

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA - DEPTO. DE ESTATÍSTICA.

TRABALHO DE CONCLUSÃO
DO BACHARELADO EM ESTATÍSTICA

UMA TÉCNICA PARA AGRUPAR
TRATAMENTOS NA ANÁLISE DE VARIÂNCIA

MARIA CRISTINA PEREIRA PIMENTEL

PROFESSORA ORIENTADORA: ELSA C. DE MUNDSTOCK

Porto Alegre, novembro de 1987.

AGRADECIMENTOS

- À Prof.^a Elsa C. de Mundstock pela excelente orientação recebida.
- Ao pessoal do Centro de Processamento de Dados da U.F.R.G.S., pela colaboração e recursos fornecidos para a realização deste trabalho.

SUMÁRIO

1. INTRODUÇÃO	7
2. METODOLOGIA	9
3. UM MÉTODO DE ANÁLISE DE CLUSTER	11
3.1 - CONSIDERAÇÕES GERAIS	11
3.2 - MÉTODO DE EDWARDS E CAVALLI-SFORZA	11
3.3 - UM CRITÉRIO PARA A DIVISÃO DE UM GRUPO	12
3.4 - UMA FORMA OPERACIONAL DE CALCULAR AS SOMAS DE QUADRADOS	14
3.5 - QUANDO INTERROMPER O PROCESSO DE DIVISÃO	15
4- AGRUPAMENTO DE TRATAMENTOS NA ANÁLISE DE VARIÂNCIA ..	16
4.1 - DEFINIÇÕES	16
4.2 - OBTENÇÃO DE DOIS GRUPOS DE TRATAMENTOS	18
4.3 - OBTENÇÃO DE TRÊS OU MAIS GRUPOS DE TRATAMENTOS.	19
5. EXEMPLO DE APLICAÇÃO	21
5.1 - DELINEAMENTO EXPERIMENTAL	21
5.2 - ANÁLISE DE VARIÂNCIA	23
5.3 - OBTENÇÃO DOS GRUPOS	23
5.3.1 - Estimativa da Variância das Médias	24
5.3.2 - Matriz das Distâncias entre as Médias .	24
5.3.3 - Matriz dos Quadrados das Distâncias en- tre as Médias	26

5.3.4 - Partição do Grupo 1 2 3 4 5 6	26
5.3.5 - Partição do Grupo 1 2 3 4 5	28
5.3.6 - Partição do Grupo 1 2 3	29
5.3.7 - Partição do Grupo 4 5	30
5.3.8 - Grupos Resultantes	31
6. PROGRAMA DE COMPUTADOR	33
6.1 - UTILIZAÇÃO DO PROGRAMA	33
6.2 - ARQUIVO DE DADOS	34
6.3 - OBSERVAÇÕES	36
6.4 - EXEMPLO DE USO DO PROGRAMA	36
7. CONCLUSÃO	46
BIBLIIGRAFIA	47
APÊNDICES	48

LISTA DE APÊNDICES

- APÊNDICE A: LISTAGENS DO PROGRAMA ANOVAR	49
- APÊNDICE B: LISTAGENS DO PROGRAMA FONTE	52
- APÊNDICE C: ALGORITMO DAS PARTIÇÕES	65
- APÊNDICE D: FLUXOGRAMA	69

LISTA DE TABELAS

- TABELA I:	OBSERVAÇÕES SIMULADAS DOS TRATAMENTOS ...	22
- TABELA II:	TABELA ANOVA DO EXPERIMENTO	23
- TABELA III:	PARTIÇÕES DO GRUPO 1 2 3 4 5 6	27
- TABELA IV:	PARTIÇÕES DO GRUPO 1 2 3 4 5	29
- TABELA V:	PARTIÇÕES DO GRUPO 1 2 3	30
- TABELA VI:	PARTIÇÕES DO GRUPO 4 5	31
- TABELA VII:	GRUPOS RESULTANTES	32

1 - INTRODUÇÃO

Uma pesquisa científica muitas vezes envolve a comparação de tratamentos para verificar se eles apresentam ou não um efeito diferenciado sobre uma variável de resposta. Uma das técnicas estatísticas adequadas a este propósito, desde que verificadas certas condições, é a análise de variância. Nesta técnica, a variabilidade total da resposta é decomposta em componentes associadas a fontes distintas de variação, definidas no delineamento do experimento. A partir da comparação destas fontes é que testamos a homogeneidade dos tratamentos.

Entretanto, para que a análise da variância possa ser empregada adequadamente, é importante que a variável de resposta tenha distribuição normal e que a sua variância seja a mesma para todos os tratamentos (homoscedasticidade), embora as médias possam diferir entre si. Se isto não se verificar podemos recorrer a uma transformação da variável original de tal forma que a nova variável obedeça a estas condições.

Uma das limitações da análise de variância é que ela nos oferece uma visão global do experimento. Ou seja, é verificado se as amostras colhidas apresentam evidências estatísticas de que existem tratamentos que diferem entre si. Porém, se tais diferenças existirem, não nos é informado onde elas estão localizadas. Assim sendo, devemos recorrer a outras técnicas estatísticas que nos permitam dar continuidade à análise.

As técnicas de comparações múltiplas são muito úteis quando estamos apenas interessados em localizar as diferenças.

Mas, se desejarmos obter grupos de tratamentos similares, elas já não serão tão adequadas, pois os grupos formados nem sempre são disjuntos, o que dificulta a interpretação dos resultados.

Em um artigo publicado na Revista Biometrics (1974) A.J.Scott e M.Knott sugerem uma técnica para dividir os tratamentos em grupos homogêneos e disjuntos com base no método de Cluster, sugerido por Edwards e Cavalli-Sforza (1965). No presente estudo, procuraremos expor brevemente tal técnica e apresentar um programa de computador para a sua aplicação.

2 - METODOLOGIA

Com a finalidade de tornar viável a aplicação do método de agrupar tratamentos apresentado neste estudo, desenvolvemos um programa na linguagem FORTRAN IV compilado no computador A^{9P} da linha Burroughs com versão de software 3.6. Procuramos fazer a lógica o mais estruturada possível, de forma a tornar fácil a compreensão do programa.

Como exemplo de aplicação, simulou-se um experimento com um fator fixo de seis níveis. Estes níveis foram simulados de tal forma que pertençam a populações com médias distintas, ou seja, três delas com média 15, duas com média 27, uma com média 49 e todas com variância 25, configurando assim, um delineamento completamente casualizado com seis tratamentos.

Para cada tratamento, gerou-se dez observações independentes de acordo com a distribuição normal correspondente. Para tanto, utilizou-se a rotina RANDOM NUMBER GENERATOR da calculadora programável TEXAS TI-58C. Nesta rotina, primeiro é gerado um par de números aleatórios uniformemente distribuí-

dos no intervalo (0 ; 1). Então, usando esses números (μ_1 e μ_2) um valor $X \sim N(0 ; 1)$ é calculado através da expressão:

$$X = \sqrt{-2 \ln \mu_1} \cdot \cos(2 \pi \mu_2).$$

Finalmente, é feito um ajuste para a média μ e desvio padrão σ informados no início do uso da rotina: $X' = \mu + \sigma X$.

O desenvolvimento das fórmulas e do algoritmo desta rotina encontram-se na seguinte referência bibliográfica:

KNUTH, Donald E. *The Art of Computer Programming*.
USA, Addison-Wesley Publishing Co., 1969.

Uma vez obtidas as amostras, aplicou-se a análise de variância e o agrupamento de tratamentos de forma a tornar clara a técnica de Scott e Knott, bem como o uso do programa.

3 - UM MÉTODO DE ANÁLISE DE CLUSTER

3.1 - CONSIDERAÇÕES GERAIS

A análise de Cluster é uma técnica que nos permite classificar indivíduos não previamente classificados em grupos que sejam homogêneos internamente e heterogêneos entre si. Tal classificação pode ser baseada na descrição dos indivíduos por um conjunto de medidas (caso multivariado) ou por apenas uma medida (caso univariado). Neste estudo, ficaremos restritos ao caso univariado.

3.2 - MÉTODO DE EDWARDS E CAVALLI-SFORZA

O método apresentado a seguir, é um método hierárquico divisivo, sugerido por Edwards e Cavalli-Sforza no qual, inicialmente consideramos todos os indivíduos como pertencentes a um único grupo. Então, dividimos este grupo em dois. Cada um dos novos grupos será dividido em outros dois grupos, e assim sucessivamente até que cada grupo contenha um único indi-

víduo. O resultado final pode ser representado através de um diagrama de árvore. Por exemplo, se temos cinco indivíduos, um dos possíveis resultados seria o seguinte:

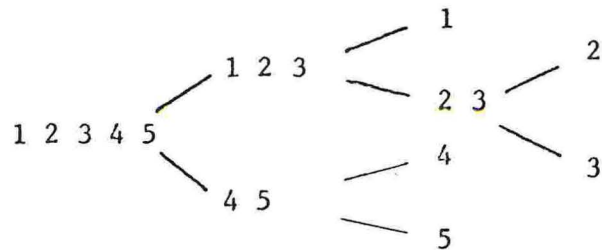


FIGURA 1 - DIAGRAMA DE ÁRVORE

O problema principal resume-se, portanto, em obter um critério para se dividir um grupo em dois grupos o mais distintos possível entre si. E, através da aplicação sucessiva deste critério, obteremos a análise completa.

3.3 - UM CRITÉRIO PARA A DIVISÃO DE UM GRUPO

Quando os indivíduos estão classificados em dois grupos, sabemos da análise de variância, que a soma dos quadrados das distâncias de cada indivíduo à média geral (SQT) pode ser decomposta em três parcelas:

- SQW1: soma dos quadrados das distâncias de cada indivíduo do primeiro grupo à média deste grupo (soma de quadrados dentro do primeiro grupo);
- SQW2: soma dos quadrados das distâncias de cada indivíduo do segundo grupo à média deste grupo (soma

de quadrados dentro do segundo grupo);

- SQB: soma dos quadrados das distâncias da média de cada grupo à média geral, ponderados pelo tamanho do grupo (soma de quadrados entre os grupos).

Se representarmos a soma de SQW1 e SQW2 por SQW (soma de quadrados dentro dos grupos), podemos escrever:

$$SQT = SQW + SQB$$

Um critério natural para obtermos os dois grupos mais distintos entre si, é escolhermos a partição que maximiza a soma de quadrados entre grupos (SQB), cujo valor será também, uma medida de "importância" da divisão. Os grupos obtidos, continuarão a ser divididos até o final do processo quando cada grupo conterà apenas um indivíduo, a soma de quadrados dentro de cada grupo será nula e o somatório das SQB's associadas a cada partição feita será igual a SQT inicial.

A variância de um grupo pode ser tomada como uma medida da sua homogeneidade interna e é calculada dividindo-se a sua SQT pelo número de indivíduos que o compõem (divide-se por n e não por $n-1$ porque o objetivo principal não é a estimação). Devemos observar que a SQW1 e SQW2 serão as SQT's dos respectivos grupos na próxima divisão.

Do exposto, conclui-se que em cada divisão teremos como objetivo maximizar a SQB e conseqüentemente, minimizar a SQW.

3.4 - UMA FORMA OPERACIONAL DE CALCULAR AS SOMAS DE QUADRADOS

Pode-se demonstrar que a variância de n pontos é igual a soma de quadrados das distâncias entre cada dois pontos (sendo cada distância tomada uma única vez) dividida pelo quadrado de n . Portanto, se dividirmos esse somatório apenas por n , obteremos a SQT. Assim, torna-se extremamente simples calcular as SQT's e as SQW's a partir da matriz simétrica dos quadrados das distâncias entre os indivíduos.

Quando vamos dividir os n indivíduos de um grupo em dois outros grupos, cada indivíduo poderá pertencer tanto ao grupo um quanto ao grupo dois, existindo $2 \times 2 \times \dots \times 2 = 2^n$ formas diferentes de aloca-los. Como nenhum grupo poderá ficar vazio, deveremos retirar estas duas situações:

$$(2^n - 2) = 2(2^{n-1} - 1).$$

Finalmente, como a ordem dos grupos não importa, devemos dividir esse total por dois, obtendo-se $(2^{n-1} - 1)$ partições possíveis dos indivíduos em grupos não vazios e disjuntos.

Porém, como estamos restritos ao caso univariado, se ordenarmos os indivíduos de acordo com a variável que os está descrevendo, basta pesquisarmos as $n - 1$ partições sucessivas. Por exemplo, se temos cinco indivíduos ordenados, basta pesquisarmos as seguintes partições:

1 : 2 3 4 5
1 2 : 3 4 5
1 2 3 : 4 5
1 2 3 4 : 5

3.5 - QUANDO INTERROMPER O PROCESSO DE DIVISÃO

Devemos lembrar que o objetivo da análise de Cluster é obter grupos heterogêneos entre si e não isolar os indivíduos. Portanto, agora que já temos um critério para determinar a melhor partição de um grupo, devemos ter outro critério para decidir quando um grupo não deve mais ser dividido.

No caso específico em que desejamos agrupar tratamentos na análise de variância, Scott e Knott sugerem como critério um teste estatístico de hipóteses que será descrito a seguir.

4 - AGRUPAMENTO DE TRATAMENTOS NA ANÁLISE DE VARIÂNCIA

4.1 - DEFINIÇÕES

Seja um experimento balanceado onde y_{ij} é o valor da variável de resposta da j -ésima observação do i -ésimo tratamento e tal que:

- $i = 1, 2, \dots, k$ e $j = 1, 2, \dots, n$;
- $y_{ij} \sim N(\mu_i ; \sigma^2)$ independentes;
- μ_i pode ou não variar de um tratamento para outro;
- σ^2 é comum a todos os tratamentos;
- y_1, y_2, \dots, y_k são as médias ordenadas dos tratamentos;
- $y_i \sim N(\mu_i ; \sigma^2/n)$ independentes ($i = 1, 2, \dots, k$);
- \bar{y} é a média geral das observações:

$$\bar{y} = \frac{\left(\sum_{i=1}^k \sum_{j=1}^n y_{ij} \right)}{k \cdot n}$$

Na análise de variância, um estimador não viciado para σ^2 é dado pelo quadrado médio residual (QMR) com ν graus de liberdade. Portanto, podemos definir $s^2 = \text{QMR}/n$ como sendo um estimador não viciado da variância de y_i ($i = 1, 2, \dots, k$), tal que

$$\frac{\nu s^2}{(\sigma^2/n)} \sim \chi^2(\nu).$$

Sob a hipótese de que todos os tratamentos possuem a mesma média, o estimador de máxima verossimilhança para a variância das médias amostrais y_i ($i = 1, 2, \dots, k$) é dado por:

$$\hat{\sigma}_0^2 = \frac{1}{n} \left(\frac{\text{SQB} + \nu \text{QMR}}{k + \nu} \right), \text{ onde}$$

$$\text{SQB} = \sum_{i=1}^k n (y_i - \bar{y})^2, \text{ ou ainda,}$$

$$\hat{\sigma}_0^2 = \frac{\sum_{i=1}^k (y_i - \bar{y})^2 + \nu s^2}{k + \nu}.$$

O somatório $\sum_{i=1}^k (y_i - \bar{y})^2$ corresponde a SQT do grupo das k médias e portanto, podemos escrever:

$$\hat{\sigma}_0^2 = \frac{\text{SQT} + \nu s^2}{k + \nu}$$

4.2 - OBTENÇÃO DE DOIS GRUPOS DE TRATAMENTOS

Considerando as definições acima, vamos analisar inicialmente o problema onde suspeitamos que as médias dos tratamentos podem ser divididas em dois grupos distintos e internamente homogêneos. As hipóteses estatísticas neste caso, são:

$$\left\{ \begin{array}{l} H_0: \mu_i = \mu \quad (i = 1, 2, \dots, k) \\ H_1: \mu_i = \alpha_1, \quad \mu_i = \alpha_2 \text{ e } \mu_i = \alpha_1 \text{ ou } \alpha_2 \\ \quad (i = 1, 2, \dots, k) \end{array} \right.$$

onde α_1 e α_2 representam as médias desconhecidas dos dois grupos.

É razoável admitirmos que de todas as partições possíveis das k médias em dois grupos, aquela que atribui maior verossimilhança à amostra dada sob H_1 é a que apresenta uma maior soma de quadrados entre os grupos (SQB).

Vamos denotar o valor máximo de SQB por B_0 , calculado como na técnica de Cluster apresentada anteriormente. Como B_0 é uma medida da diferença dos grupos mais verossímeis sob H_1 , um teste para as hipóteses dadas é equivalente a um teste que rejeita H_0 se $B_0/\hat{\sigma}_0^2$ for suficientemente grande.

Para realizarmos o teste, utilizaremos a estatística modificada

$$\lambda = \frac{\pi}{2(\pi-2)} \cdot \frac{B_0}{\hat{\sigma}_0^2},$$

cuja distribuição nula de probabilidades é assintoticamente igual a uma χ^2 com $\nu_0 = \frac{k}{\pi-2}$ graus de liberdade. Esta estatística é adequada mesmo para k igual a dois.

Portanto, se rejeitarmos H_0 , temos evidências estatísticas de que os k tratamentos podem ser divididos nos dois grupos indicados por B_0 .

4.3 - OBTENÇÃO DE TRÊS OU MAIS GRUPOS DE TRATAMENTOS

Se não for suficiente dividirmos os tratamentos em apenas dois grupos, então aplicamos o mesmo procedimento a cada um dos dois novos grupos e assim sucessivamente até que os grupos resultantes sejam considerados homogêneos pela aplicação do teste da estatística λ .

Em cada novo grupo de l médias devemos recalcular $\hat{\sigma}_0^2$ da seguinte forma:

$$\hat{\sigma}_0^2 = \frac{\sum_{i=1}^l (y_i - \bar{y})^2 + \nu s^2}{l + \nu}, \text{ onde,}$$

- y_1, y_2, \dots, y_l são as médias ordenadas que compõem o novo grupo;

$$- \bar{y} = \frac{\sum_{i=1}^l y_i}{l};$$

- $\sum_{i=1}^{\ell} (y_i - \bar{y})^2$ corresponde a SQT do grupo das ℓ médias.

O número de graus de liberdade da estatística λ será dado por:

$$v_0 = \frac{\ell}{\pi - 2} .$$

5 - EXEMPLO DE APLICAÇÃO

5.1 - DELINEAMENTO EXPERIMENTAL

Para exemplificar o método, analisaremos um experimento balanceado no delineamento completamente casualizado com um fator fixo A, obtido através de um processo de simulação. O experimento contém seis tratamentos, dez observações em cada parcela e o seu modelo matemático é:

$$y_{ij} = \mu + \alpha_i + e_{ij}, \text{ onde}$$

- y_{ij} ($i = 1, 2, \dots, 6$ $j = 1, 2, \dots, 10$) é a j -ésima observação do i -ésimo nível de A;
- $y_{ij} \sim N(\mu_i ; \sigma^2)$ independentes;
- μ é a média geral;
- α_i é o efeito do i -ésimo nível de A;
- e_{ij} é um erro casual $\sim N(0 ; \sigma^2)$.

Na simulação das amostras dos tratamentos, foram utilizadas as seguintes distribuições de probabilidades:

- y_{1j}, y_{2j} e y_{3j} ($j = 1, 2, \dots, 10$) $\sim N(15;25)$;
- y_{4j} e y_{5j} ($j = 1, 2, \dots, 10$) $\sim N(27;25)$;
- y_{6j} ($j = 1, 2, \dots, 10$) $\sim N(49;25)$.

Portanto, após a análise, espera-se obter três grupos de tratamentos como definidos acima.

A seguir, apresentamos os valores obtidos na simulação.

TABELA I - OBSERVAÇÕES SIMULADAS DOS TRATAMENTOS

ESTATÍSTICAS	TRATAMENTOS						TOTAL
	A ₁	A ₂	A ₃	A ₄	A ₅	A ₆	
	16,66	10,04	19,81	23,85	23,92	41,11	
	14,95	6,86	11,39	27,37	25,21	45,57	
	24,14	11,14	20,42	28,72	31,26	52,70	
	23,00	18,90	15,21	30,15	29,73	42,90	
	26,34	12,90	18,87	16,88	31,23	48,76	
	12,03	20,21	19,72	20,08	31,46	51,79	
	11,79	20,19	14,53	30,08	29,97	46,19	
	15,65	14,94	11,68	23,91	31,82	56,75	
	12,46	23,48	17,07	37,69	22,29	42,50	
	11,58	15,22	23,09	28,75	32,49	48,94	
TOTAL AMOSTRAL	168,60	154,18	171,79	267,48	289,38	477,21	1.528,64
MÉDIA AMOSTRAL	16,86	15,42	17,18	26,75	28,94	47,72	25,48
TAMAN. AMOSTRA	10	10	10	10	10	10	60

OBS.: Neste exemplo, todos os cálculos serão feitos com precisão de máquina, porém, os resultados serão transcritos com apenas duas casas decimais (arredondados).

5.2 - ANÁLISE DE VARIÂNCIA

A análise de variância foi feita mediante o uso do software ANOVAR, desenvolvido por Linn Shaw (1973) na Universidade da Califórnia. As listagens do computador encontram-se no Apêndice A.

TABELA II - TABELA ANOVA DO EXPERIMENTO

FONTE DE VARIAÇÃO	g.l.	S.Q.	Q.M.	F _o	SIGNIFICÂNCIA.
Tratamentos	5	7.526,82	505,36	61,27	0,00
Resíduo	54	1.326,72	24,57		
Total.	59	8.853,54			

De acordo com os dados acima, podemos afirmar que, ao nível de significância 0,05 as amostras simuladas oferecem evidências estatísticas de que existem tratamentos que diferem significativamente entre si.

Assim sendo, é adequado aplicarmos a técnica de Scott e Knott para obtermos grupos de tratamentos.

5.3 - OBTENÇÃO DOS GRUPOS

Vamos representar pelos algarismos 1, 2, ..., 6 as médias amostrais ordenadas y_1, y_2, \dots, y_6 :

$$(1) y_1 = \bar{y}_{A2} = 15,42$$

$$(2) y_2 = \bar{y}_{A1} = 16,86$$

$$(3) y_3 = \bar{y}_{A3} = 17,18$$

$$(4) y_4 = \bar{y}_{A4} = 26,75$$

$$(5) y_5 = \bar{y}_{A5} = 28,94$$

$$(6) y_6 = \bar{y}_{A6} = 47,72$$

5.3.1 - Estimativa da Variância das Médias (σ^2/n)

$$s^2 = \frac{\text{QMR}}{n} = \frac{24,57}{10} = 2,46$$

O quadrado médio residual (QMR) possui 54 graus de liberdade, ou seja: $v = 54$.

5.3.2 - Matriz das Distâncias entre as Médias

Seja $D = (d_{ij})$ $i, j = 1, 2, \dots, 6$ a matriz simétrica das distâncias entre as médias, onde $d_{ij} = |y_i - y_j|$. Como nos cálculos das SQT's e SQW's a distância entre duas médias é tomada uma única vez, colocaremos zeros nos elementos abaixo da diagonal principal.

$$d_{12} = |15,42 - 16,86| = 1,44$$

$$d_{13} = |15,42 - 17,18| = 1,76$$

$$d_{14} = |15,42 - 26,75| = 11,33$$

$$d_{15} = |15,42 - 28,94| = 13,92$$

$$d_{16} = |15,42 - 47,72| = 32,30$$

$$d_{23} = |16,86 - 17,18| = 0,32$$

$$d_{24} = |16,86 - 26,75| = 9,89$$

$$d_{25} = |16,86 - 28,94| = 12,08$$

$$d_{26} = |16,86 - 47,72| = 30,86$$

$$d_{34} = |17,18 - 26,75| = 9,57$$

$$d_{35} = |17,18 - 28,94| = 11,76$$

$$d_{36} = |17,18 - 47,72| = 30,54$$

$$d_{45} = |26,75 - 28,94| = 2,19$$

$$d_{46} = |26,75 - 47,72| = 20,97$$

$$d_{56} = |28,94 - 47,72| = 18,78$$

$$D = \begin{bmatrix} 0,00 & 1,44 & 1,76 & 11,33 & 13,92 & 32,30 \\ 0,00 & 0,00 & 0,32 & 9,89 & 12,08 & 30,86 \\ 0,00 & 0,00 & 0,00 & 9,57 & 11,76 & 30,54 \\ 0,00 & 0,00 & 0,00 & 0,00 & 2,19 & 20,97 \\ 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 18,78 \\ 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 0,00 \end{bmatrix}$$

5.3.3 - Matriz dos Quadrados das distâncias entre as Médias

Esta matriz é obtida elevando-se ao quadrado os elementos da anterior:

$$Q = (q_{ij}) = (d_{ij}^2) \quad i, j = 1, 2, \dots, 6.$$

$$Q = \begin{bmatrix} 0,00 & 2,07 & 3,10 & 128,37 & 182,79 & 1\ 043,29 \\ 0,00 & 0,00 & 0,10 & 97,81 & 145,93 & 952,34 \\ 0,00 & 0,00 & 0,00 & 91,58 & 138,30 & 932,69 \\ 0,00 & 0,00 & 0,00 & 0,00 & 4,80 & 439,74 \\ 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 352,69 \\ 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 0,00 \end{bmatrix}$$

5.3.4 - Partição do Grupo 1 2 3 4 5 6

Representaremos por k o número de médias do grupo a ser partido e por k_1 e k_2 , o dos seus respectivos subgrupos.

$$\begin{aligned} SQT &= \frac{\sum_{i=1}^5 \sum_{j>i}^6 q_{ij}}{k} = \frac{(2,07 + 3,10 + \dots + 352,69)}{6} \\ &= 752,60 \end{aligned}$$

$$\hat{\sigma}_0^2 = \frac{SQT + v s^2}{k + v} = \frac{752,60 + 54 \times 2,46}{6 + 54} = 14,75$$

Como primeira partição temos 1 : 2 3 4 5 6
 ($k_1 = 1$ e $k_2 = 5$) e, sendo o subgrupo 1 unitário, temos que
 $SQW_1 = 0$.

$$SQW_2 = \frac{\sum_{i=2}^5 \sum_{j>i}^6 q_{ij}}{k_2} = \frac{(0,10 + 97,81 + \dots + 352,69)}{5}$$

$$= 631,20$$

$$SQW = SQW_1 + SQW_2 = 0 + 631,20 = 631,20$$

$$SQB = SQT - SQW = 752,60 - 631,20 = 121,40$$

Estes cálculos são feitos para todas as partições possíveis, obtendo-se os seguintes resultados:

— TABELA III - PARTIÇÕES DO GRUPO 1 2 3 4 5 6

PARTIÇÃO	SQW1	SQW2	SQW	SQB
1 : 2 3 4 5 6	0	631,20	631,20	121,40
1 2 : 3 4 5 6	1,04	489,95	490,99	261,61
1 2 3 : 4 5 6	1,76	265,74	267,50	485,10
1 2 3 4 : 5 6	80,76	176,34	257,10	495,50
1 3 3 4 5 : 6	158,97	0	158,97	593,63

O valor máximo da SQB (B_0) é igual a 593,63 para a partição 1 2 3 4 5 : 6.

$$\lambda = \frac{\pi}{2(\pi - 2)} \cdot \frac{B_0}{\hat{\sigma}_0^2} = \frac{\pi}{2(\pi - 2)} \cdot \frac{593,63}{14,75} = 55,36$$

$$v_0 = \frac{k}{\pi - 2} = \frac{6}{\pi - 2} = 5,26 \text{ (g.l.)}$$

Sob a hipótese nula de indivisibilidade do grupo, $\lambda \sim \chi^2(5)$ e a $P_0(\lambda \geq 55,36) = 0,00$. Portanto, ao nível de significância 0,05, os grupos de médias 1 2 3 4 5 e 6 diferem significativamente entre si.

Como o segundo grupo obtido é unitário, não poderá mais ser dividido e continuaremos a análise apenas com o primeiro.

5.3.5 - Partição do Grupo 1 2 3 4 5

$$SQT = \frac{\sum_{i=1}^4 \sum_{j>i}^5 q_{ij}}{k} = \frac{(2,07 + 3,10 + \dots + 4,80)}{5}$$

$$= 158,97$$

$$\hat{\sigma}_0^2 = \frac{SQT + v s^2}{k + v} = \frac{158,57 + 54 \times 2,46}{5 + 54} = 4,94$$

TABELA IV - PARTIÇÕES DO GRUPO 1 2 3 4 5.

PARTIÇÃO	SQW1	SQW2	SQW	SQB
1 : 2 3 4 5	0	119,63	119,63	39,34
1 2 : 3 4 5	1,04	78,23	79,26	79,71
1 2 3 : 4 5	1,76	2,40	4,16	154,81
1 2 3 4 : 5	80,76	0	80,76	78,21

$$B_o = \text{MAX (SQB)} = 154,81$$

$$\text{Partição} = 1 \quad 2 \quad 3 : 4 \quad 5$$

$$\lambda = \frac{\pi}{2(\pi - 2)} \cdot \frac{B_o}{\hat{\sigma}_o^2} = \frac{\pi}{2(\pi - 2)} \cdot \frac{154,81}{4,94} = 43,09$$

$$v_o = \frac{k}{\pi - 2} = \frac{5}{\pi - 2} = 4,38$$

Sob a hipótese nula de indivisibilidade do grupo, $\lambda \sim \chi^2(4)$ e $P_o(\lambda \geq 31,32) = 0,00$. Portanto, ao nível de significância 0,05 os grupos de médias 1 2 3 e 4 5 diferem significativamente entre si.

5.3.6 - Partição do Grupo 1 2 3

$$SQT = \frac{\sum_{i=1}^2 \sum_{j>i}^3 q_{ij}}{k} = \frac{(2,07 + 3,10 + 0,10)}{3} = 1,76$$

$$\hat{\sigma}_0^2 = \frac{SQT + v s^2}{k + v} = \frac{1,76 + 54 \times 2,46}{3 + 54} = 2,36$$

TABELA V - PARTIÇÕES DO GRUPO 1 2 3

PARTIÇÃO	SQW1	SQW2	SQW	SQB
1 : 2 3	0	0,05	0,05	1,71
1 2 : 3	1,04	0	1,04	0,72

$$B_0 = \text{MAX (SQB)} = 1,71$$

$$\text{Partição} = 1 : 2 \quad 3$$

$$\lambda = \frac{\pi}{2(\pi - 2)} \cdot \frac{B_0}{\hat{\sigma}_0^2} = \frac{\pi}{2(\pi - 2)} \cdot \frac{1,71}{2,36} = 0,99$$

$$v_0 = \frac{k}{\pi - 2} = \frac{3}{\pi - 2} = 2,63$$

Sob a hipótese nula de indivisibilidade do grupo, $\lambda \sim \chi^2(2)$ e $P_0(\lambda \geq 0,99) = 0,61$. Portanto, ao nível de significância 0,05 não temos evidências estatísticas para crermos que o grupo possa ser dividido.

5.3.7 - Partição do Grupo 4 5

$$SQT = \frac{\sum_{i=4}^4 \sum_{j>i}^5 q_{ij}}{k} = \frac{4,80}{2} = 2,40$$

$$\hat{\sigma}_0^2 = \frac{SQT + v s^2}{k + v} = \frac{2,40 + 54 \times 2,46}{2 + 54} = 2,41$$

TABELA VI - PARTIÇÃO DO GRUPO 4 5

PARTIÇÃO	SQW1	SQW2	SQW	SQB
4 : 5	0	0	0	2,40

$$B_0 = \text{MAX (SQB)} = 2,40$$

Partição: 4 : 5

$$\lambda = \frac{\pi}{2(\pi - 2)} \cdot \frac{B_0}{\hat{\sigma}_0^2} = \frac{\pi}{2(\pi - 2)} \cdot \frac{2,40}{2,41} = 1,37$$

$$v_0 = \frac{k}{\pi - 2} = \frac{2}{\pi - 2} = 1,75$$

Sob a hipótese nula de indivisibilidade do grupo, $\lambda \sim \chi^2(1)$ e $P_0(\lambda \geq 1,37) = 0,24$. Portanto, ao nível de significância 0,05, não temos evidências estatísticas para crermos que o grupo possa ser dividido.

5.3.8 - Grupos Resultantes

Vamos representar através de um diagrama de árvore, as partições efetuadas que foram significativas ao nível de

significância 0,05:

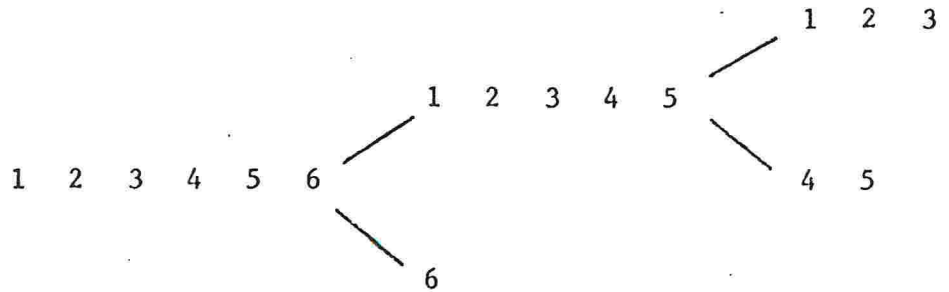


FIGURA 2 - DIAGRAMA EM ÁRVORE DAS PARTIÇÕES SIGNIFICATIVAS

Como podemos observar, foram obtidos três grupos de tratamentos que correspondem exatamente aos grupos definidos previamente na simulação do experimento.

TABELA VII - GRUPOS RESULTANTES

GRUPOS OBTIDOS	TRATAMENTOS	MÉDIA AMOSTRAL DO GRUPO
1	A_2, A_1 e A_3	16,49
2	A_4, A_5	27,85
3	A_6	47,72

6 - PROGRAMA DE COMPUTADOR

Os cálculos para a obtenção de grupos de tratamentos são bastante trabalhosos como podemos observar no exemplo anterior. Por esta razão, desenvolvemos um programa na linguagem FORTRAN IV para a aplicação deste procedimento. No apêndice B, encontra-se uma listagem do programa fonte.

6.1 - UTILIZAÇÃO DO PROGRAMA

Os cartões de controle em WFL (Work Flow Language), para a execução do programa em computadores da linha Burroughs, com versão de software 3.6 são:

```
BEGIN JOB <nome do job>;  
COMPILE OBJETO FORTRAN LIBRARY;  
FORTRAN DATA  
<programa fonte>  
?RUN OBJETO ;  
DATA FILE5
```

<arquivo de dados>

? END JOB

Obs.: Utilizamos a palavra cartão no sentido de registro de arquivo.

6.2 - ARQUIVO DE DADOS

CARTÃO 1

Colunas	Descrição
1 - 2	Número de médias
4 - 6	Número de observações de cada parcela
8 - 12	Nível de significância
14 - 15	Valor de w no formato Fw.d do FORTRAN para leitura das médias
17 - 17	Valor de d no formato Fw.d do FORTRAN para leitura das médias.
19 - 21	Graus de liberdade do resíduo (v).
23 - 32	Valor do QMR

34 - 34	1: a matriz das distâncias entre as médias é impressa. 0: a matriz não é impressa.
36 - 36	1: a matriz dos quadrados das distâncias entre as médias é impressa. 0: a matriz não é impressa.
38 - 38	1: o relatório das partições efetuadas é impresso. 0: o relatório não é impresso.
40 - 40	1: o resumo das partições efetuadas é impresso. 0: o resumo não é impresso.

— CARTÃO 2

Colunas	Descrição
1 - 42	Título do trabalho.

CARTÕES 3, 4, ...

Cartões contendo as médias (uma em cada cartão)

Colunas	Descrição
1 - 6	Nome do tratamento.
8 - 72	Valor da média no formato Fw.d definido no cartão 1.

Se houver mais de um problema a ser analisado, repetimos toda esta seqüência de cartões para cada um dos outros problemas.

6.3 - OBSERVAÇÕES

O número de médias que o programa pode receber é de no máximo 50.

As médias não precisam estar ordenadas previamente, pois o programa possui uma rotina de ordenação.

Se o nível de significância não for informado ou for inválido, o programa utilizará o nível 0,05.

Os relatórios foram desenhados para serem impressos com a densidade de seis linhas por polegada.

6.4 - EXEMPLO DE USO DO PROGRAMA

Para ilustrar o uso do programa, apresentamos o arquivo de dados e as saídas relativas ao experimento simulado do exemplo anterior.

ARQUIVO DE DADOS

06 010 0.050 05 2 054 00024.5688 1 1 1 1

EXPERIMENTO SIMULADO

A1	16.86
A2	15.42
A3	17.18
A4	26.75
A5	28.94
A6	47.72

AGRPAMENTO DE TRATAMENTOS NA ANALISE DE VARIANCIA

METODO DE A. J. SCOTT E. P. KNOTT (1974)

PROBLEMA : EXPERIMENTO SIMULADO

NUMERO DE OBSERVACOES EM CADA PARCELA : 10

NUMERO DE MEDIAS : 6

QUADRADO MEDIO RESIDUAL (QMR) : 24.5688

GRAUS DE LIBERDADE DO QMR : 54

VARIANCIA ESTIMADA DAS MEDIAS : 2.4569

NIVEL DE SIGNIFICANCIA : 0.050

MEDIAS A SEREM AGRUPADAS

1.	A2	=	15.42
2.	A1	=	16.86
3.	A3	=	17.18
4.	A4	=	26.75
5.	A5	=	28.94
6.	A6	=	47.72

MATRIZ DAS DISTANCIAS ENTRE AS MEDIAS (ACIMA DA DIAGONAL PRINCIPAL)

	1	2	3	4	5	6
1	0.00	1.44	1.76	11.33	13.52	32.30
2	0.00	0.00	0.32	9.89	12.08	31.36
3	0.00	0.00	0.00	9.57	11.76	30.54
4	0.00	0.00	0.00	0.00	2.19	20.97
5	0.00	0.00	0.00	0.00	0.00	13.78
6	0.00	0.00	0.00	0.00	0.00	0.00

MATRIZ DOS QUADRADOS DAS DISTANCIAS ENTRE AS MEDIAS (ACIMA DA DIAGONAL PRINCIPAL)

	1	2	3	4	5	6
1	0.0000	2.0736	3.0976	128.3689	182.7904	1043.2900
2	0.0000	0.0000	0.1024	97.3121	145.9264	952.3396
3	0.0000	0.0000	0.0000	91.5849	138.2976	932.6916
4	0.0000	0.0000	0.0000	0.0000	4.7961	439.7409
5	0.0000	0.0000	0.0000	0.0000	0.0000	352.6884
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

PARTICULO 1

1 - 5 : 6 - 5

SQT= 722.500

SQB= 573.630

SCW= 159.970

SIGMA= 14.7545

LAMBDA= 57.3604

SIGNIFICANCIA= 0.000 (5 G.L.)

PARTICULO 2

1 - 3 : 4 - 5

SQT= 158.970

SQB= 154.814

SCW= 4.15592

SIGMA= 4.94308

LAMBDA= 43.0945

SIGNIFICANCIA= 0.000 (4 G.L.)

PARTICULO 3

1 - 1 : 2 - 3

SQT= 1.75737

SQB= 1.70667

SCW= .512000E+01

SIGMA= 2.35841

LAMBDA= 0.995722

SIGNIFICANCIA= 0.608 (2 G.L.)

PARTICAD 4

4 - 4 : 5 - 5

SQT= 2.39805

SQB= 2.39805

SQW= 0.

SIGMA= 2.41196

LAMBDA= 1.35804

SIGNIFICANCIA= 0.242 (1 G.L.)

RESUMO DAS PARTICOES

<u>GRUPO</u>	<u>SUBGRUPO 1</u>	<u>SUBGRUPO 2</u>	<u>ETAPA DA PARTICAO</u>	<u>LAMBDA</u>	<u>SIGNIFICANCIA</u>	<u>GRAUS DE LIBERDADE</u>
1 - 6	1 - 5	6 - 6	1	55.3504	0.000	5
1 - 5	1 - 3	4 - 5	2	43.0945	0.000	4
1 - 3	1 - 1	2 - 3	3	0.5957	0.508	2
4 - 5	4 - 4	5 - 5	3	1.3680	0.242	1

GRUPPO 1 - MEDIA = 16.49

A2

A1

A3

GRUPPO 2 - MEDIA = 27.35

A4

A5

GRUPPO 3 - MEDIA = 47.72

A6

ANALISI
SISTEMI DI QUALITÀ - 2
MATEMATICA STATISTICA

7 - CONCLUSÃO

Este trabalho procurou expor o método de agrupar tratamentos sugerido por Scott e Knott. Esta técnica é bastante simples, porém trabalhosa, fazendo-se necessário o uso de recursos computacionais. Assim sendo, desenvolvemos um programa em FORTRAN IV para computadores de grande porte da linha Burroughs com versão de software 3.6. Não obstante, este programa é de fácil adaptação para outros equipamentos, inclusive para microcomputadores.

É interessante salientar a importância da análise da normalidade e da homoscedasticidade, uma vez que a técnica foi desenvolvida tendo por base estes dois pressupostos. No exemplo apresentado neste trabalho, esta etapa foi suprimida, uma vez que as condições citadas já estavam implícitas no processo de simulação das amostras. Outro aspecto a ser destacado, é que a aplicação desta técnica é restrita a experimentos com delineamentos balanceados.

Nos apêndices C e D, apresentamos o algoritmo e o fluxograma de tal forma que o programa possa ser transcrito para outra linguagem se necessário for.

BIBLIOGRAFIA

1. BURROUGHS. A Series FORTRAN. U.S.A., Burroughs Corporation, 1985.
2. BURROUGHS. A Series Work Flow Language (WFL). U.S.A., Burroughs Corporation, 1985.
3. EDWARDS, A. W. e CAVALLI-SFORZA, L. L. A Method for Cluster Analysis. In: _____. Biometrics. U.S.A., june, 1965, p.363-375.
4. PERES, C.A. e SALDIVA C.D. Planejamento de Experimentos. In: 5º Simpósio Nacional de Probabilidade e Estatística, 1982.
5. SCOTT, C.A. e KNOTT, M. A Cluster Analysis Method for Grouping Means in the Analysis of Variance. In.: _____. Biometrics. U.S.A., 30:507-12, sept., 1974.
6. TEXAS Instruments Incorporated. TI Programável 58/59 - Biblioteca Geral. Campinas, Texas Instrumentos Eletrônicos do Brasil Ltda., 1977.

APÊNDICES

APÊNDICE A
LISTAGENS DO PROGRAMA ANOVAR

UCD ANOVAR, VERSION 2.4.0, 09/73

PROBLEM EXPERIMENTO SIMLLADO
 N-WAY, 1
 VARIATES, Y
 LIMITS, 6, 10
 ORDER, Y
 FORMAT(10R5.2)
 MEANS, ALL
 *DATA, LIST, CARDS

NO. OF VARIATES 1
 NO. OF TRANSFORMATIONS 0
 NO. OF OBSERVATIONS 60

FACTOR RANGE
 A 6

VARIATE LISTING

A	Y
1	16.66
1	14.95
1	24.14
1	23.00
1	26.34
1	12.03
1	11.79
1	15.65
1	12.46
1	11.58
2	10.34
2	6.86
2	11.14
2	18.90
2	12.90
2	20.21
2	20.19
2	14.94
2	23.46
2	15.52
3	19.81
3	11.39
3	20.42
3	15.21
3	18.87
3	19.72
3	14.53
3	11.68
3	17.07
3	23.09
4	23.35
4	27.37
4	28.72
4	30.15
4	16.88
4	20.08
4	30.08
4	23.91
4	37.69
4	28.75
5	23.92
5	25.21
5	31.26
5	29.73
5	31.23
5	31.46
5	29.97
5	31.82
5	22.29
5	32.49
6	41.11
6	45.57
6	52.70
6	42.90
6	48.76
6	51.79
6	46.19
6	56.75
6	42.50
6	48.94

ANALYSIS OF VARIANCE TABLE FOR (Y) - OVERALL MEAN IS 25.477

SOURCE	D.F.	S.S.	M.S.	F
A	5	7526.8247	1505.3649	61.271
ERROR	54	1326.7171	24.5688	
TOTAL	59	8853.5418		

VARIATE MEANS

A	N	Y
1	10	16.860
2	10	15.418
3	10	17.179
4	10	26.743
5	10	28.938
6	10	47.721

APÊNDICE B
LISTAGENS DO PROGRAMA FONTE

```

C                               P R O G R A M A   F O N T E
C                               -----

C
C -----
C  DECLARACOES
C -----
C
C      INTEGER D,FASE,F,GLR,GLO,PART,OP(4),PILHA1(50,3),Q,QUEBRA,R
C      REAL    LABEL(50),GRUPO(50,4),PILHA2(50,7)
C      REAL    CAB(7),DE(50,50),LAMEDA,MEDIAS(50),NSIG,QDE(50,50)
C      LOGICAL B
C
C -----
C  CALCULO DOS VALORES CONSTANTES
C -----
C
C      PI=ARCOS(-1.)
C      PI2=PI-2.
C      CSIE=PI/(2*PI2)
C
C -----
C  LEITURA E IMPRESSAO DOS DADOS INICIAIS
C -----
C
C      1 READ (5,100,END=37) NM,N,NSIG,F,D,GLR,QMR,OP,CAB
C      S2=QMR/N
C      DO 2 I=1,NM
C      READ (5,101) LABEL(I),F,D,MEDIAS(I)
C      2 CONTINUE
C      IF (NSIG.LE.0.09.NSIG.GT.1.0) NSIG=0.05
C      WRITE (6,200) CAB,N,NM,QMR,GLR,S2,NSIG
C
C

```

```

C -----
C  ORDENACAO E IMPRESSAO DAS MEDIAS
C  -----
C
      DO 4 I=1,NM-1
        NSMALL=I
        ISTART=I+1
        DO 3 J=ISTART,NM
          IF (MEDIAS(J).LT.MEDIAS(NSMALL)) NSMALL=J
        3 CONTINUE
        SAVE=MEDIAS(NSMALL)
        MEDIAS(NSMALL)=MEDIAS(I)
        MEDIAS(I)=SAVE
        SAVE=LABEL(NSMALL)
        LABEL(NSMALL)=LABEL(I)
        LABEL(I)=SAVE
      4 CONTINUE
      WRITE (6,201) NM,(J,LABEL(J),F,D,MEDIAS(J),J=1,NM)
C
C -----
C  CALCULO DAS MATRIZES DE BISTANCIAS ENTRE AS MEDIAS ("Q" E "D")
C  -----
C
      DO 5 I=1,NM
        DO 5 J=1,I
          DE(J,I)=MEDIAS(I)-MEDIAS(J)
          QDE(J,I)=DE(J,I)**2
        5 CONTINUE
C
C -----
C  IMPRESSAO DA MATRIZ "D"
C  -----
C
      NMAT=58/(4+NM)
      IF (OP(1).NE.1) GO TO 9
      NCCL=122/(6+F)

```

```

NREP=NM/NCOL
NRESTO=MOD(NM,NCOL)
R=F+1
J=1
N1=999
IF (NREP.EQ.0) GO TO 8
DO 8 I=1,NREP
6 IF (N1.LT.NMAT) GO TO 7
N1=0
WRITE (6,202)
7 WRITE (6,203)
*NCOL,(R,J1,J1=J,(J+NCOL-1))
WRITE (6,204) NM,(J2,NCOL,
*(F,C,DE(J2,J3),J3=J,(J+NCOL-1)),J2=1,NM)
N1=N1+1
J=J1
8 CONTINUE
IF (NRESTO.EQ.0) GO TO 9
NCOL=NRESTO
NRESTO=0
GO TO 6
C
C -----
C IMPRESSAO DA MATRIZ "Q"
C -----
C
9 IF (OP(2).NE.1) GO TO 13
NCOL=122/(6+2*F)
NREP=NM/NCOL
NRESTO=MOD(NM,NCOL)
R=2*F+1
J=1
N1=999
IF (NREP.EQ.0) GO TO 12
DO 12 I=1,NREP
10 IF (N1.LT.NMAT) GO TO 11

```

```

      NI=0
      WRITE (6,206)
11  WRITE (6,203) NCOL,(R,J1,J1=J,(J+NCOL-1))
      WRITE (6,204) NM,(J2,NCOL,(F*2,D*2,
      *SDE(J2,J3),J3=J,(J+NCOL-1)),J2=1,NM)
      NI=NI+1
      J=J1
12  CONTINUE
      IF (NRESTO.EQ.0) GO TO 13
      NCOL=NRESTO
      NRESTO=0
      GO TO 10

C
C -----
C PARTICAD DAS MEDIAS EM GRUPOS
C -----
C

13  L=0
      FASE=0
      PART=4
      DO 14 J=1,L1
      GRUPO(L1,4)=0
14  CONTINUE
      L1=0
      PILHA1(1,1)=1
      PILHA1(1,2)=NM
      PILHA1(1,3)=0
      I=1
15  MI=PILHA1(I,1)
      MF=PILHA1(I,2)
      ETAPA=PILHA1(I,3)+1
      I=I+1
      SQT=0.
      BQ=0.
      K=MF-MI+1
      DO 16 J=MI,MF

```



```

DO 16 J1=MI,MF
SQT=SQT+QDE(J,J1)
16 CONTINUE
SQT=SQT/(MF-MI+1)
DO 19 Q=MI,(PF-1)
SQW1=0.
SQW2=0.
DO 17 J=MI,Q
DO 17 J1=MI,Q
SQW1=SQW1+QDE(J,J1)
17 CONTINUE
SQW1=SQW1/(Q-MI+1)
DO 18 J=(Q+1),MF
DO 18 J1=(Q+1),MF
SQW2=SQW2+QDE(J,J1)
18 CONTINUE
SQW2=SQW2/(MF-Q)
SQW=SQW1+SQW2
SQB=SQT-SQW
IF (SQB.LE.B0) GO TO 19
EO=SQB
QUEBRA=Q
19 CONTINUE
IF (SQT.EQ.0) QUEBRA=MF-1
SIGMA0=(SQT+GLR+S2)/(K+GLR)
LAMBDA=CSTE*EO/SIGMA0
GLO=IFIX(K/PI2)
SQWM=SQT-B0
GUI=CHIPRB(LAMBDA,GLO)
IF (GUI.GT.NSIG) GO TO 21
IF ((QUEBRA+1).EQ.MF) GO TO 20
I= I+1
PILFA1(I,1)=QUEBRA+1
PILFA1(I,2)=MF
PILFA1(I,3)=ETAPA
20 IF (MI.EQ.QUEBRA) GO TO 21

```

```

I=I+1
PILHA1(I,1)=MI
PILHA1(I,2)=QUEBRA
PILHA1(I,3)=ETAPA
21 L=L+1
PILHA2(L,1)=MI
PILHA2(L,2)=QUEBRA
PILHA2(L,3)=MF
PILHA2(L,4)=ETAPA
PILHA2(L,5)=LAMBDA
PILHA2(L,6)=QUI
PILHA2(L,7)=GLO
B=((MI.EQ.QUEBRA).AND.(QUI.LE.NSIG))
IF (.NOT.B) GO TO 22
L1=L1+1
GRUPO(L1,1)=MI
GRUPO(L1,2)=MI
GRUPO(L1,3)=1
GRUPO(L1,4)=MEDIAS(MI)
22 B=(MF.EQ.(QUEBRA+1)).AND.(QUI.LE.NSIG)
IF (.NOT.B) GO TO 23
L1=L1+1
GRUPO(L1,1)=MF
GRUPO(L1,2)=MF
GRUPO(L1,3)=1
GRUPO(L1,4)=MEDIAS(MF)
23 IF (QUI.LE.NSIG) GO TO 25
L1=L1+1
GRUPO(L1,1)=MI
GRUPO(L1,2)=MF
GRUPO(L1,3)=K
DO 24 J=MI,MF
GRUPO(L1,4)=GRUPO(L1,4)+MEDIAS(J)
24 CONTINUE
GRUPO(L1,4)=GRUPO(L1,4)/K
25 IF (OP(3).NE.1) GO TO 27

```

```

FASE=FASE+1
IF (PART.NE.4) GO TO 26
WRITE (6,205)
PART=0
26 WRITE (6,207) FASE,MI,QUEBRA,QUEERA+1,
  *MF,SQT,B0,SQWM,SIGMA0,LAMBDA,QUI,GLO
PART=PART+1
27 IF (I.GT.0) GO TO 15
C
C -----
C IMPRESSAO DO RESUMO DAS PARTICOES
C -----
C
28 IF (OP(4).NE.1) GO TO 32
IF (L.EQ.1) GO TO 31
DO 30 I=1,L-1
NSMALL=I
ISTART=I+1
DO 29 J=ISTART,L
IF (PILHA2(J,4).LT.PILHA2(NSMALL,4)) NSMALL=J
29 CONTINUE
DO 30 M=1,7
SAVE=PILHA2(NSMALL,M)
PILHA2(NSMALL,M)=PILHA2(I,M)
PILHA2(I,M)=SAVE
30 CONTINUE
31 WRITE (6,208)
  *L,(PILHA2(I,1),PILHA2(I,3),PILHA2(I,1),PILHA2(I,2),
  *PILHA2(I,2)+1,PILHA2(I,3),PILHA2(I,4),PILHA2(I,5),
  *PILHA2(I,6),PILHA2(I,7),I=1,L)
C
C -----
C IMPRESSAO DOS GRUPOS RESULTANTES
C -----
C
32 IF (L1.EQ.1) GO TO 34

```

```

DO 34 I=1,L1-1
NSMALL=I
ISTART=I+1
DO 33 J=ISTART,L1
IF (GRUPO(J,1).LT.GRUPO(NSMALL,1)) NSMALL=J
33 CONTINUE
DO 34 M=1,4
SAVE=GRUPO(NSMALL,M)
GRUPO(NSMALL,M)=GRUPO(I,M)
GRUPO(I,M)=SAVE
34 CONTINUE
LINHAS=63
DO 36 J=1,L1
IF ( (GRUPO(J,3)*2+7).LE.(63-LINHAS) ) GO TO 35
WRITE (6,205)
LINHAS=0
35 LINHAS=LINHAS+GRUPO(J,3)*2+7
WRITE (6,209) J,F,D,GRUPO(J,4),GRUPO(J,3),
*(LABEL(I),I=GRUPO(J,1),GRUPO(J,2))
36 CONTINUE
C
C -----
C FORMATOS PARA LEITURA E IMPRESSAO
C -----
C
100 FORMAT (I2,1X,I3,1X,F5.0,1X,I2,1X,I1,1X,I3,1X,F10.0,
*4(1X,I1) / 7A6)

101 FORMAT (A6,1X,F*,*)

200 FORMAT ('1', 132('*') /// T43,
*'AGRUPAMENTO DE TRATAMENTOS NA ANALISE DE VARIANCIA' /
*/ T49,'METODO DE A. J. SCOTT E. M. KNOTT (1974)',
*/// 132('*'),21 (/),
*T43,'PROBLEMA : ',7A6 //
*T43,'NUMERO DE OBSERVACOES EM CADA PARCELA : ',I3 //

```

```

*T43,'NUMERO DE MEDIAS : ',I2 //
*T43,'QUADRADO MEDIO RESIDUAL (QMR) : ',F10.4 //
*T43,'GRAUS DE LIBERDADE DO QMR : ',I3 //
*T43,'VARIANCIA ESTIMADA DAS MEDIAS : ',F10.4 //
*T43,'NIVEL DE SIGNIFICANCIA : ',F5.3)

```

```

201 FORMAT ('1' /// T10,'MEDIAS A SEREM AGRUPADAS' /
*T10,24(' ') //) // * (T10,F3.0,2X,A6,' = ',F*.* //)

```

```

202 FORMAT ('1' /// T2,'MATRIZ DAS DISTANCIAS ENTRE AS'
*' MEDIAS (ACIMA DA DIAGONAL PRINCIPAL)' /
*T2,68(' ') //)

```

```

203 FORMAT (// T10,*(X,I2) //)

```

```

204 FORMAT (*(T8,I2,*(3X,F*.* ) /) )

```

```

205 FORMAT ('1')

```

```

206 FORMAT ('1' /// T2,'MATRIZ DOS QUADRADOS DAS '
*' DISTANCIAS ENTRE AS MEDIAS (ACIMA DA DIAGONAL'
*' PRINCIPAL)' / T2,83(' ') //)

```

```

207 FORMAT (T10,' P A R T I C A C ',I3 / T10,1X,
*'19(' ') /// T11,I2,' - ',I2,' : ',I2,' - ',
*I2 /// T10,' SQT= ',G12.6,T50,' SOB= ',G12.6,
*T90,' SQW= ',G12.6 // T10,' SIGMA= ',G12.6,
*T51,' LAMBDA= ',G12.6,T90,' SIGNIFICANCIA= ',
*F5.3,' ( ',I2,' G.L. )' //)

```

```

208 FORMAT ('1' // T51,' R E S U M O   D A S   P A R T I '
*' C C E S' / T51,39(' ') //) // T11,'GRUPO',T22,
*'SUBGRUPO', ' 1',T37,'SUBGRUPO 2',T52,'ETAPA DA '
*'PARTICAO',T73,'LAMBDA',T87,'SIGNIFICANCIA',T106,
*'GRAUS DE LIBERDADE' / T11,5(' '),T22,10(' '),T37,
*10(' '),T52,17(' '),T73,6(' '),T87,13(' '),T106,

```

```
*18('-' ) // *(T10,I2,' - ',I2,T24,I2,' - ',I2,T39,
*I2,' - ',I2,T57,I2,T72,F8.4,T91,F5.3,T114,I2 /))
```

```
209 FORMAT (T10,'G R U P O ',I2,' - ',
*'M E D I A = ',F*.* // // *(T25,A6 // ) // // )
```

```
C
C -----
C FINAL DO PROBLEMA
C -----
C
```

```
GO TO 1
37 CALL EXIT
END
```

FUNCTION CHIPRB(XS,N)

```

C
C -----
C  CALCULO DA SIGNIFICANCIA DE LAMBDA
C  -----
C
      REAL A,C,D,E,Z,S,Y,X
      X=XS
      IF (X.LT.E=35) GO TO 13
      I=(N/2)*2
      A=X*.5E0
      IF (I.EQ.N) GO TO 1
      S=1.E1-ERF(SQRT(X)/1.4142135624E0)
      Y=EXP(-A)
      GO TO 2
1  Y=EXP(-A)
      S=Y
2  IF (N.GT.2) GO TO 6
      CHIPRB=S
      RETURN
6  Z=.5E0
      J=3
      IF (I.NE.N) GO TO 7
      Z=1.E0
      J=4
7  X=(N-1.E0)*.5E0
      IF (X.LT.1.E35) GO TO 11
      E=0.D0
      IF (I.NE.N) E=.5723649429300
      C=ALOG(A)
      DO 8 NM=J,N,2
      Z=Z+1.E0
      E=ALOG(Z)+E
8  S=EXP(C*Z-A-E)*S
      CHIPRB=S

```

```
      RETURN
11  E=1.E0
      IF (I.NE.N) E=.5641895835E0/SQRT(A)
      C=C.E0
      DO 12 NM=J,N,2
      E=E*A/Z
      C=C+E
12  Z=Z+1.E0
      CHIPRB=C*Y+S
      RETURN
13  CHIPRB=1.EC
      RETURN
      END
```

OBS: esta rotina foi transcrita de notas de
aula (distribuição qui-quadrado).

APÊNDICE C
ALGORITMO DAS PARTIÇÕES

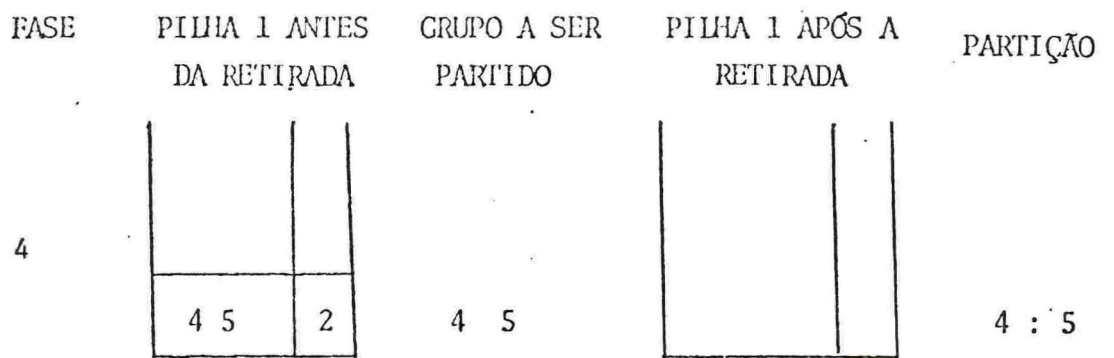
Para fazer as partições, utilizamos duas pilhas de trabalho: a PILHA 1 que contém os grupos de médias a serem partidas juntamente com a etapa na qual foram geradas; e a PILHA 2, que contém as informações sobre as partições efetuadas.

GRUPO	ETAPA	SUBGRUPO 1	SUBGRUPO 2	ETAPA	λ	χ^2	ν_0

PILHA 1
PILHA 2

Para mostrar a lógica utilizada, vamos representar, graficamente, a execução das partições no problema do experimento simulado, apresentado anteriormente,

FASE	PILHA 1 ANTES DA RETIRADA	GRUPO A SER PARTIDO	PILHA 1 APÓS A RETIRADA	PARTIÇÃO										
1	<table border="1"> <tr><td></td><td></td></tr> <tr><td>1 2 3 4 5 6</td><td>0</td></tr> </table>			1 2 3 4 5 6	0	1 2 3 4 5 6	<table border="1"> <tr><td></td><td></td></tr> <tr><td></td><td></td></tr> </table>					1 2 3 4 5 : 6 *		
1 2 3 4 5 6	0													
2	<table border="1"> <tr><td></td><td></td></tr> <tr><td>1 2 3 4 5</td><td>1</td></tr> </table>			1 2 3 4 5	1	1 2 3 4 5	<table border="1"> <tr><td></td><td></td></tr> <tr><td></td><td></td></tr> </table>					1 2 3 : 4 5 *		
1 2 3 4 5	1													
3	<table border="1"> <tr><td></td><td></td></tr> <tr><td>1 2 3</td><td>2</td></tr> <tr><td>4 5</td><td>2</td></tr> </table>			1 2 3	2	4 5	2	1 2 3	<table border="1"> <tr><td></td><td></td></tr> <tr><td>4 5</td><td>2</td></tr> </table>			4 5	2	1 : 2 3 *
1 2 3	2													
4 5	2													
4 5	2													



(*) Partições significativas

PILHA 2 NO FINAL DO PROCESSO

4	5	3	1,37	0,24	1
1	23	3	0,99	0,61	2
1 2 3	45	2	43,09	0,00	4
1 2 3 4 5	6	1	55,36	0,00	5
SUBGRUPO 1	SUBGRUPO 2,	ETAPA	λ	χ^2	v_0

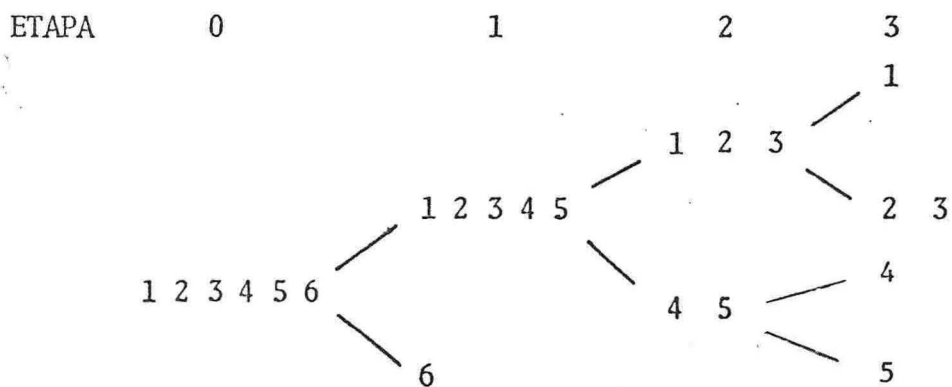


FIGURA 3 - PARTIÇÕES EFETUADAS

Observação:

Um grupo de médias fica perfeitamente identificado pelo par ordenado (MI, MF) , onde MI é a ordem da primeira média do grupo e MF , a da última. Por exemplo, o grupo de médias 2 3 4 5 é equivalente ao par ordenado $(2,5)$.

APÊNDICE D

FLUXOGRAMA

