

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA  
Departamento de Métodos Quantitativos

O PROCEDIMENTO CLMEANS

Moacir Pedroso Júnior

DMQ/A/34 - Janeiro, 1979

O procedimento CLMEANS.  
1979 FL - FOL.0810



10818 - 1

## O PROCEDIMENTO CLMEANS

Moacir Pedroso Júnior\*

### 1. INTRODUÇÃO.

O procedimento CLMEANS é um programa escrito em linguagem PL/I (F) e implementado como procedimento do "Statistical Analysis System" (Barr et alli (1)), através do qual ele pode ser acessado. Este procedimento procura particionar um conjunto de médias de tratamentos em subconjuntos homogêneos, de acordo com o algoritmo descrito por Scott e Knott (2).

Brevemente, o algoritmo procede do seguinte modo: dado um conjunto de  $k$  médias de tratamentos,  $y_1, y_2, \dots, y_k$ , com  $y_i \sim N(\mu_i, \sigma^2)$ , e uma estimativa independente,  $s^2$ , da variância comum, nos podemos descobrir se o conjunto pode ser particionado em, pelo menos, dois grupos homogêneos pela aplicação do teste  $F$ , da maneira usual. Podemos então usar um teste da razão de verossimilhança para testar essa hipótese. Como foi desenvolvido na referência dada acima, nós consideraremos uma estatística  $\lambda$ , cuja porcentagem de pontos sob a hipótese nula (nenhuma diferença entre as médias) pode ser aproximada por aquela de uma distribuição  $\chi^2$  com  $\mu = k / (\pi - 2)$  graus de liberdade. Então, se nós aceitamos a partição do grupo original, os subconjuntos resultantes poderão ser particionados de um modo semelhante, hierarquicamente.

### 2. RESULTADOS.

Para cada execução do procedimento CLMEANS, os seguintes resultados serão impressos:

- uma listagem das médias originais em ordem crescente;
- um mapa dos grupos formados, semelhantes a um dendograma em análise de conglomerados ("cluster analysis");

---

\* Pesquisador do Departamento de Métodos Quantitativos da Empresa Brasileira de Pesquisa Agropecuária, Brasília, DF.

- um sumário estatístico, compreende o valor de  $\lambda$ , os graus de liberdade e a probabilidade de um valor maior que uma variável  $\chi^2$ , para cada partição.

### 3. USO.

#### O COMANDO DO PROCEDIMENTO CLMEANS:

```
PROC CLMEANS opções-e-parâmetros;
```

As opções e parâmetros que podem aparecer no comando PROC CLMEANS estão listadas abaixo;

```
N=___  
NUMBER=___
```

Este parâmetro deve sempre estar presente. Ele especifica o número de médias existentes no grupo a ser analisado.

```
VAR=___  
VARMEAN=___
```

Este parâmetro deve sempre estar presente. Ele especifica o valor de uma estimativa independente da variância comum das médias. (Em muitos casos uma estimativa independente da variância comum das médias pode ser obtida multiplicando-se o quadrado médio do resíduo de uma análise de variância pelo número de repetições)

```
DF=___
```

Este parâmetro deve sempre estar presente. Ele especifica o número de graus de liberdade da estimativa de variância comum.

```
L=.___  
LEVEL=.___
```

Este parâmetro especifica o nível a ser usado no julgamento da significância de uma partição. Quando este parâmetro é omitido ele é assumido como sendo igual a 0.05.

## NAMED

Quando esta opção é omitida, o procedimento atribuirá os nomes "MEAN#001", "MEAN#002",... às médias, na ordem em que estas aparecem. Quando esta opção está presente, o procedimento atribuirá às médias os nomes dados pelo usuário.

O COMANDO PARMCARDS:

```
PARMCARDS;
```

Este comando indica que as médias e os nomes (quando a opção NAMED foi usada) são dados abaixo.

As médias podem ser dadas em qualquer ordem, devendo estar separada uma das outras por brancos e/ou uma vírgula. Quando a opção NAMED foi usada, os nomes devem ser dados após todas as médias em ordem correspondente. Cada nome, que consiste de até oito caracteres envolvidos por apóstrofes, deve estar separado dos outros por brancos e/ou uma vírgula.

## 4. REFERÊNCIAS.

1. Barr, A.J. et alli (1976), "A User's Guide to SAS76", SAS Institute, Inc., Raleigh, N.C.
2. Scott, A.J. e M. Knott (1974), "A cluster analysis method for grouping means in the analysis of variance". Biometrics 30, pp. 507-512.

## 5. EXEMPLO.

NOTE: THE JOB PAKAA91X HAS BEEN RUN UNDER RELEASE 76.60 OF SAS AT EMBRAPA/JICA.

```
1      PROC CLMEANS N=15 VAR=16.3 DF=49 NAMED;  
2      PARMCARDS;  
3      90.4 70.8 158.6 25.9 103.2 31.2 120.6 50.03 88.1 43.2 65.7 110.0 135.  
4      180.4 160.7  
5      'UFRS' 'UFSC' 'PUCRJ' 'UFRJ' 'USP' 'UNB' 'UFMG' 'UFMT' 'UFPB' 'UFPE'  
6      'PUCDF' 'PUCSP' 'UEPR' 'UFPE' 'UFPA'  
7      ;
```

NOTE: THE PROCEDURE CLMEANS USED 2.91 SECONDS AND 448K AND PRINTED PAGES 1 TO 3

NOTE: BARR, GOODNIGHT, SALL AND HELWIG  
SAS INSTITUTE INC.  
P.O. BOX 10066  
RALPHIGH, N.C. 27605



STATISTICAL SUMMARY

| PARTITION OF THE GROUP | LAMBDA | OF | PRCB > LAMBDA |
|------------------------|--------|----|---------------|
| UFRJ -UFPE             | 62.099 | 14 | 0.000100      |
| UFRJ -UFRS             | 53.077 | 8  | 0.000100      |
| USP -UFPE              | 54.247 | 7  | 0.000100      |
| UFRJ -UFMT             | 20.474 | 4  | 0.000403      |
| PUCDF -UFRS            | 25.619 | 4  | 0.000100      |
| USP -UFPR              | 23.939 | 4  | 0.000100      |
| PUCRJ -UFPE            | 20.103 | 3  | 0.000182      |
| UFMG -UFPR             | 6.262  | 2  | 0.016089      |

\*\*\* PROCEDURE CLMEANS \*\*\*

REFERENCE: SCOTT, A. J. AND M. KNOTT (1974),  
 'A CLUSTER ANALYSIS METHOD FOR  
 GROUPING MEANS IN THE ANALYSIS  
 OF VARIANCE'. BIOMETRICS, 30.

PROGRAMMER: MOACIR PEDROSO JUNIOR  
 DMO/EMBRAPA  
 SRTV. RD. VERANCIO 2000, S/722  
 CEP 70333  
 BRASILIA, DF  
 BRASIL

\*\*\*