} \end{cases}$$

Suponha-se, ainda, que a tendência seja dada por um polinômio de grau  $m$

$$T_t = \sum_{i=0}^m \beta_i t^i$$

Então,

$$z_t = \sum_{j=1}^s \alpha_j d_{jt} + \sum_{i=0}^m \beta_i t^{i+a_t}$$

e os parâmetros  $\alpha$  e  $\beta$  podem ser estimados, por exemplo, por mínimos quadrados.

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

```
libname <nome da biblioteca> 'a:';
data <nome da biblioteca>.reg;
drop i;
set <nome da biblioteca>.<nome do arquivo>;
array d[11] d1-d11;
if mes < 12 then do;
do i=1 to 11;
if i=mes then d[i]=1;
else d[i]=0;
end;
end;
else
do i=1 to 11;
d[i]=-1;
end;
t=_n_;
t2=t*t;
run;
```

```
proc glm data = <nome da biblioteca>.reg;
model valor = t d1-d11;
run;
proc glm data = <nome da biblioteca>.reg;
model valor = t t2 d1-d11;
run;
```

## 5 - MODELOS ARIMA

Supõe-se que a série seja gerada por um processo auto-regressivo integrado de médias móveis, ARIMA (p,d,q) x (P,D,Q)<sub>s</sub>, conforme BOX & JENKINS (1976):

$$\phi_p(B) \Phi_P(B^s) (1-B)^d (1-B^s)^D z_t = \theta_q(B) \Theta_Q(B^s) a_t$$

Para a identificação do modelo no SAS®, sugere-se:

```
libname <nome da biblioteca> 'a:';
proc arima data = <nome da biblioteca>.<nome do
arquivo>;
identify var = valor;
run;
```

Analisar as autocorrelações e verificar se a série é estacionária. Se houver necessidade de diferença de ordem  $n$ , sugere-se:

```
libname <nome da biblioteca> 'a:';
proc arima data = <nome da biblioteca>.<nome do
arquivo>;
identify var = valor(n);
run;
```

Voltar a analisar as autocorrelações e a estacionariedade da série. A seguir para a estimação sugere-se:

```
libname <nome da biblioteca> 'a:';
proc arima data = <nome da biblioteca>.<nome do
arquivo>;
identify var = valor(n) noprint;
estimate q = <parâmetros de médias móveis>
p = <parâmetros auto-regressivos> noconstant
method = uls;
run;
```

## LITERATURA CITADA

- BOX, George E. P. & JENKINS, Gwilym M. **Times series analysis: forecasting and control.** San Francisco, Holden-Day, 1976.
- PINO, Francisco A.; NOGUEIRA JUNIOR, Sebastião; TOLOI, Clélia M. C.. Relações dinâmicas entre os preços da soja brasileira. **Pesquisa Agropecuária Brasileira**, Brasília, 18(11):1163-1173, 1983.
- \_\_\_\_\_. et alii. Cursos introdutórios de informática. **Informações Econômicas**, SP, 22 (Supl,1): 1-176, 1992.
- PINO, Francisco A. et alii. **Sazonalidade em séries temporais econômicas: teoria.** São Paulo, IEA/CMQ, 1994. no prelo.
- SAS INSTITUTE. **SAS<sup>®</sup> language guide for personal computers: release 6.** 3.ed. Cary, NC, SAS Institute, 1988a.
- \_\_\_\_\_. **SAS/ETS<sup>®</sup> user's guide.** Cary, NC, SAS Institute, 1988b.
- \_\_\_\_\_. **SAS/STAT<sup>™</sup> user's guide.** Cary, NC, SAS Institute, 1988c.

