



**RAD1509 – Estatística Aplicada à Administração II**  
**Prof. Dr. Evandro Marcos Saidel Ribeiro**  
**RESUMO para software MINITAB**

**1 Matriz de Covariância**

Organize os dados na forma de matriz, por exemplo, na coluna 1 liste as 5 observações da variável  $x_1$  e na coluna 2 liste as 5 observações da variável  $x_2$ :

$$X = \begin{bmatrix} 2 & 1 \\ 9 & 8 \\ 7 & 5 \\ 3 & 4 \\ 4 & 2 \end{bmatrix}. \quad (1)$$

No Minitab:

|   | C1 | C2 | C3 | C4 |
|---|----|----|----|----|
|   | x1 | x2 |    |    |
| 1 | 2  | 1  |    |    |
| 2 | 9  | 8  |    |    |
| 3 | 7  | 5  |    |    |
| 4 | 3  | 4  |    |    |
| 5 | 4  | 2  |    |    |
| 6 |    |    |    |    |
| 7 |    |    |    |    |

Calcule as médias das variáveis:

$$\bar{x}_1 = 5 \quad \bar{x}_2 = 4. \quad (2)$$

**No Minitab, estatísticas descritivas:**

Stat > Basic Statistics > Display Descriptive Statistics >

- Clique no quadro [Variables];
- Marque as variáveis;
- Clique no botão [select]

Resultado:

| Descriptive Statistics: x1; x2 |   |    |      |         |       |          |         |         |      |        |
|--------------------------------|---|----|------|---------|-------|----------|---------|---------|------|--------|
| Variable                       | N | N* | Mean | SE Mean | StDev | Variance | CoefVar | Minimum | Q1   | Median |
| x1                             | 5 | 0  | 5,00 | 1,30    | 2,92  | 8,50     | 58,31   | 2,00    | 2,50 | 4,00   |
| x2                             | 5 | 0  | 4,00 | 1,22    | 2,74  | 7,50     | 68,47   | 1,00    | 1,50 | 4,00   |

| Variable | Q3   | Maximum |
|----------|------|---------|
| x1       | 8,00 | 9,00    |
| x2       | 6,50 | 8,00    |

A matriz de covariância,  $S$ , é dada por:

$$S = \begin{bmatrix} 8,50 & 7,25 \\ 7,25 & 7,50 \end{bmatrix}. \quad (3)$$

**No Minitab, covariância:**

Stat > Basic Statistics > Covariance

Selecionar variáveis:

| Covariances: x1; x2 |         |         |
|---------------------|---------|---------|
|                     | x1      | x2      |
| x1                  | 8,50000 |         |
| x2                  | 7,25000 | 7,50000 |

**2 Matriz de correlação**

A matriz de correlação é escrita na forma:

$$R = \begin{bmatrix} 1 & r_{12} & \dots & r_{1k} \\ r_{21} & 1 & \dots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \dots & 1 \end{bmatrix}. \quad (4)$$

No exemplo a matriz de correlação será:

$$R = \begin{bmatrix} 1 & 0,908025 \\ 0,908025 & 1 \end{bmatrix}. \quad (5)$$

**No Minitab, correlação:**

Stat > Basic Statistics > Correlation

Resultado (observe também o valor-P)

| Correlations: x1; x2             |         |
|----------------------------------|---------|
| Pearson correlation of x1 and x2 | = 0,908 |
| P-Value                          | = 0,033 |

**2 Autovalores e autovetores**

1º passo: Determinar autovalores. Para a matriz de covariância  $S$  os autovalores são:

$$\lambda = \frac{(s_{11} + s_{22})}{2} \pm \frac{\sqrt{(s_{11} + s_{22})^2 - 4(s_{11}s_{22} - s_{12}^2)}}{2} \quad (6)$$

No caso específico da matriz **S** obtida em (3), os autovalores são:

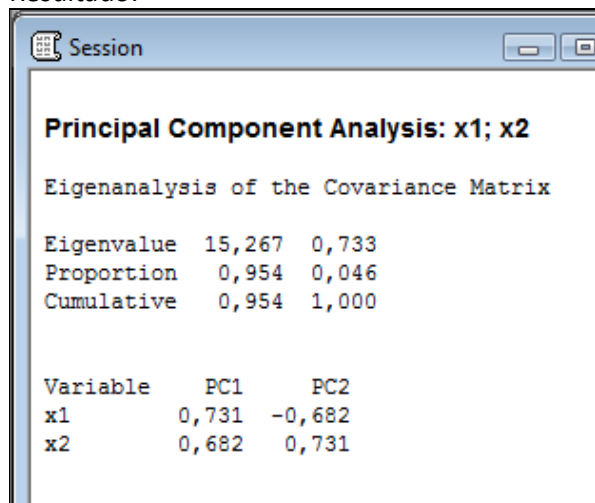
$$\lambda_1 = 15,267221 \quad \lambda_2 = 0,732779. \quad (7)$$

**No Minitab, Comp. Princip., autovalores da matriz de covariância:**

Stat > Multivariate > Principal Components

- Selecione as variáveis
- Selecione (x) Covariance
- Graphs > (Scree Plot)

Resultado:



**Principal Component Analysis: x1; x2**

Eigenanalysis of the Covariance Matrix

|            | Eigenvalue | 15,267 | 0,733 |
|------------|------------|--------|-------|
| Proportion | 0,954      | 0,046  |       |
| Cumulative | 0,954      | 1,000  |       |

| Variable | PC1   | PC2    |
|----------|-------|--------|
| x1       | 0,731 | -0,682 |
| x2       | 0,682 | 0,731  |

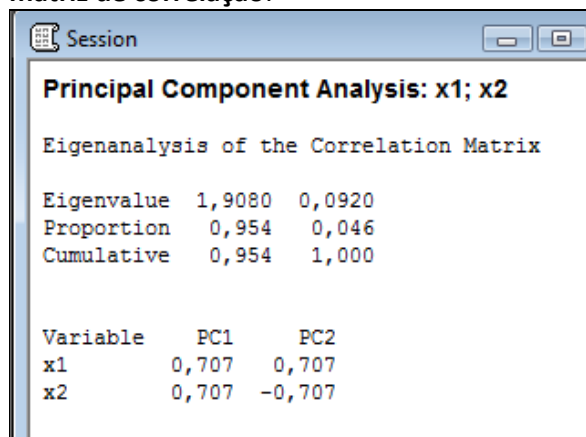
Os autovalores da matriz de correlação **R** são:

$$\lambda = 1 \pm r_{12}. \quad (8)$$

No caso específico da matriz **R** obtida em (5), os autovalores são:

$$\lambda_1 = 1,908025 \quad \lambda_2 = 0,091975. \quad (9)$$

**No Minitab: Comp. Princip., autovalores da matriz de correlação:**



**Principal Component Analysis: x1; x2**

Eigenanalysis of the Correlation Matrix

|            | Eigenvalue | 1,9080 | 0,0920 |
|------------|------------|--------|--------|
| Proportion | 0,954      | 0,046  |        |
| Cumulative | 0,954      | 1,000  |        |

| Variable | PC1   | PC2    |
|----------|-------|--------|
| x1       | 0,707 | 0,707  |
| x2       | 0,707 | -0,707 |

### 3 Componentes Principais

Componentes Principais são escritos em termos das variáveis originais através de ponderações, dadas por componentes dos autovetores **e**. Se

tivermos duas variáveis originais teremos dois componentes principais: **CP = e'X**

Sendo,  $e = \begin{bmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{bmatrix}$  ou seja  $e' = \begin{bmatrix} e_{11} & e_{21} \\ e_{12} & e_{22} \end{bmatrix}$ .

Assim,  $\begin{bmatrix} CP_1 \\ CP_2 \end{bmatrix} = \begin{bmatrix} e_{11} & e_{21} \\ e_{12} & e_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$  e então:

$$\begin{aligned} CP_1 &= e_{11}x_1 + e_{21}x_2 \\ CP_2 &= e_{12}x_1 + e_{22}x_2 \end{aligned} \quad (10)$$

No caso específico do nosso problema 2x2, utilizando os resultados obtidos dos autovetores temos (veja as duas figuras anteriores), para o caso da matriz de covariância:

$$\begin{aligned} CP_1 &= 0,731027 x_1 + 0,682348 x_2 \\ CP_2 &= -0,682348 x_1 + 0,731027 x_2 \end{aligned} \quad (11)$$

- Percentual da variação explicada.

$$\text{Variação explicada por } CP_i = \frac{\lambda_i}{\sum \lambda} \times 100. \quad (12)$$

No exemplo em questão, considerando os autovalores, obtemos:

$$\text{Variação explicada por } CP_1 = 95,42 \%, \quad (13)$$

$$\text{Variação explicada por } CP_2 = 4,58 \%. \quad (14)$$

Veja as duas figuras anteriores.

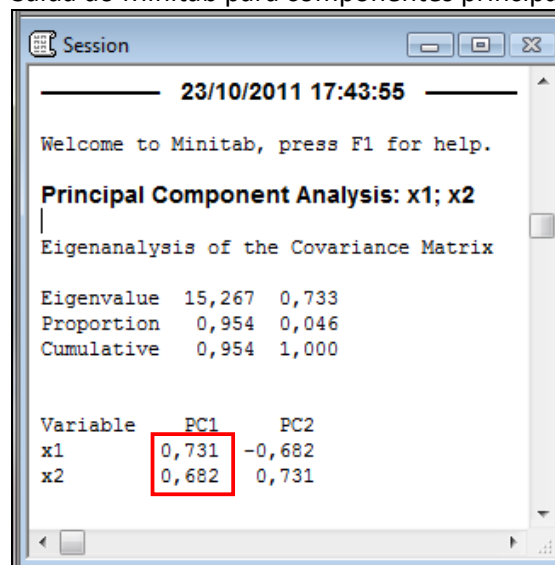
### 3.1 Componentes Principais no Minitab:

No menu do Minitab:

Stat ► Multivariate ► Principal Components

Para que a análise fique semelhante escolha "Covariance"

Saída do Minitab para componentes principais:



**Principal Component Analysis: x1; x2**

Eigenanalysis of the Covariance Matrix

|            | Eigenvalue | 15,267 | 0,733 |
|------------|------------|--------|-------|
| Proportion | 0,954      | 0,046  |       |
| Cumulative | 0,954      | 1,000  |       |

| Variable | PC1   | PC2    |
|----------|-------|--------|
| x1       | 0,731 | -0,682 |
| x2       | 0,682 | 0,731  |

Observe o autovetor  $e_1$  em vermelho acima.

Observe o sinal na CP<sub>2</sub>. Isto se deve à escolha da componente do autovetor que será igual a 1, o Minitab escolhe a segunda componente, se for escolhida a primeira o sinal ficará trocado.

#### 4 Análise Fatorial

O objetivo é descrever um **conjunto de  $k$  variáveis originais** ( $x_1, x_2, \dots, x_k$ ) através da criação de um número menor de variáveis (**fatores**).

Cada variável é descrita em termos de fatores:

$$x_i = a_{i1}F_1 + a_{i2}F_2 + \dots + a_{in}F_n + \varepsilon_i \quad (15)$$

$a_{ij}$  são cargas fatoriais.

A soma das cargas ao quadrado para uma determinada variável é igual a comunalidade da variável:

$$h_i = \sum_{j=1}^n a_{ij}^2 \quad (16)$$

Se a soma for feita para um determinado índice relacionado a um fator então obtemos o autovalor correspondente ao fator em questão:

$$\lambda_j = \sum_{i=1}^k a_{ij}^2 \quad (17)$$

Como resultado, cada Fator pode ser escrito em termos das variáveis sendo possível assim obter uma função para o Score Fatorial  $j$ :

$$F_j = b_{j1}z_1 + b_{j2}z_2 + \dots + b_{jk}z_k \quad (18)$$

Para o cálculo correto do Score Fatorial as variáveis devem ser padronizadas.

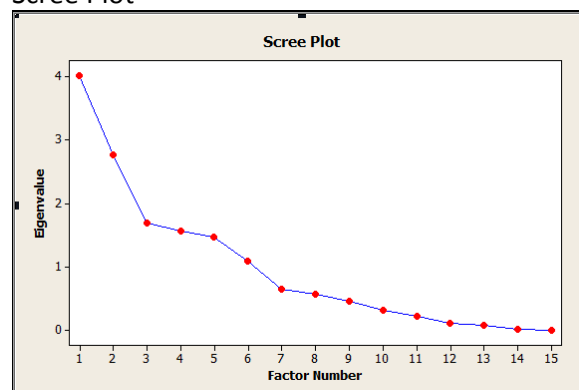
Assim, em (18)  $z_i = (x_i - \bar{x}_i) / s_i$  é o Score-z.

##### 4.1 Análise Fatorial no Minitab:

No menu do Minitab:

Stat ► Multivariate ► Factor Analysis

##### Scree Plot



Considerando apenas 6 fatores a serem extraídos (Critério Kaiser: autovalores > 1):

##### Cargas Fatoriais:

| Variable | Factor1 | Factor2 | Factor3 | Factor4 | Factor5 | Factor6 | Communality |
|----------|---------|---------|---------|---------|---------|---------|-------------|
| ICOM     | 0,966   | -0,082  | 0,151   | 0,021   | -0,012  | -0,083  | 0,971       |
| ICOA     | -0,105  | -0,059  | 0,072   | -0,836  | -0,136  | -0,106  | 0,748       |
| ICAP     | 0,293   | -0,847  | -0,210  | -0,054  | 0,035   | -0,029  | 0,853       |
| IEND     | 0,026   | 0,790   | -0,259  | -0,051  | -0,094  | 0,183   | 0,736       |
| IRPG     | 0,129   | -0,084  | 0,044   | -0,126  | 0,073   | -0,834  | 0,743       |
| IIMR     | -0,082  | -0,118  | -0,063  | -0,205  | 0,222   | 0,796   | 0,750       |
| ISIN     | -0,051  | -0,046  | -0,047  | -0,096  | -0,818  | -0,071  | 0,690       |
| ICOL     | -0,045  | 0,144   | 0,022   | 0,071   | 0,812   | 0,028   | 0,688       |
| IDAD     | 0,938   | -0,125  | 0,207   | -0,023  | -0,009  | -0,083  | 0,945       |
| ILPG     | 0,973   | -0,092  | 0,145   | 0,061   | 0,024   | -0,090  | 0,990       |
| IRFL     | -0,060  | 0,024   | 0,045   | 0,895   | 0,047   | -0,151  | 0,832       |
| PRFL     | -0,044  | 0,787   | -0,232  | 0,097   | 0,263   | -0,109  | 0,765       |
| IALI     | -0,075  | 0,906   | 0,151   | 0,043   | 0,204   | -0,121  | 0,908       |
| ILCO     | 0,225   | -0,031  | 0,963   | -0,018  | 0,046   | -0,049  | 0,984       |
| ILGE     | 0,239   | -0,039  | 0,955   | -0,012  | 0,036   | -0,068  | 0,977       |
| Variance | 3,0046  | 2,8631  | 2,1299  | 1,5934  | 1,5280  | 1,4589  | 12,5778     |
| % Var    | 0,200   | 0,191   | 0,142   | 0,106   | 0,102   | 0,097   | 0,839       |

Observe que neste exemplo foram “extraídos” seis fatores. A primeira coluna tem os nomes de cada variável, as seis colunas seguintes têm as cargas fatoriais (veja 15) e a última coluna apresenta as comunalidades (veja 16). As últimas duas linhas apresentam os autovalores e a variabilidade explicada por cada um, no final é apresentada a soma, ou seja variabilidade total explicada, neste caso igual a 0,839.

##### Composição dos Scores Fatoriais (Fatores):

| Variable | Factor1 | Factor2 | Factor3 | Factor4 | Factor5 | Factor6 |
|----------|---------|---------|---------|---------|---------|---------|
| ICOM     | 0,355   | 0,049   | -0,062  | -0,003  | -0,022  | 0,034   |
| ICOA     | -0,047  | 0,013   | 0,032   | -0,536  | 0,012   | -0,126  |
| ICAP     | 0,074   | -0,307  | -0,168  | -0,018  | 0,125   | -0,047  |
| IEND     | 0,134   | 0,324   | -0,120  | -0,050  | -0,146  | 0,150   |
| IRPG     | -0,033  | -0,052  | -0,062  | -0,144  | 0,144   | -0,618  |
| IIMR     | 0,035   | -0,040  | 0,018   | -0,105  | 0,129   | 0,538   |
| ISIN     | -0,005  | 0,062   | 0,013   | 0,027   | -0,558  | 0,003   |
| ICOL     | -0,025  | -0,033  | -0,016  | -0,049  | 0,554   | -0,044  |
| IDAD     | 0,332   | 0,032   | -0,029  | -0,028  | -0,013  | 0,030   |
| ILPG     | 0,355   | 0,040   | -0,068  | 0,019   | 0,001   | 0,027   |
| IRFL     | -0,061  | -0,042  | 0,042   | 0,576   | -0,059  | -0,059  |
| PRFL     | 0,065   | 0,264   | -0,130  | -0,010  | 0,118   | -0,091  |
| IALI     | 0,010   | 0,319   | 0,082   | -0,034  | 0,053   | -0,069  |
| ILCO     | -0,054  | 0,010   | 0,482   | 0,000   | -0,008  | 0,046   |
| ILGE     | -0,050  | 0,008   | 0,474   | 0,004   | -0,014  | 0,034   |

A figura acima apresenta a “receita” para compor os Fatores, cada Fator é escrito como um coeficiente vezes a variável padronizada (veja 18), ou seja os valores apresentados em cada coluna de um determinado Fator  $j$  são os valores de  $b_j$  para cada variável de acordo com a equação 18.

## 5 Análise de Agrupamentos (Cluster Analysis)

Objetivo dessa técnica é agrupar objetos (itens) semelhantes segundo suas características (variáveis)

Procedimento:

- Padronizar as variáveis:  $z_i = (x_i - \bar{x}_i) / s_i$
- Obter uma matriz de distâncias
- Aplicar um método de aglomeração, por exemplo os métodos hierárquicos
  - Ligação Simples (vizinho mais próximo), ou ainda Single Linkage
  - Ligação Completa (vizinho mais distante), ou ainda Complete Linkage
  - Ligação Média (distância média), ou ainda Average linkage
- Considerar a menor distância para formar um grupo.
- Recalcular a matriz de distâncias
- No final esboçar um Dendrograma

### 5.1 Análise de Agrupamentos no Minitab:

No menu do Minitab:

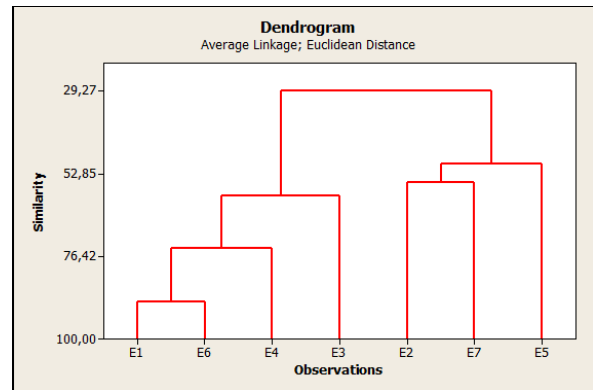
Stat ► Multivariate ► Cluster Observations

Escolha o método hierárquico; a medida de distância; escolha padronizar as variáveis; show dendrogram; Customize: Case Labels.

No Minitab é possível ver o Agglomeration Schedule (Amalgamation Steps):

| Standardized Variables, Euclidean Distance, Average Linkage |                    |                  |                |                 |             |                               |
|-------------------------------------------------------------|--------------------|------------------|----------------|-----------------|-------------|-------------------------------|
| Amalgamation Steps                                          |                    |                  |                |                 |             |                               |
| Step                                                        | Number of clusters | Similarity level | Distance level | Clusters joined | New cluster | Number of obs. in new cluster |
| 1                                                           | 6                  | 89,3891          | 0,42796        | 1               | 6           | 2                             |
| 2                                                           | 5                  | 74,2684          | 1,03781        | 1               | 4           | 3                             |
| 3                                                           | 4                  | 59,2646          | 1,64294        | 1               | 3           | 4                             |
| 4                                                           | 3                  | 55,4507          | 1,79676        | 2               | 2           | 2                             |
| 5                                                           | 2                  | 50,0920          | 2,01289        | 2               | 5           | 3                             |
| 6                                                           | 1                  | 29,2746          | 2,85250        | 1               | 2           | 7                             |

O Minitab também fornece o Dendrograma



## 6 Análise Discriminante

Utilizada para determinar quais variáveis discriminam entre dois ou mais grupos já conhecidos.

**A variável dependente é categórica** e indica os grupos já conhecidos. As variáveis independentes podem ser métricas ou categóricas, estas deverão ter poder de discriminar os indivíduos entre dois ou mais grupos.

O primeiro passo é compor para cada grupo ( $g_i$ ), as funções discriminantes ( $d_i$ ):

$$\text{Para } g_1: d_1 = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \dots + \beta_{1n}x_n$$

$$\text{Para } g_2: d_2 = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \dots + \beta_{2n}x_n$$

⋮

$$\text{Para } g_k: d_k = \beta_{k0} + \beta_{k1}x_1 + \beta_{k2}x_2 + \dots + \beta_{kn}x_n$$

Para classificar um indivíduo num determinado grupo devemos substituir as variáveis  $x_1, x_2, \dots, x_n$  em todas as funções discriminantes  $d_i$ . A função que obtiver o maior valor indica a classificação do indivíduo para o grupo correspondente.

Os coeficientes  $\beta$ 's podem ser determinados pelo SPSS ou pelo Minitab.

### 6.1 Análise Discriminante no Minitab:

No menu do Minitab:

Stat ► Multivariate ► Discriminant Analysis

O Minitab fornece os valores dos  $\beta$ 's no quadro Linear Discriminant Function for Groups.

| Linear Discriminant Function for Groups |         |         |         |
|-----------------------------------------|---------|---------|---------|
|                                         | 1       | 2       | 3       |
| Constant                                | -26,888 | -43,228 | -35,491 |
| Vendas                                  | 0,000   | 0,000   | 0,000   |
| Rentabilidade                           | 0,409   | 0,651   | 0,682   |
| Liquidez                                | 23,688  | 43,778  | 22,095  |
| Endividamento                           | 0,335   | 0,462   | 0,583   |

Saída Minitab para Análise Discriminante:  
Valores de  $\beta$ 's para as funções discriminantes.

Para verificar o percentual de acerto na classificação o Minitab fornece um resumo da classificação. Neste resumo podemos observar as categorias verdadeiras (coluna: True Group), bem como, as Classificações feitas de acordo com as funções discriminantes (Put into Group). Com isto podemos verificar a proporção de acertos (0,952 no exemplo abaixo).

| Summary of classification                                  |            |       |       |
|------------------------------------------------------------|------------|-------|-------|
| Put into Group                                             | True Group |       |       |
|                                                            | 1          | 2     | 3     |
| 1                                                          | 6          | 0     | 0     |
| 2                                                          | 0          | 10    | 0     |
| 3                                                          | 0          | 1     | 4     |
| Total N                                                    | 6          | 11    | 4     |
| N correct                                                  | 6          | 10    | 4     |
| Proportion                                                 | 1,000      | 0,909 | 1,000 |
| N = 21      N Correct = 20      Proportion Correct = 0,952 |            |       |       |

Saída Minitab para Análise Discriminante:  
Tabela de Classificação.

A **precisão dos coeficientes  $\beta$ 's** deve ser verificada pois no exemplo o Minitab forneceu 0,000 (veja os coeficientes da segunda linha) mas como a variável Vendas assume valores altos devemos saber o valor de  $\beta$  com mais casas decimais. Neste caso, opção é gravar os coeficientes numa coluna e modificar o formato da coluna. Selecione uma célula da coluna que quer modificar o formato e em seguida selecione no menu do Minitab:

Editor ► Format Column ► Numeric ►  
Category: ( • ) Exponential.

No exemplo acima a alternativa é gravar os coeficientes  $\beta$ 's nas colunas disc1, disc2 e disc3:

| ↓ | C7          | C8          | C9          |
|---|-------------|-------------|-------------|
|   | disc1       | disc2       | disc3       |
| 1 | -2,6888E+01 | -4,3228E+01 | -3,5491E+01 |
| 2 | 3,1581E-06  | 1,2579E-06  | 9,7455E-07  |
| 3 | 4,0853E-01  | 6,5124E-01  | 6,8182E-01  |
| 4 | 2,3688E+01  | 4,3778E+01  | 2,2095E+01  |
| 5 | 3,3481E-01  | 4,6246E-01  | 5,8346E-01  |
| 6 |             |             |             |

## 6.2 Análise Discriminante no SPSS:

No SPSS: Analyze ► Classify ► Discriminant

Os  $\beta$ 's são fornecidos no quadro "Classification Function Coefficients". O SPSS denomina as funções discriminantes de "Funções discriminantes lineares de Fisher". Os coeficientes pequenos são apresentados no formato científico: 3.2E-006 = 0,0000032.

### Classification Function Coefficients

|               | grupo    |          |          |
|---------------|----------|----------|----------|
|               | 1        | 2        | 3        |
| vendas        | 3.2E-006 | 1.2E-006 | 9.7E-007 |
| rentabilidade | .408     | .651     | .682     |
| liquidez      | 23.681   | 43.774   | 22.094   |
| endividamento | .335     | .463     | .584     |
| (Constant)    | -27.969  | -44.315  | -36.584  |

Fisher's linear discriminant functions

O resultado da classificação é apresentado em colunas (no Minitab é em linhas)

### Classification Results<sup>a</sup>

|          |       | gru po | Predicted Group Membership |      |       | Total |
|----------|-------|--------|----------------------------|------|-------|-------|
|          |       |        | 1                          | 2    | 3     |       |
| Original | Count | 1      | 6                          | 0    | 0     | 6     |
|          |       | 2      | 0                          | 10   | 1     | 11    |
|          |       | 3      | 0                          | 0    | 4     | 4     |
|          | %     | 1      | 100,0                      | ,0   | ,0    | 100,0 |
|          |       | 2      | ,0                         | 90,9 | 9,1   | 100,0 |
|          |       | 3      | ,0                         | ,0   | 100,0 | 100,0 |

a. 95,2% of original grouped cases correctly classified.

