



Mem/IPEA

Rio de Janeiro, 25 de setembro de 1968

De : Processamento de Dados  
Para: Coordenadores e Técnicos

Ao ensejo da realização do I Congresso Nacional de Processamento de Dados, no qual o IPEA se fez representar por sua equipe de processamento de dados, transmitimos aos técnicos do IPEA resumo da tese apresentada - TRADUTOR DA LINGUAGEM ESTATÍSTICA (TRALE), trabalho elaborado em conjunto com o Departamento de Cálculo Científico da Coordenação dos Programas Pós-Graduados de Engenharia da U.F.R.J.

O programa, pela grande versatilidade e facilidade de utilização, vem sendo eficientemente utilizado em mais de mil variados tipos de trabalhos, que não poderiam ser realizados sem o uso do computador.

Apesar da quantidade de tarefas executadas, não conseguimos exaurir, até hoje, todos os recursos que a linguagem do TRALE pode oferecer.

IPEA  
031

M. P. C. G. - Instituto de Pesquisa  
Econômico-Social Aplicada - (IPEA)  
SETOR DE DOCUMENTAÇÃO

ex-2

MINISTÉRIO DO PLANEJAMENTO INSTITUTO DE PESQUISA ECONÔMICO-SOCIAL APLICADA (IPEA) F N.º 2977 Data 16 / 10 / 68
---

## TRADUTOR DE LINGUAGEM ESTATÍSTICA - TRALE

O TRALE é um programa para ser utilizado nas análises de correlação e regressão de dados estatísticos.

O programa é comandado por pseudo-instruções que tornem bastante fácil a sua utilização por pessoal técnico não acostumado com o uso de computadores.

### Características

O programa considera que os dados podem ser sempre representados na forma de uma matriz, onde as colunas correspondem às várias observações de uma mesma variável aleatória. Assim, chamamos as colunas da matriz de dados de variáveis e as linhas de observações. O TRALE permite o máximo de 1.000 elementos dentro da matriz, embora não haja limitação quanto ao número de variáveis ou de observações, isto é, o produto

$$n^{\circ} \text{ de variáveis} \cdot n^{\circ} \text{ de observações} \leq 1.000$$

É possível manipular-se com as variáveis isoladamente, em qualquer fase de execução do programa.

### Utilização

Comandos- os comandos possuem uma forma fácil de ser assimilada pelos técnicos que deles se utilizarem. Compõem-se de duas partes: o código e os parâmetros.

O código é escrito com caracteres alfabéticos e lembram o nome do procedimento a ser efetuado. Somente as três primeiras letras dos comandos bastam para que o computador os execute. Os parâmetros são números inteiros ou reais que fornecem informações adicionais requeridas pelo procedimento.

Número inteiro é um número que necessariamente se escreve sem o ponto decimal enquanto que o número real é tratado como fração e deve ser escrito com ponto decimal. O computador só reconhece o ponto como sinal gráfico para separar a parte inteira da fracionária.

Cabe aqui lembrar que muito dos resultados são impressos no formato E. Este tipo de representação numérica é composto de duas partes: mantissa e expoente. A mantissa é a parte decimal (positiva ou negativa) e o expoente é a potência de 10 que se deve multiplicar pela mantissa, para obter o número. Exemplos:

$$0.9104910E 01 = 0.9104910 \times 10^1 = 9.104910$$

$$0.9104910E-03 = 0.9104910 \times 10^{-3} = 0,0009104910$$

$$-0.2600000E 00 = -0.2600000 \times 10^0 = -0,2600000$$

Passaremos agora a descrever os comandos:

1) LEIA NV NO

Este comando efetua a leitura dos dados construindo uma matriz com NV colunas e NO linhas.

Os dados devem ser números plotados em formulário próprio, em anexo, e cada observação deve estar contida num campo de dez colunas. As dez primeiras colunas devem ser utilizadas como identificação, sendo lícito o uso de qualquer abreviatura alfa-numérica que identifique o trabalho e a observação respectiva.

Cada linha do formulário comporta somente 7 variáveis. No caso da leitura de mais de 7 será necessário usar uma segunda folha de formulário, obedecendo, entretanto, às mesmas normas, como abaixo explicando:

Primeira folha do formulário

colunas	1	a	10	-	Identificação
colunas	11	a	20	-	Variável 1
colunas	21	a	30	-	Variável 2
.....					
colunas	71	a	80	-	Variável 7

Segunda folha do formulário

colunas	1	a	10	-	Identificação
colunas	11	a	20	-	Variável 8
colunas	21	a	30	-	Variável 9
.....					
colunas	71	a	80	-	Variável 14
					etc. ...

Se o dado for um número inteiro, o algarismo das unidades deverá estar na última coluna do campo correspondente. Se decimal, deve-se reservar uma coluna do formulário para o ponto decimal, posicionando-o sempre na mesma coluna no sentido vertical do formulário.

Após o último dado deverá vir o número nove na primeira coluna, para indicar o fim dos dados.

2) LISTE

Este comando não possui parâmetros. Ele fará com que toda a matriz de dados seja listada.

3) SOME m n v ou SOME m a v

SUBR m n v ou SUBR m a v

MULT m n v ou MULT m a v

DIV m n v ou DIV m a v

onde: m, n, v são números inteiros que se referem aos números das variáveis e a é um número real, positivo.

Estes comandos correspondem a, respectivamente:

$m + n = v$  ou  $m + a = v$

$m - n = v$  ou  $m - a = v$

$m \cdot n = v$  ou  $m \cdot a = v$

$m \div n = v$  ou  $m \div a = v$

No primeiro caso, os comandos operam sobre todas as observações das variáveis (m) e (n), pondo o resultado na variável v. No segundo caso todas as observações das variáveis m são operadas com o número real a e o resultado é posto em v.

NOTAS: 1) Se v for uma variável que não exista na matriz originalmente lida, ela é criada por esse comando, e a matriz passa a se estender até essa variável.

2) Os valores de m, n ou v podem ser iguais entre si.

4) EXP m v ou EXP m a v

LOG m v

onde: m, v são números de variáveis (números inteiros)

a é um número decimal, positivo.

Êstes comandos têm as seguintes representações matemáticas:

<u>EXP</u> <u>m</u> <u>v</u>	corresponde a	$e^m = v$
<u>EXP</u> <u>m</u> <u>a</u> <u>v</u>	corresponde a	$m^a = v$
<u>LOG</u> <u>m</u> <u>v</u>	corresponde a	$\log_e m = v$

No primeiro comando temos  $e^m = v$ , ou seja, tôdas as observações da variável m são expoentes de e e o resultado é pôsto em v. Ou então temos  $m^a = v$ , ou seja, tôdas as observações de m são elevadas ao número decimal a e o resultado colocado em v.

No comando LOG, o logaritmo neperiano de tôdas as observações de m é pôsto em v.

OBSERVAÇÃO:

As notas do item III são aplicáveis também a êsses comandos.

5) RET m n

onde m, n são números das variáveis (números inteiros).

Êsse comando retarda tôdas as observações da variável m e as coloca a partir da segunda observação da variável n. A primeira observação da variável n é feita igual a zero. Assim, por exemplo:

RETARDE A VAR 2 E PONHA NA VAR 3 faz com que a observação nº i da variável 2 seja posta na observação  $i + 1$  da variável 3.

OBSERVAÇÃO:

As notas do item III são aplicáveis também a êsse comando.

6) GERE I COM VALÔRES R1 R2

onde I é o número da variável a ser gerada;

R1 o valor da primeira observação (número real)

R2 o valor do incremento (número real).

Este comando permite que se construa uma variável com os valores de  $R_1$ ,  $R_1 + R_2$ ,  $21 + 22$ , ... nas observações número 1, 2, 3, ... etc., até a observação NO (número de observações da matriz lida).

7) CORRELAÇÃO NV NO  $n_1, n_2, n_3, \dots$

onde NV é o número de variáveis que devem entrar na correlação;

NO é o número de observações que devem ser incluídas na correlação; e

$n_1, n_2, n_3, \dots$  são os números das variáveis que entram no cálculo da matriz de correlação.

O valor de NO indica que se deve utilizar todas as variáveis com as últimas NO observações.

Esse comando faz com que sejam calculados e impressos a matriz de correlação, média, desvio-padrão e coeficiente de variação de cada variável.

NOTAS - 1) São permitidas no máximo 18 variáveis na correlação.

2) É necessário calcular sempre a matriz de correlação antes de mandar calcular a regressão.

3) Se a correlação visa a construir a matriz para o cálculo da regressão em seguida, é necessário que as variáveis estejam na ordem crescente a partir da variável 1, sem saltar nenhuma.

8) REGRESSÃO NV,  $I_1, I_2, \dots$  D, F

onde NV é o número de variáveis independentes + 1;

$I_1, I_2, \dots$  são os números das variáveis independentes

D é o número da variável dependente; e

F é o número da variável cujos elementos serão substituídos pelos valores ajustados da variável D.

Para exemplificar o uso, vejamos dois exemplos:

CORRELAÇÃO 4 VARIÁVEIS 18 OBSERVAÇÕES 1, 2, 3, 4  
REGRESSÃO 3 VARIÁVEIS INDEF 1,4 DEP 3 ESTIMADA 5.

Esses dois comandos farão com que seja calculada a regressão da variável 3 em função de 1 e 4 e os valores ajustados pela regressão ficarão na variável 5.

O comando de regressão faz com que sejam calculados e impressos os valores dos coeficientes da regressão, os valores observados e ajustados da variável dependente, e os desvios respectivos, bem como o coeficiente de correlação múltipla  $R^2$ , o valor  $F$  da regressão, erro-padrão da regressão, erro-padrão de cada coeficiente e o valor de t de Student para cada coeficiente e a estatística de Von Neumann dos resíduos obtidos (veja Handbook of Statistical Tables de D. B. Owen).

#### 9) GRÁFICO NV, I, D1, D2, ... COM M LINHAS

onde NV é o número de variáveis independentes e dependentes envolvidos no gráfico.

I é o número da variável independente.

D1, D2, ... são os números das variáveis dependentes do gráfico

M é o número de linhas que o gráfico deve tomar e deve ser maior ou igual ao número de linhas da matriz de dados.

Este comando faz com que sejam "plotadas" variáveis em função de uma só variável independente, pela impressora do computador.

A variável D1 será impressa com o número 1, D2 com o número 2, etc.

A seguir apresentamos um exemplo da saída da impressora do computador, onde se deve notar que os comandos estão precedidos de dois asteriscos (\*\*). As informações precedidas de um asterisco (\*) não são processadas, servindo, apenas, para orientação do técnico usuário.



\* PLAN GERAL - INVESTIMENTO PRIVADO EXCLUSIVO CIA.S. MISTAS - PERIODO 1948/54 - FGV

\* VARIAVEL 1 - INVESTIMENTO PRIVADO EXCLUSIVO CIA.S MISTAS

\* 2 - FLUXO DE CREDITO AO SETOR PRIVADO EM CR BILHOES DE 1949

\* 3 - LUCRO TRIBUTADO EM CR\$ BILHOES DE 49

\* 4 - TAXA DE INFLACAO EM PERCENTAGENS

\* 5 - RELACAO DE TROCAS

\*\*LEIA 5 VARIAVEIS COM 17 OBSERVACOES

\*\*SOME A VARIAVEL 2 COM A 3 E O RESULTADO SERA A 6

\*\*GERE A VARIAVEL TEMPO 7 A PARTIR DE 1948. COM ACRESCIMO DE 1.

\*\*CORRELACAO DE 7 VARIAVEIS COM 17 OBSERVACOES A SABER 1 2 3 4 5 6 7

NO. DE VAR.	7	NO. DE OBS.	17		
VARIAVEL	MEDIA		DESV. PADRAO		COEF. DE VAR.
1	0.3695880E 02	02	0.7639065E 01		0.2066914
2	0.3270584E 01	01	0.9817613E 01		3.0017919
3	0.3449998E 02	02	0.8290672E 01		0.2403094
4	0.2693528E 02	02	0.2268554E 02		0.8422241
5	0.2130586E 02	02	0.4510178E 01		0.2116871
6	0.3777056E 02	02	0.9785623E 01		0.2590807
7	0.1956000E 04	04	0.4898980E 01		0.0025045

#### MATRIZ DE CORRELACAO

	1	2	3	4	5	6	7
1	1.0000						
2	-0.4033	1.0000					
3	0.8210	-0.4260	1.0000				
4	0.5607	-0.5064	0.7189	1.0000			
5	0.7580	-0.4193	0.7577	0.4185	1.0000		
6	0.2909	0.6422	0.4197	0.1010	0.2212	1.0000	
7	0.7692	-0.4826	0.8547	0.8138	0.7566	0.2400	1.0000

REGRESSAO MULTIPLA

VAR.INDEPENDENTES- 2, 3, 5,

VAR.DEPENDENTE- 1

RQUADRADO= 0.7180953E 00 ERRO PADRAO EST.= 0.4638144E 01 VALOR DE F= 0.1103828E 02

COEFICIENTES DA REGRESSAO	ERRO PADRAO	VALORDE T
B( 1, 1)= 0.7565419E 01		
B( 1, 2)= -0.2214375E-01	0.1283810E 00	-0.1724846E 00
B( 1, 3)= 0.5269886E 00	0.2114899E 00	0.2491790E 01
B( 1, 5)= 0.5296530E 00	0.3874238E 00	0.1367115E 01

OBS.	VERDADEIRO	ESTIMADO	DESVIO
1	0.2180000E 02	0.2113381E 02	0.6661912E 00
2	0.2200000E 02	0.2322528E 02	-0.1225284E 01
3	0.2490000E 02	0.3047446E 02	-0.5574467E 01
4	0.3880000E 02	0.3720612E 02	0.1593879E 01
5	0.4050000E 02	0.3269144E 02	0.7808564E 01
6	0.3220000E 02	0.3710234E 02	-0.4902344E 01
7	0.4040000E 02	0.3856571E 02	0.1834289E 01
8	0.3710000E 02	0.3708440E 02	0.1560211E-01
9	0.4040000E 02	0.3847457E 02	0.1925430E 01
10	0.3590000E 02	0.3745839E 02	-0.1558395E 01
11	0.3370000E 02	0.4144058E 02	-0.7740586E 01
12	0.4200000E 02	0.4130174E 02	0.6982651E 00
13	0.4280000E 02	0.4036340E 02	0.2436600E 01
14	0.4710000E 02	0.4313608E 02	0.3963921E 01
15	0.3960000E 02	0.4046256E 02	-0.8625565E 00
16	0.4870000E 02	0.4247882E 02	0.6221184E 01
17	0.4040000E 02	0.4570016E 02	-0.5300164E 01



**Firma Freire (Grupo de Processamento de Dados)**

**Para :** Demais setores do IPEA, em particular Educação, Planejamento Geral e Demografia.

**Assunto:** Processamento de uma amostra do Censo Demográfico de 1960

ABRIL 1969

1 - Existem atualmente no IPEA 29 caixas de cartões, totalizando aproximadamente 56.000 cartões, representando uma amostra do Censo Demográfico de 1960.

As amostras são referentes às regiões : Nordeste, Sul, e Leste, sendo o número de caixas por região igual a :

Nordeste : 8 caixas

Sul : 10 caixas

Leste : 11 caixas

A amostra de cada região foi feita tendo por base o grau de instrução, concluído com aprovação, do chefe de família. Dependendo do grau de instrução, a seleção foi feita segundo o critério dado pela tabela abaixo :

Tabela 1

GRUPO	Grau de Instrução	Uma família de N em N Chefes		
		Leste	Sul	Nordeste
0 e 1	s/ A alfabetização	25	25	20
2	Primário	25	30	15
3	1º Ciclo	6	7	2
4	2º Ciclo	3	2	1
5	Superior	1	1	1

Por exemplo, para região Sul, foi escolhida uma família de 15 em 15, cujo grau de instrução do chefe pertence ao Grupo 5, isto é, possui instrução primária.

Em cada grupo constituído como descrito, as famílias recebem uma ordenação numérica crescente. Assim, todos os membros de uma mesma família, pertencentes a um dos 5 grupos, possuem o mesmo número de identificação.

## 2 - Modificações feitas :

Os cartões das regiões Nordeste e Sul estão dispostos como descrito em 1, porém, algumas modificações foram feitas nos cartões referentes à amostra da região Leste :

2.1 - A amostra foi subdividida em 10 subamostras.

Subamostra  $n$  é constituída por todos os cartões cujo último dígito de enumeração é  $n$ ;  $n=0, 1, \dots, 9$ .

Por exemplo: subamostra 5 contém todos os cartões cujo último dígito de enumeração é 5.

2.2 - Foram excluídos da amostra, cartões referentes a:

2.2.1 - Pessoas não residentes

2.2.2 - Convidados e empregados

2.2.3 - Pessoas que responderam que não sabiam a resposta, e cartões com códigos impossíveis.

Esta última separação foi feita apenas nas colunas 1, 5, 6, 15, 16, 20, 21, 24, 31, 32, 36, e 40. Os cartões excluídos estão na caixa 11 desta região.

2.3 - As subamostras 2, 3 e 4 mantêm o agrupamento familiar; as demais estão misturadas.

## 3 - Gravação dos cartões da amostra em fita magnética.

Os cartões representantes da amostra do Censo foram gravados em fita magnética (IPMA 10), a qual se encontra no Rio Data-Centro da Pontifícia Universidade Católica do Rio de Janeiro.

O número de observações por região é dado na tabela 2 e o número de observações por subamostra da região Leste na tabela 3.

## 4 - Informações Disponíveis

Vide Apêndice

## 5 - Programa de Processamento

Em fasc final de implantação. Para maiores detalhes, procurar responsável pelo projeto.

Tabela 2

Região	Nº Obs.
Nordeste	15979
Sul	19662
Leste	19211

Tabela 3

Região Leste	
Sub-regiões	Nº Obs.
1	1873
2	1877
3	2011
4	1903
5	1932
6	1951
7	1942
8	1949
9	1924
10	1949

## APÊNDICE

Cada cartão, da amostra do Censo Demográfico de 1960, diz respeito a 1(un) indivíduo e fornece as seguintes informações a respeito do mesmo :

### a) Características Individuais

<u>Nº</u>	<u>Variável</u>
1	Situação (urbana, suburbana, rural)
2	Estado
3	Sexo e condição presente (morador, visitante, empregado ...)
4	Parentesco
5	Idade
6	Religião
7	Côr
8	Naturalidade
9	Nacionalidade
10	Procedência ( de onde veio)
11	Tempo de residência
12	Residência anterior
13	Alfabetização
14	Última série concluída
15	Grau de conhecimento
16	Espécie de estudo
17	Est. do conjugal
18	Ano do casamento
19	Filhos tidos
20	Filhos vivos
21	Rendimento
22	Atividade não econômica
23	Ocupação habitual
24	Última semana de ocupação
25	Ramo e classe de atividade
26	Posição na ocupação

### b) Características Domiliares

<u>Nº</u>	<u>Variável</u>
1	Espécie (particular, coletivo ...)
2	Tipo (durável, rústico ...)
3	Condição de ocupação
4	Aluguel mensal
5	Água

NR

Variável

6	Instalação Sanitária
7	Fogão
8	Iluminação
9	Rádio (tem, não tem)
10	Geladeira (tem, não tem)
11	Televisão (tem, não tem)
12	Total de cômodos
13	Total de dormitórios