



Centro de Informática

★ • • • • • • • • • • UFPE

Aprendizagem de Dados Simbólicos e/ou Numéricos

Francisco de A.T. de Carvalho

Estatísticas Descritivas: Dados Usuais 1/

➤ Notação

- Y : variável categórica (nominal ou ordinal) ou quantitativa discreta
- $E = \{1, \dots, n\}$: conjunto de n itens
- $Y(k)$, $k \in E$: valor de Y em k
- $y = (Y(1), \dots, Y(n))$: vetor de valores de Y para os n elementos

Estatísticas Descritivas: Dados Usuais 2/

➤ Notação

➤ ξ_1, \dots, ξ_t conjunto dos t diferentes valores tomados por Y em E ($t \leq n$)

➤ Frequência observada

➤ $n_i := \#\{k \in E \mid Y(k) = \xi_i\}$, $i = 1, \dots, t$

➤ Distribuição de Frequências

➤ Lista de pares $(\xi_1, n_1), \dots, (\xi_t, n_t)$

Estatísticas Descritivas: Dados Usuais 3/

➤ Frequência Relativa

➤ $f_i = n_i / n, i = 1, \dots, t$

➤ Distribuição de Frequências: Representação

➤ Variáveis nominais (valores não ordenados):

➤ *setores*, cujas áreas são proporcionais a n_i

Estatísticas Descritivas: Dados Usuais 4/

➤ Distribuição de Frequências: Representação

➤ Variáveis ordinais ou quantitativas discretas:

➤ *diagrama de barras*, onde cada par (ξ_i, n_i) é localizado na coordenada ξ_i do eixo horizontal e cuja altura é proporcional a n_i

➤ Variáveis quantitativas contínuas:

➤ muitos valores distintos ξ_i agrupados em classes de valores

Estatísticas Descritivas: Dados Usuais 5/

➤ Distribuição de Frequências: Representação

➤ Variáveis quantitativas contínuas:

➤ Seja a ordenação dos ξ_i em ordem crescente:

$$\xi_{(1)} \leq \dots \leq \xi_{(t)}$$

➤ Seja uma partição $I = \{I_1, \dots, I_m\}$ do intervalo $[\xi_{(1)}, \xi_{(t)}]$, onde

➤ I_1, \dots, I_m são intervalos contíguos de $[\xi_{(1)}, \xi_{(t)}]$

Estatísticas Descritivas: Dados Usuais 6/

➤ Distribuição de Freqüências: Representação

➤ Variáveis quantitativas contínuas:

➤ Seja $m_j = \#\{k \in E \mid Y(k) \in I_j\}$, $j = 1, \dots, m$

➤ Distribuição de Freqüências

➤ Lista de pares $(I_1, m_1), \dots, (I_j, m_j)$

Estatísticas Descritivas: Dados Usuais 7/

➤ Distribuição de Frequências: Representação

➤ Variáveis quantitativas contínuas:

➤ histograma:

➤ é uma representação gráfica em que cada par (I_j, m_j) , $j = 1, \dots, m$, é representada por uma barra vertical cuja base é o intervalo I_j e cuja área é proporcional a m_j

Estatísticas Descritivas: Dados Usuais 8/

➤ Distribuição de Frequências: Representação

➤ Variáveis quantitativas contínuas:

➤ Função de distribuição empírica F_y

$$F_y(\xi) = \frac{1}{n} \#\{k \in E \mid Y(k) \leq \xi\}$$

➤ Distribuição de Frequências Acumuladas

➤ Lista de pares $(\xi_1, F_y(\xi_1)), \dots, (\xi_t, F_y(\xi_t))$

Estatísticas Descritivas: Dados Usuais 9/

➤ Estatísticas

➤ Média amostral

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

➤ Moda – Valor + freqüente

➤ Variância

$$s_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

Estatísticas Descritivas: Dados Usuais 10/

➤ Estatísticas

➤ Desvio-padrão

$$s_y = \sqrt{s_y^2}$$

➤ Coeficiente de Variação

$$\left(\frac{s_y}{\bar{y}} \right) \times 100$$

Estatísticas Descritivas: Dados Usuais 11/

➤ Frequência Observada

➤ Frequência observada O_y é uma função

$$O_y : \mathcal{R} \rightarrow N_0$$

tal que

$$O_y(\xi) = \#\{k \in E \mid Y(k) = \xi\}$$

Estatísticas Descritivas: Dados Usuais 12/

➤ Frequência Observada

➤ Exemplo: $y = (4, 7, 6, 2, 6)$

$$O_y(\xi) = \begin{cases} 1 & \text{se } \xi \in \{2, 4, 7\} \\ 2, & \text{se } \xi = 6 \\ 0, & \text{nos outros casos} \end{cases}$$

Estatísticas Descritivas: Dados Usuais 13/

➤ Frequência Observada

➤ Número de exemplos

$$n_i = O_y(\xi_i), i = 1, \dots, l$$

➤ Frequencia acumulada

$$F_y(\xi) = \frac{1}{n} \sum_{\xi_j \leq \xi} O_y(\xi_j) \xi_j$$

Estatísticas Descritivas: Dados Usuais 14/

➤ Frequência Observada

➤ média

$$\bar{y} = \frac{1}{n} \sum_{j=1}^l O_y(\xi_j) \xi_j$$

➤ Variância

$$s_y^2 = \frac{1}{n} \sum_{j=1}^l O_y(\xi_j) (\xi_j - \bar{y})^2$$

Estatísticas Descritivas: Dados Simbólicos 1/

➤ Tabela de Dados

➤ Variáveis simbólicas: Y_1, \dots, Y_p

➤ Itens: $E = \{1, \dots, n\}$

➤ Cada célula (k,j) da tabela de dados contém o valor de Y_j para k

$$\xi_{kj} = Y_j(k), \quad k = 1, \dots, n \quad j = 1, \dots, p$$

Estatísticas Descritivas: Dados Simbólicos 2/

➤ Tabela de Dados

➤ Descrição simbólica de $k \in E$ (linha da tabela)

$$d_k = (\xi_{k1}, \dots, \xi_{kp})', \quad k = 1, \dots, n$$

➤ Extensão virtual de d_1

$$\text{vir}(d_1) = \{ x \mid x \in \xi_{k1} \times \dots \times \xi_{kp} \}$$

Estatísticas Descritivas: Dados Simbólicos 3/

Exemplo

	Y_1	Y_2
1	{a, b}	{3}
2	{b, c}	{2, 3}
3	{b}	{1, 2}
4	{c}	{1, 3}
5	{a}	{1, 3}

Estatísticas Descritivas: Dados Simbólicos 4/

➤ Exemplo

- Variáveis Simbólicas: Y_1 e Y_2
- D_1 e D_2 : domínios de Y_1 e Y_2
- $E = \{1,2,3,4,5\}$: conjunto de itens
- $B_1 = P(D_1)$ e $B_2 = P(D_2)$
- Descrição da linha 1:

$$d_1 = (\{ a, b \}, \{ 3 \})'$$

Estatísticas Descritivas: Dados Simbólicos 5/

➤ Exemplo

➤ Descrição da linha 1:

$$d_1 = (\{ a, b \}, \{ 3 \})'$$

➤ Extensão virtual de d_1

$$\begin{aligned} \text{vir}(d_1) &= \{ x \in \{ a, b \} \times \{ 3 \} \} = \\ &= \{ (a, 3)', (b, 3)' \} \end{aligned}$$

Estatísticas Descritivas: Dados Simbólicos 6/

➡ Distribuição de Frequências: variáveis multivaloradas (categóricas ou quantitativas)

➡ Frequência Observada

➡ Dada uma variável multivalorada Z , a *frequência observada* O_Z é uma função que associa a cada valor $\xi \in D_Z$ a soma de $\pi(\xi; k)$ para $k \in E$.

Estatísticas Descritivas: Dados Simbólicos 8/

➡ Distribuição de Frequências: variáveis multivaloradas (categóricas ou quantitativas)

➡ Frequência Observada

$$\sum_{\xi \in D_Z} O_Z(\xi) = \sum_{k \in E} \sum_{\xi \in D_Z} \frac{\#\{x \in \text{vir}(d_k) \mid x_{[Z]} = \xi\}}{\#\text{vir}(d_k)} = n$$

Estatísticas Descritivas: Dados Simbólicos 9/

➤ Distribuição de Frequências: variáveis multivaloradas (categóricas ou quantitativas)

➤ Distribuição da Frequência Observada

➤ lista de pares

$$(\xi, O_Z(\xi)), \xi \in D_Z$$

➤ Frequência Observada relativa

$$freq_Z(\xi) = \frac{1}{n} O_Z(\xi)$$

Estatísticas Descritivas: Dados Simbólicos 10/

Exemplo 1

$$\text{O}_{Y_1}(\mathbf{a}) = \sum_{k \in E} \left(\frac{\#\{x \in \text{vir}(d_k) \mid x = a\}}{|\text{vir}(d_k)|} \right) = \frac{1}{1} + \frac{0}{2} + \frac{0}{1} + \frac{2}{2} = 2$$

$$\text{O}_{Y_1}(\mathbf{b}) = \sum_{k \in E} \left(\frac{\#\{x \in \text{vir}(d_k) \mid x = b\}}{|\text{vir}(d_k)|} \right) = \frac{0}{1} + \frac{1}{2} + \frac{1}{1} + \frac{0}{2} = 1,5$$

Estatísticas Descritivas: Dados Simbólicos 11/

Exemplo 1

$$\mathbf{O}_{Y_1}(c) = \sum_{k \in E} \left(\frac{\#\{x \in \text{vir}(d_k) \mid x = c\}}{|\text{vir}(d_k)|} \right) = \frac{0}{1} + \frac{1}{2} + \frac{0}{1} + \frac{0}{2} = 0,5$$

$$f_{Y_1}(\xi) = \begin{array}{ll} \frac{2}{4} = 0,500 & \text{se } \xi = a \\ \frac{(1,5)}{4} = 0,375 & \text{se } \xi = b \\ \frac{(0,5)}{4} = 0,125 & \text{se } \xi = c \end{array}$$

Estatísticas Descritivas: Dados Simbólicos 12/

➤ Outras medidas para variáveis multivaloradas numéricas

➤ Distribuição acumulada

$$F_Z(\xi) = \frac{1}{n} \sum_{\xi_j \leq \xi} O_Z(\xi_j)$$

➤ Média

$$\bar{z} = \frac{1}{n} \sum_{j=1}^l O_Z(\xi_j) \xi_j$$

Estatísticas Descritivas: Dados Simbólicos 13/

➡ Outras medidas para variáveis multivaloradas numéricas

➡ Desvio padrão

$$s_Z = \sqrt{\frac{1}{n} \sum_{j=1}^l O_Z(\xi_j) (\xi_j - \bar{z})^2}$$

Estatísticas Descritivas: Dados Simbólicos 14/

Exemplo 2

Variável Y_2 : $D_2 = \{1, 2, 3\}$

$O_{Y_2}(1) = 0,5$ $O_{Y_2}(2) = 2,0$ $O_{Y_2}(3) = 1,5$

função de densidade (f)

$$f_{Y_1}(e_i) = \begin{array}{ll} 0,125 & \text{se } e_i = 1 \\ 0,500 & \text{se } e_i = 2 \\ 0,375 & \text{se } e_i = 3 \end{array}$$

Função de Distribuição (F)

$$F_{Y_1}(e_i) = \begin{array}{ll} 0 & \text{se } e_i < 1 \\ 0,125 & \text{se } 1 \leq e_i < 2 \\ 0,625 & \text{se } 2 \leq e_i < 3 \\ 1 & \text{se } e_i \geq 3 \end{array}$$

Estatísticas Descritivas: Dados Simbólicos 15/

Exemplo 2

$$\bar{y}_2 = \frac{1}{4} \{(0,5 \times 1) + (2 \times 2) + (1,5 \times 3)\} = 2,25$$

$$S_{Y_2}^2 = \frac{1}{(4-1)} \{[(1-2,25)^2 \times 0,5] + (2-2,25)^2 \times 2,0\} + \\ + (3-2,25)^2 \times 1,5\} = 0,583$$

$$S_{Y_2} = \sqrt{0,604} = 0,764$$

Estatísticas Descritivas: Dados Simbólicos 16/

➤ Variável intervalo

➤ Seja Z uma variável intervalo, para cada $k \in E$ temos:

$$➤ Z(k) = [z_k, \bar{z}_k]$$

➤ Hipóteses:

➤ (i) $k \in E$ é selecionado com mesma probabilidade $(1/n)$

➤ (ii) Os valores dos $x_{[z]}$ para $x \in \text{vir}(d_k)$ são uniformemente distribuídos no intervalo $Z(k) = [z_k, \bar{z}_k]$

Estatísticas Descritivas: Dados Simbólicos 17/

➡ Variável intervalo

$$\Pr(x_{[z]} \leq e \mid x \in \text{vir}(d_k)) = \begin{cases} 0 & \text{se } e < \underline{z}_k \\ \frac{e - \underline{z}_k}{\bar{z}_k - \underline{z}_k} & \text{se } \bar{z}_k \leq e < \underline{z}_k \\ 1 & \text{se } e \geq \underline{z}_k \end{cases}$$

Estatísticas Descritivas: Dados Simbólicos 18/

➤ Variável intervalo

➤ Definição: A *função de distribuição empírica* de Z , expressa por F_Z , é considerada como

➤ uma mistura de n distribuições uniformes definidas nos intervalos $Z(k) = [z_k, \bar{z}_k]$, para $k \in E$.

➤ A partir da expressão $\Pr(x_{[z]} \leq e \mid x \in \text{vir}(d_k))$, obtemos

Estatísticas Descritivas: Dados Simbólicos 19/

➡ Variável intervalo

$$F_Z(e) = \frac{1}{n} \left[\sum_{e \in Z(k)} \left(\frac{e - \underline{z}_k}{\bar{z}_k - \underline{z}_k} \right) \right] + \frac{\#\{k \mid e \geq \bar{z}_k\}}{n}$$

➡ Derivando a expressão anterior em e , obtemos a **função de densidade empírica** de Z

$$f_Z(e) = \frac{1}{n} \sum_{e \in Z(k)} \left(\frac{1}{l(Z(k))} \right), \text{ onde } l(Z(k)) = \bar{z}_k - \underline{z}_k$$

Estatísticas Descritivas: Dados Simbólicos 20/

➡ Variável intervalo

➡ Média Amostral:

$$\bar{z} = \int_{-\infty}^{+\infty} e f_Z(e) de = \dots = \frac{1}{n} \sum_{k \in E} \frac{z_k + \bar{z}_k}{2}.$$

➡ Variância Amostral:

$$S_Z^2 = \int_{-\infty}^{+\infty} (e - \bar{z})^2 f_Z(e) de$$

Estatísticas Descritivas: Dados Simbólicos 21/

➡ Variável intervalo

➡ Variância Amostral:

$$S_Z^2 = \frac{1}{3n} \sum_{k \in E} (\bar{z}_k^2 + \bar{z}_k \underline{z}_k + \underline{z}_k^2) - \frac{1}{4n^2} \left[\sum_{k \in E} (\bar{z}_k + \underline{z}_k) \right]^2$$

Estatísticas Descritivas: Dados Simbólicos 22/

➤ Variável intervalo

➤ Exemplo: $E = \{1, \dots, 8\}$

➤ $Z(E) = \{[0,2]; [1,3]; [1.5,2.5]; [2,4]; [3.5,5]; [4.5,5.5]; [5,7]; [6.5,7.5]\}$

Z(E)	[0,2]	[1,3]	[1.5,2.5]	[2,4]	[3.5,5]	[4.5,5.5]	[5,7]	[6.5,7.5]
$\frac{\underline{z}_k + \bar{z}_k}{2}$	1	2	2	3	4,25	5	6	7

Média Amostral = 3,781

Variância Amostral = 4,182

Estatísticas Descritivas: Dados Simbólicos 23/

➤ Variável intervalo

➤ Histograma

- Seja $I = [\min \{z_k \mid k \in E\}, \max \{\bar{z}_k \mid k \in E\}]$ um intervalo que contém os valores de $Z(k)$.
- Considere uma partição de I em m subintervalos disjuntos $I_j = [u_{j-1}, u_j[$ para $j=1, \dots, m-1$ e $I_m = [u_{m-1}, u_m]$ para $j=m$.

Estatísticas Descritivas: Dados Simbólicos 24/

➤ Variável intervalo

➤ Histograma

➤ Então, o *Histograma* associado a partição $\{I_1, \dots, I_m\}$ é uma representação gráfica da distribuição de frequência $\{(I_j, p_j)\}$ da variável Z , onde p_j representa a probabilidade da classe I_j .

$$p_j = \frac{1}{n} \sum_{k \in E} \frac{l(Z(k) \cap I_j)}{l(Z(k))}.$$

Estatísticas Descritivas: Dados Simbólicos 25/

➤ Variável intervalo

➤ Exemplo: $E = \{1, \dots, 8\}$

➤ (1) $Z(E) = \{[0, 2]; [1, 3]; [1.5, 2.5]; [2, 4]; [3.5, 5]; [4.5, 5.5]; [5, 7]; [6.5, 7.5]\}$

➤ (2) $I = \{0.0, 7.5\}$

➤ (3) Considere m partições de tamanho $(7.5 - 0)/8 = 0.938$

Estatísticas Descritivas: Dados Simbólicos 26/

➤ Variável intervalo

➤ Exemplo:

$$➤ I_1 = [0, 0.938[; I_2 = [0.938, 1.875[; \dots ; I_8 = [6.563, 7.5]$$

➤ (4) Cálculo dos p_j 's.

$$➤ I_2 = [0.938, 1.875[$$

$$p_2 = \frac{1}{n} \left\{ \frac{l([0,2] \cap [0.938,1.875[)}{l([0,2])} + \frac{l([1,3] \cap [0.938,1.875[)}{l([1,3])} + \frac{l([1.5,2.5] \cap [0.938,1.875[)}{l([1.5,2.5])} \right\} =$$

Estatísticas Descritivas: Dados Simbólicos 27/

➡ Variável intervalo

➡ Exemplo:

➡ (4) Cálculo dos p_j 's.

$$p_2 = \frac{1}{8} \left\{ \frac{0.938}{2} + \frac{0.875}{2} + \frac{0.625}{1} \right\} = \frac{1}{8} \{0.469\} \cong 0.06$$

Estatísticas Descritivas: Dados Simbólicos 28/

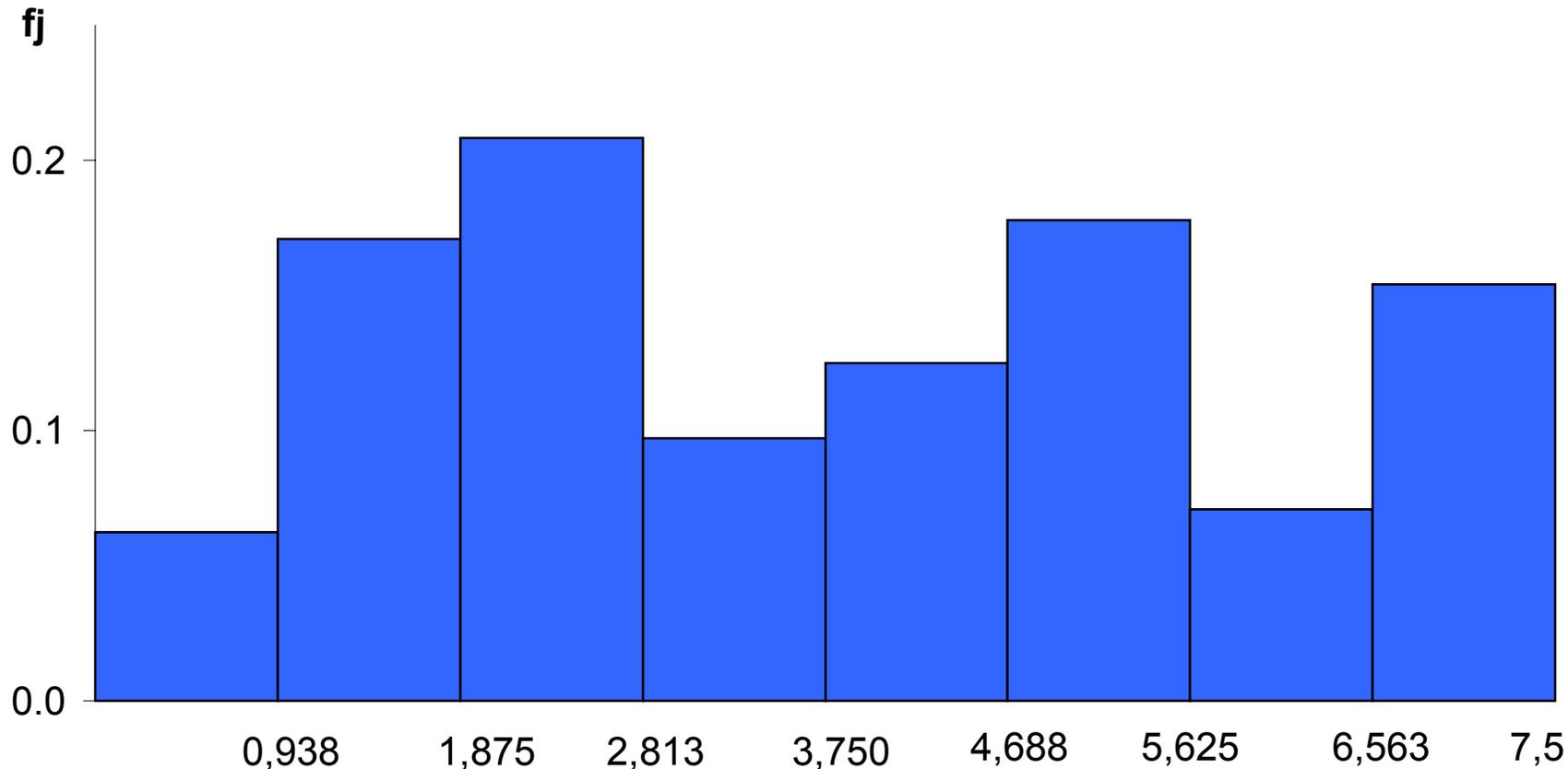
→ $Z(E) = \{[0,2]; [1,3]; [1.5,2.5]; [2,4]; [3.5,5]; [4.5,5.5]; [5,7]; [6.5,7.5]\}$

I	Inf.	Sup	O _j	p _j	pm _j	Média	Var.	f _j
I1	0.000	0.938	0.469	0.06	0.47	0.03	0.64	0.06
I2	0.938	1.875	1.281	0.16	1.41	0.23	0.91	0.17
I3	1.875	2.813	1.563	0.20	2.34	0.46	0.41	0.21
I4	2.813	3.750	0.729	0.09	3.28	0.30	0.02	0.10
I5	3.750	4.688	0.938	0.12	4.22	0.49	0.02	0.13
I6	4.688	5.625	1.333	0.17	5.16	0.86	0.31	0.18
I7	5.625	6.563	0.531	0.07	6.09	0.40	0.35	0.07
I8	6.563	7.500	1.156	0.14	7.03	1.02	1.52	0.15
			8.0	1.0	3.784		4.192	

Estatísticas Descritivas: Dados Simbólicos 29/

Histograma para variável Z

Variável intervalo



Estatísticas Descritivas: Dados Simbólicos 30/

➤ Variável categórica modal

➤ Variável Z , objeto u

➤ valores possíveis: ξ_k

➤ com frequências p_{uk} , $p_{uk} \geq 0$ e $\sum_k p_{uk} = 1$

➤ Frequência observada

$$O_Z(\xi_k) = \sum_{u \in E} \pi_Z(\xi_k; u) = \sum_{u \in E} p_{uk}$$

Estatísticas Descritivas: Dados Simbólicos 31/

➡ Variável categórica modal

➡ Frequência observada

➡ Propriedades

$$\sum_{k=1}^K O_Z(\xi_k) = \sum_{k=1}^K \sum_{u \in E} p_{uk} = \sum_{u \in E} \sum_{k=1}^K p_{uk} = n$$

➡ Frequência relativa

$$freq_Z(\xi_k) = \frac{O_Z(\xi_k)}{n}$$

Estatísticas Descritivas: Dados Simbólicos 32/

➤ Variável categórica modal

➤ Exemplo

➤ Z : fonte de energia

➤ D_Z : {gas, eletricidade, outros}

Estatísticas Descritivas: Dados Simbólicos 33/

➡ Variável categórica modal

➡ Exemplo

➡ $O_Z(\text{gas}) = 0.87 + \dots + 0.76 = 3,17$

➡ $O_Z(\text{elet}) = 0.05 + \dots + 0.11 = 0.35$

➡ $O_Z(\text{outros}) = 0.08 + \dots + 0.13 = 0.48$

<i>gas</i>	<i>elet</i>	<i>outros</i>
0.87	0.05	0.08
0.71	0.10	0.19
0.83	0.09	0.08
0.76	0.11	0.13

$$\sum_{u \in E} O_Z(\xi_k) = 4$$

Estatísticas Descritivas: Dados Simbólicos 34/

➡ Variável categórica modal

➡ Exemplo

➡ $freq_Z(\text{gas}) = 3,17/4 = 0.7925$

➡ $freq_Z(\text{elet}) = 0.35/4 = 0.0875$

➡ $freq_Z(\text{outros}) = 0.48/4 = 0.12$

<i>gas</i>	<i>elet</i>	<i>outros</i>
0.87	0.05	0.08
0.71	0.10	0.19
0.83	0.09	0.08
0.76	0.11	0.13

$$\sum_{u \in E} freq_Z(\xi_k) = 1$$

Estatísticas Descritivas: Dados Simbólicos 35/

➤ Variável intervalo modal

➤ Variável Z

➤ Objeto u , $u = 1, \dots, n$

➤ $Z(u) = \xi_{uk} = [a_{uk}, b_{uk})$ com probabilidade p_{uk} , $k=1, \dots, s_u$

➤ $I = [\min_{k,u \in E} a_{uk}, \max_{k,u \in E} b_{uk}]$: intervalo que “cobre” os valores observados em Z

Estatísticas Descritivas: Dados Simbólicos 36/

➤ Variável intervalo modal

➤ Seja uma partição de I em r sub-intervalos

$$➤ I_g = [\xi_{g-1}, \xi_g), g = 1, \dots, r-1 \text{ e } I_r = [\xi_{r-1}, \xi_r)$$

➤ Frequência observada

$$O_Z(g) = \sum_{u \in E} \pi_Z(g; u)$$

Estatísticas Descritivas: Dados Simbólicos 37/

➡ Variável intervalo modal

➡ Frequência observada

$$\pi_Z(g; u) = \sum_{k \in Z(g)} \frac{|Z(k; u) \cap I_g|}{|Z(k; u)|} p_{uk}$$

➡ $Z(g)$: os intervalos $Z(k; u) = [a_{uk}, b_{uk})$ cuja intersecção com I_g é não vazia, para um dado u

Estatísticas Descritivas: Dados Simbólicos 38/

➤ Variável intervalo modal

➤ Cada termo representa a porção de $Z(k;u)$ que é coberta por I_g , isto é, a proporção de p_{uk} que “pertence” a I_g

➤ Frequência relativa

$$freq_Z(g) = \frac{O_Z(g)}{n}$$

➤ Propriedades

$$\sum_{g=1}^r O_Z(g) = n \qquad \sum_{g=1}^r freq_Z(g) = n$$

Estatísticas Descritivas: Dados Simbólicos 1/

➡ Entrada: Matriz de Dados Simbólicos

Table					
	specific gravit	freezing point	iodine value	saponification	Major Fatty aci
	[0.93 : 0.94]	[-27.00 : -18.00]	[170.00 : 204.00]	[118.00 : 196.00]	Linoleic Acid, Linolenic Acid, Oleic Acid, Palmitic Acid, Myristic Acid
	[0.93 : 0.94]	[-5.00 : -4.00]	[192.00 : 208.00]	[188.00 : 197.00]	Linoleic Acid, Linolenic Acid, Oleic Acid, Palmitic Acid, Searic Acid
	[0.92 : 0.92]	[-6.00 : -1.00]	[99.00 : 113.00]	[189.00 : 198.00]	Linoleic Acid, Oleic Acid, Palmitic Acid, Myristic Acid, Searic Acid
	[0.92 : 0.93]	[-6.00 : -4.00]	[104.00 : 116.00]	[187.00 : 193.00]	Linoleic Acid, Oleic Acid, Palmitic Acid, Searic Acid, Arachid Acid
	[0.92 : 0.92]	[-21.00 : -15.00]	[80.00 : 82.00]	[189.00 : 193.00]	Linoleic Acid, Oleic Acid
	[0.91 : 0.92]	[0.00 : 6.00]	[79.00 : 90.00]	[187.00 : 196.00]	Linoleic Acid, Oleic Acid, Palmitic Acid, Searic Acid
	[0.86 : 0.87]	[30.00 : 38.00]	[40.00 : 48.00]	[190.00 : 199.00]	Oleic Acid, Palmitic Acid, Myristic Acid, Searic Acid, Capric Acid
	[0.86 : 0.86]	[22.00 : 32.00]	[53.00 : 77.00]	[190.00 : 202.00]	Linoleic Acid, Oleic Acid, Palmitic Acid, Myristic Acid, Searic Acid, Lauric Acid

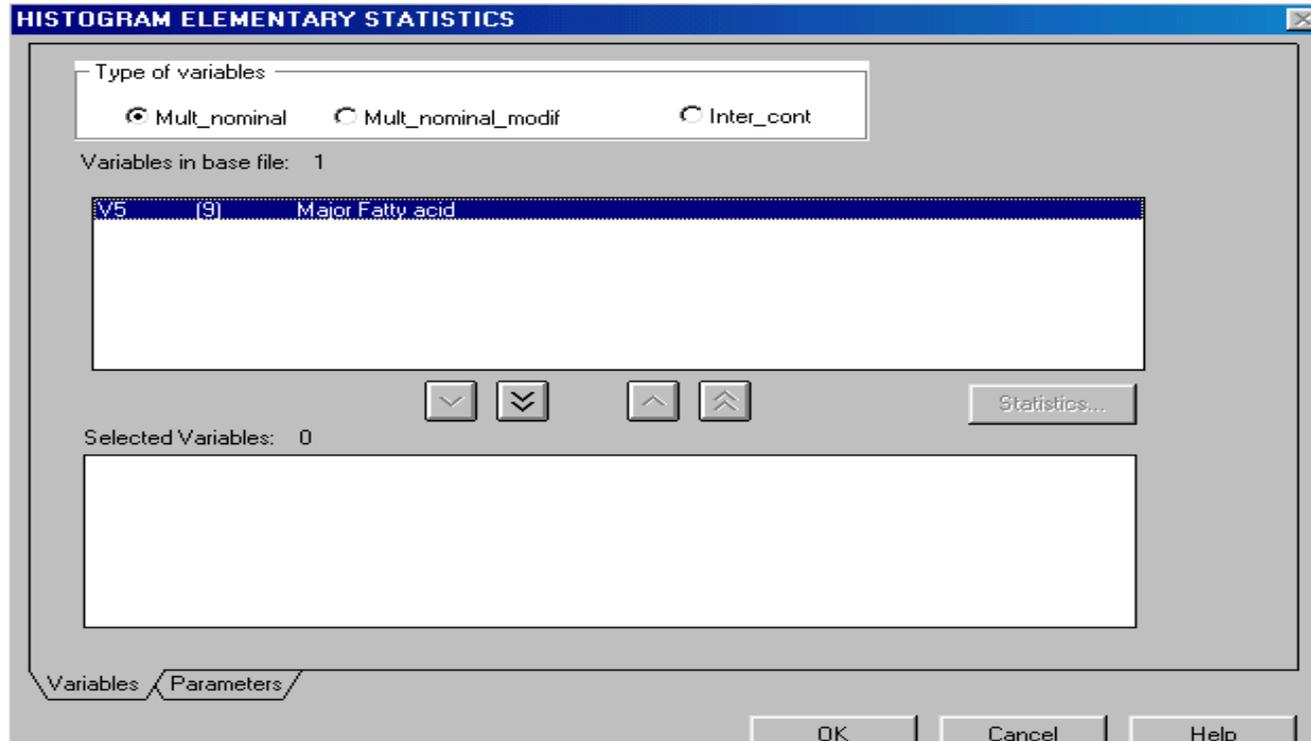
Estatísticas Descritivas: Dados Simbólicos 2/

👉 Comando STAT

The screenshot displays the SODAS software interface. The main window, titled 'VW.FIL', shows a flowchart with three nodes: 'BASE' (a blue downward-pointing triangle), 'STAT' (a grey rectangle with a dashed border and a small '1' in a box to its left), and 'END' (a blue rounded rectangle). The 'STAT' node is highlighted, and its description 'Stat Histogram, Elementary Statistics' is visible. To the left, a 'Methods' panel shows a list of 'Sodas procedures' with 'Stat' selected and highlighted in cyan. Below 'Stat', the text 'Histogram, Elementary Statistics' is also highlighted in cyan. A grid of other procedures is visible in red boxes: SOE, DIV, STAT, DKS, DI, PCM, FDA, TREE, DSD, SDT, DIM, and PYR. The menu bar at the top of the main window includes 'Chaining', 'Model', 'Method', 'Window', and 'Help'.

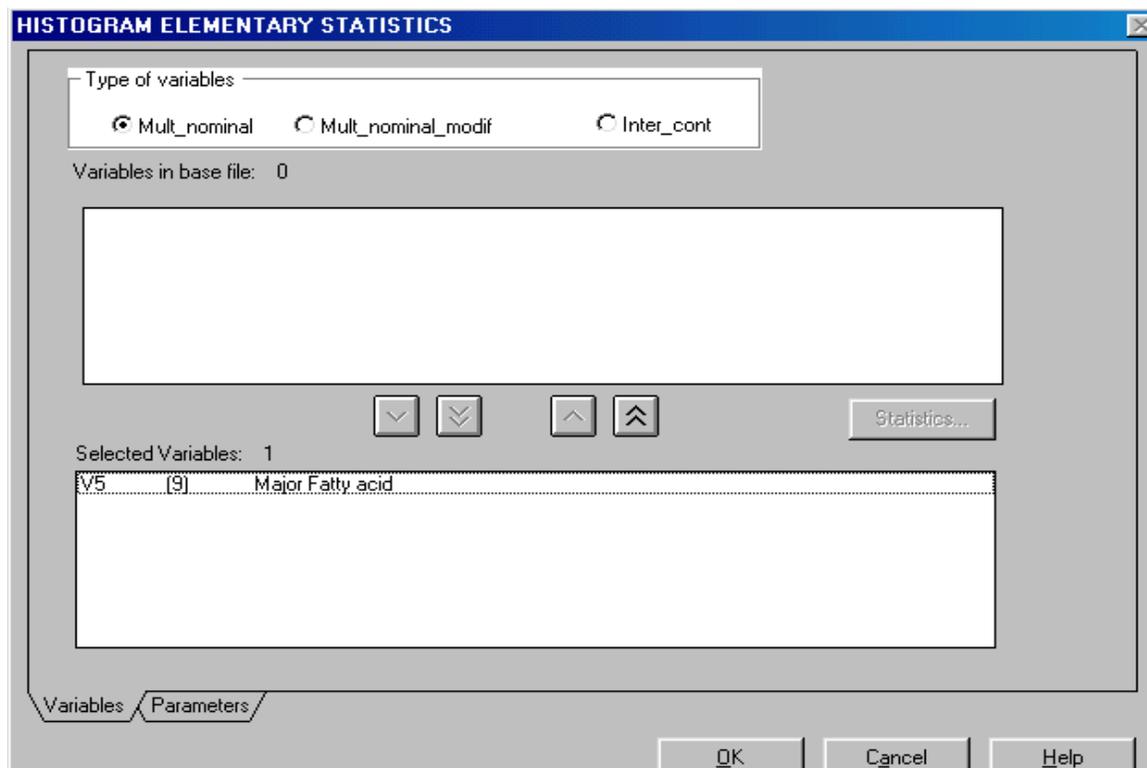
Estatísticas Descritivas: Dados Simbólicos 3/

👉 Comando STAT: Seleção de Variáveis



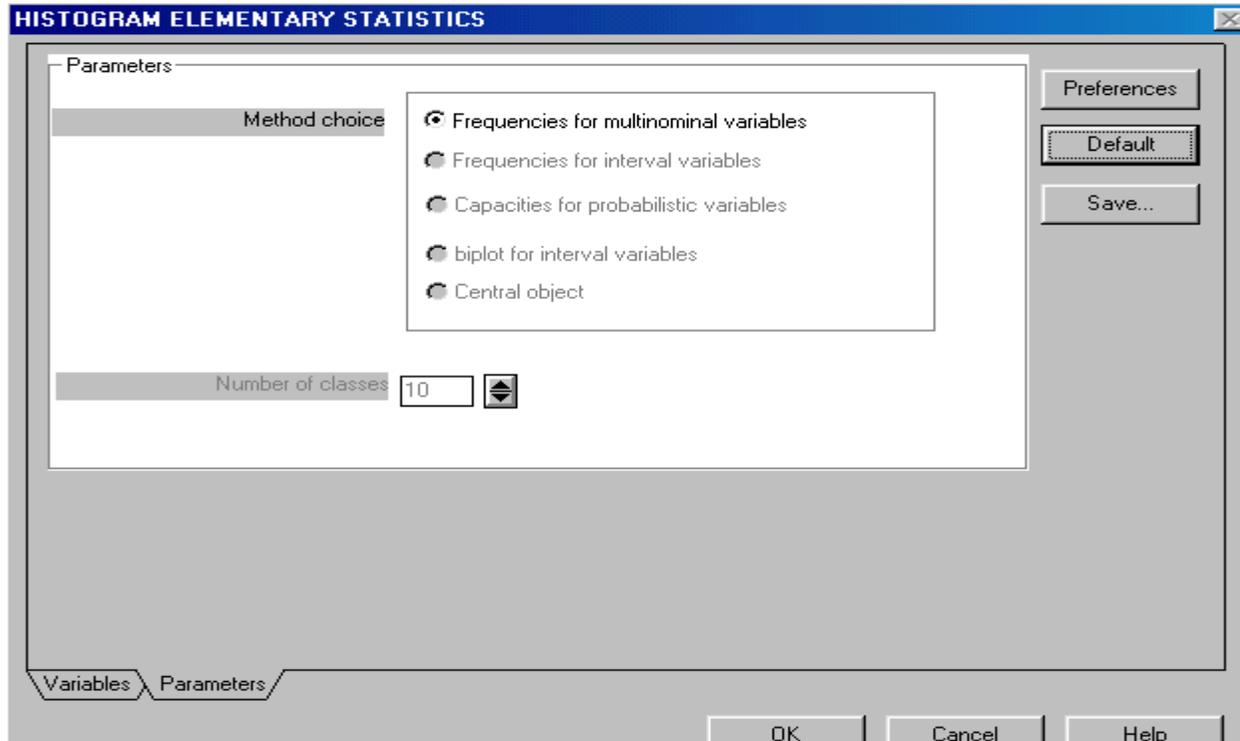
Estatísticas Descritivas: Dados Simbólicos 4/

👉 Comando STAT: Seleção de Variáveis



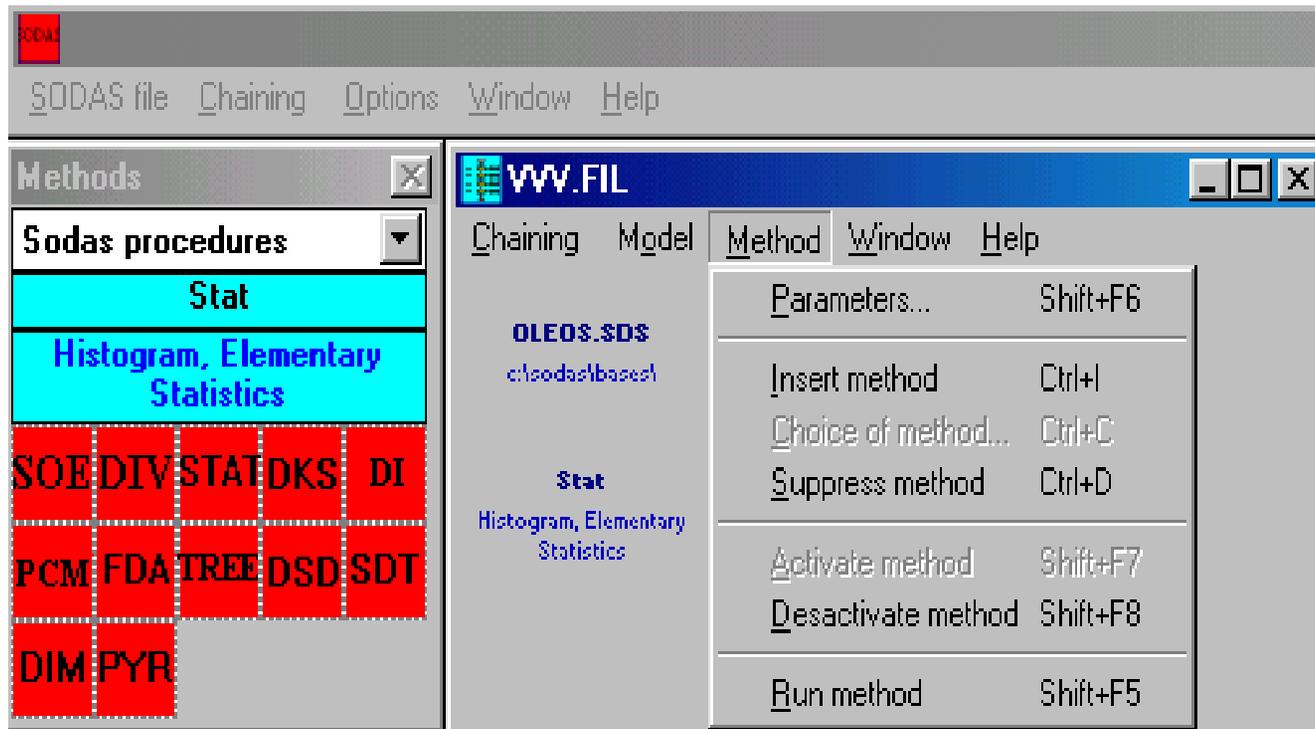
Estatísticas Descritivas: Dados Simbólicos 5/

👉 Comando STAT: Parâmetros



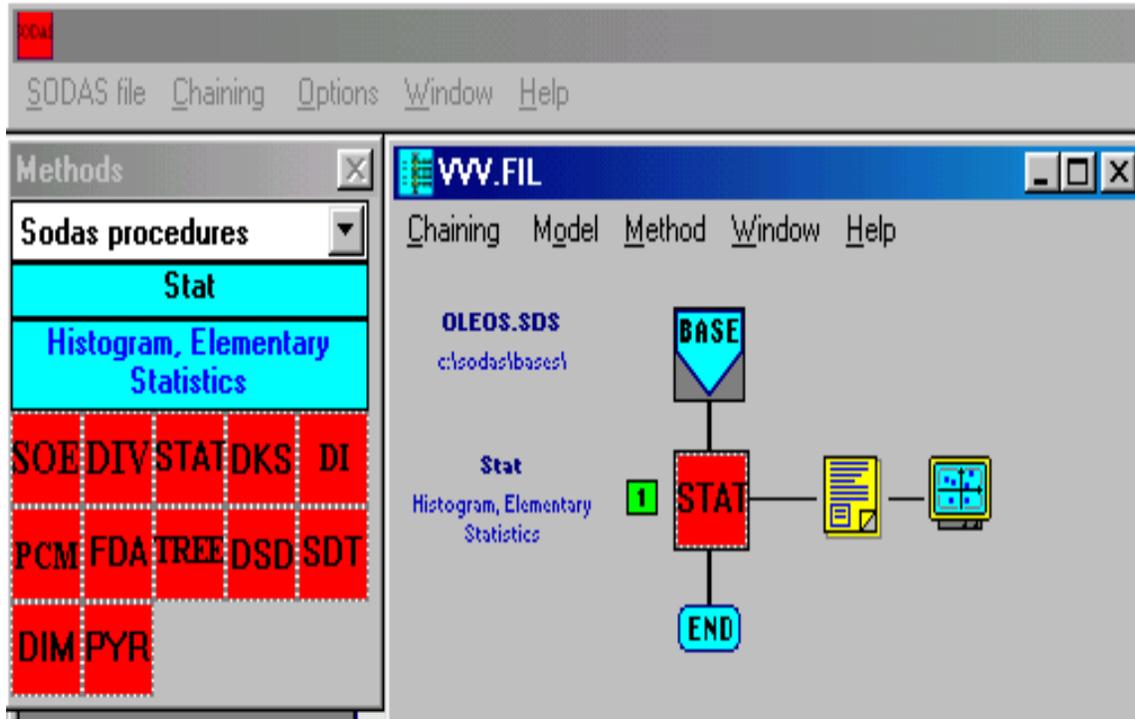
Estatísticas Descritivas: Dados Simbólicos 6/

👉 Comando STAT: Saída



Estatísticas Descritivas: Dados Simbólicos 7/

👉 Comando STAT: Saída



Estatísticas Descritivas: Dados Simbólicos 8/

👉 Comando STAT: Saída

SODAS - STAT RELATIVE FREQUENCIES (MODAL)

File: OLEOS.SDS
Title: FAT and OIL

Major Fatty acid

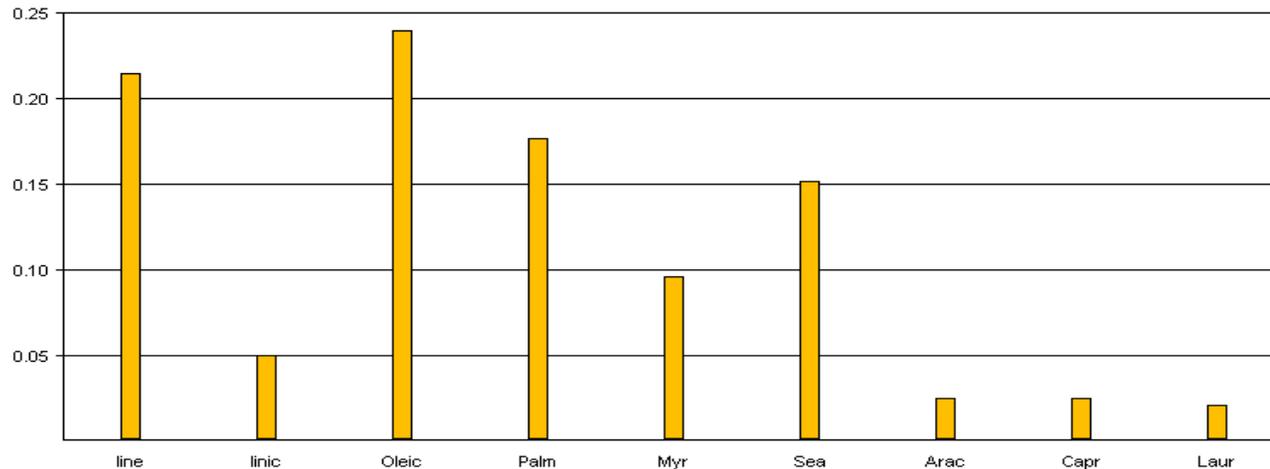
line	Linoleic Acid	0.2146
linic	Linolenic Acid	0.0500
Oleic	Oleic Acid	0.2396
Palm	Palmitic Acid	0.1771
Myr	Myristic Acid	0.0958
Sea	Searic Acid	0.1521
Arac	Arachid Acid	0.0250
Capr	Capric Acid	0.0250
Laur	Lauric Acid	0.0208

Estatísticas Descritivas: Dados Simbólicos 9/

👉 Comando STAT: Saída

FAT and OIL

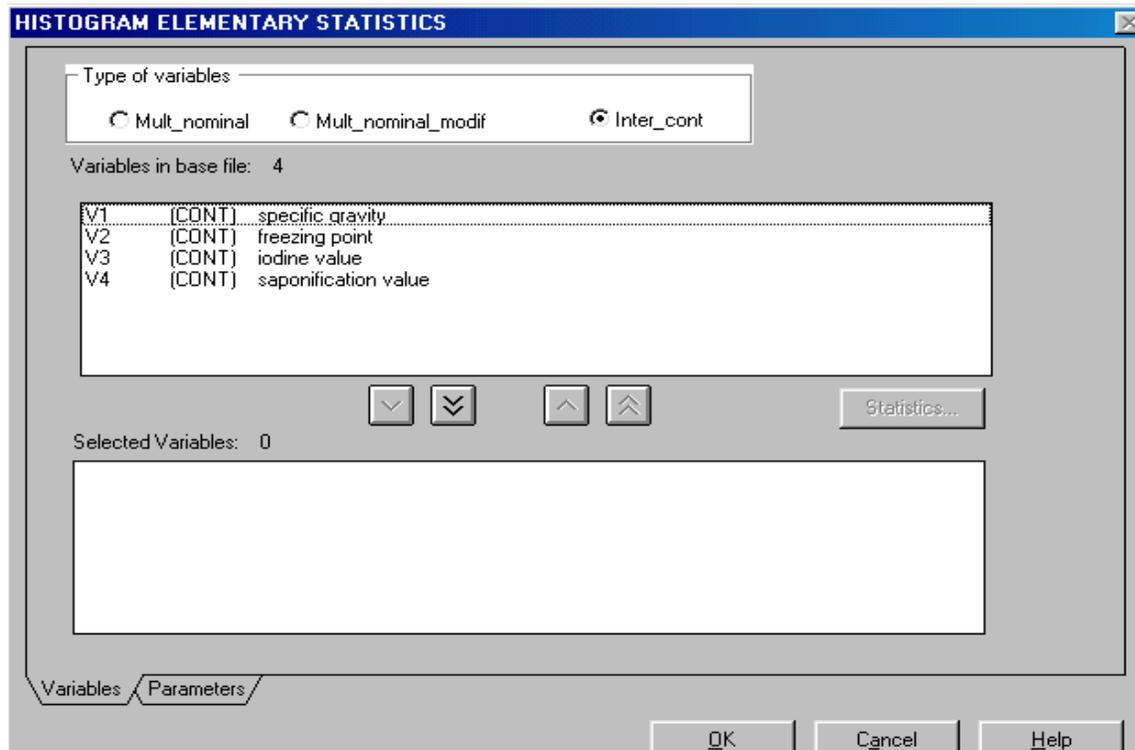
Major Fatty acid



Relative frequencies (multinomial) - file OLEOS.SDS - variable 'Major Fatty acid'

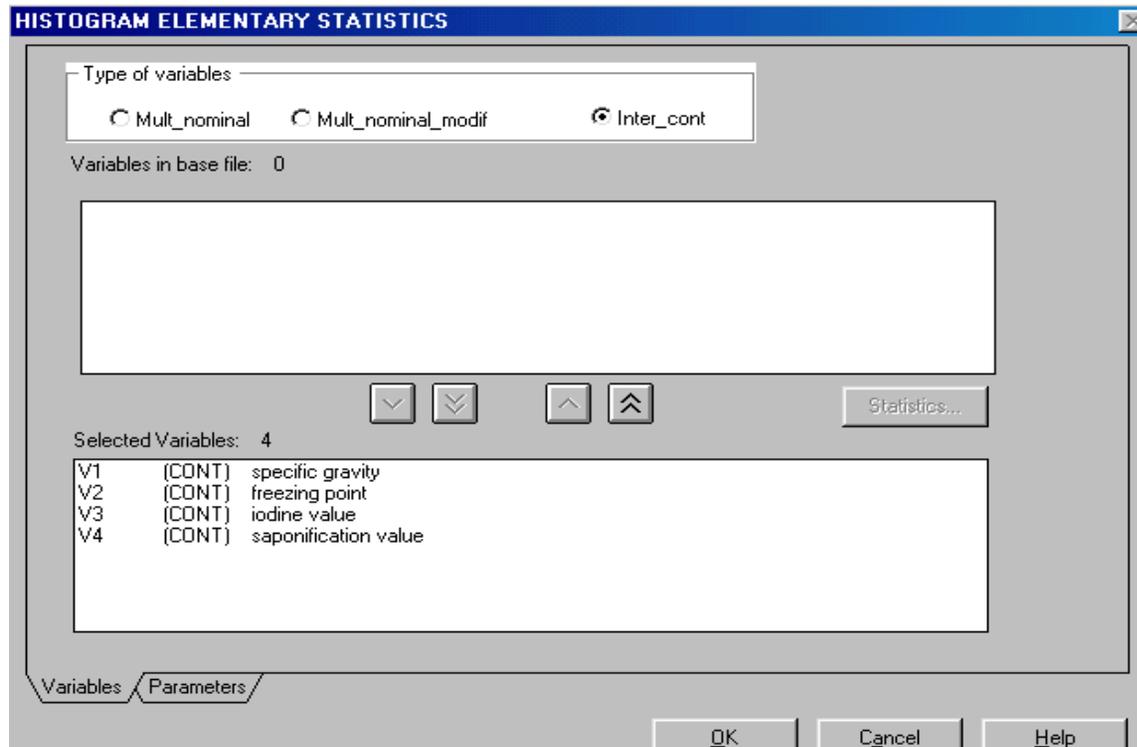
Estatísticas Descritivas: Dados Simbólicos 10/

👉 Comando STAT: Parâmetros



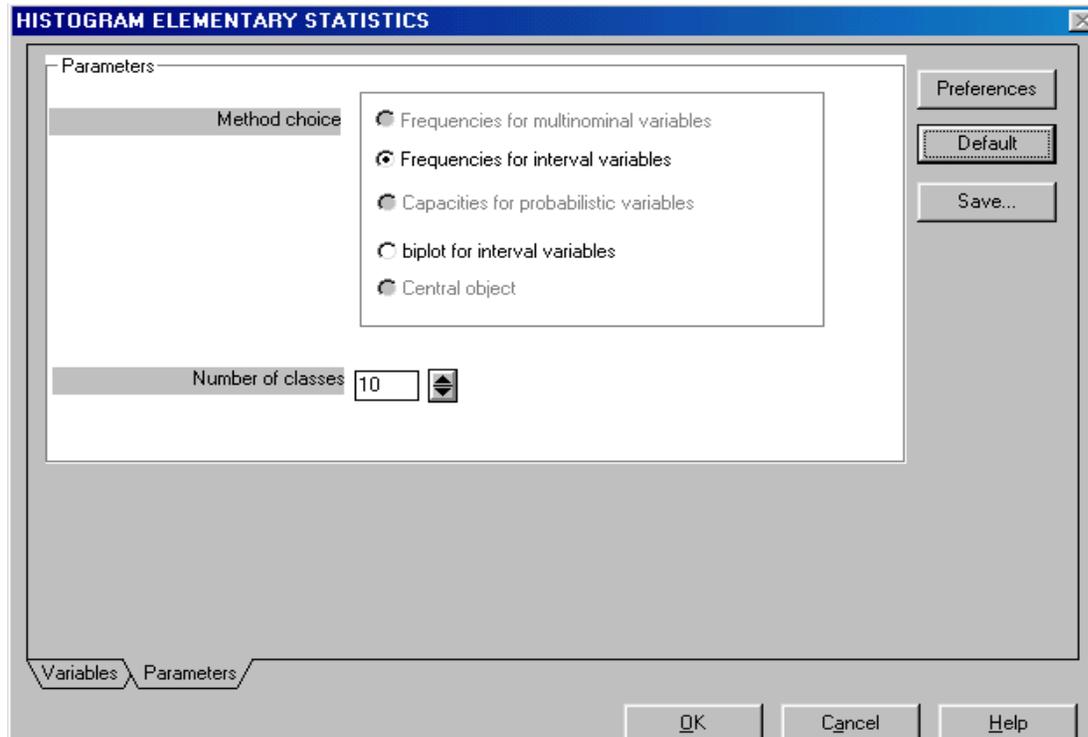
Estatísticas Descritivas: Dados Simbólicos 11/

👉 Comando STAT: Parâmetros



Estatísticas Descritivas: Dados Simbólicos 12/

👉 Comando STAT: Parâmetros



Estatísticas Descritivas: Dados Simbólicos 13/

Comando STAT: Saída

```
-----  
SODAS - STAT  RELATIVE FREQUENCIES (INTERVAL)
```

```
File:  OLEOS.SDS  
Title: FAT and OIL  
-----
```

```
specific gravity
```

```
limits: 0.858 - 0.937  class width: 0.0079
```

```
class 1  0.1987  
class 2  0.0513  
class 3  0.0000  
class 4  0.0000  
class 5  0.0000  
class 6  0.0000  
class 7  0.0000  
class 8  0.4000  
class 9  0.1000  
class 10 0.2500
```

```
Central tendency: 0.9086  
Dispersion: 0.0268
```

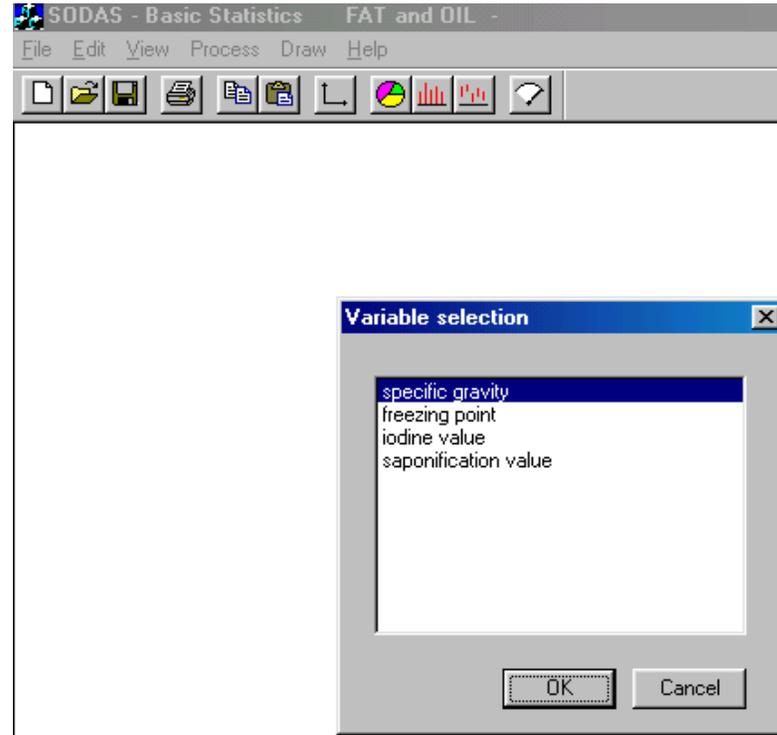
```
freezing point
```

```
limits: -27.0 - 38.0  class width: 6.5
```

```
class 1  0.1007  
class 2  0.1493
```

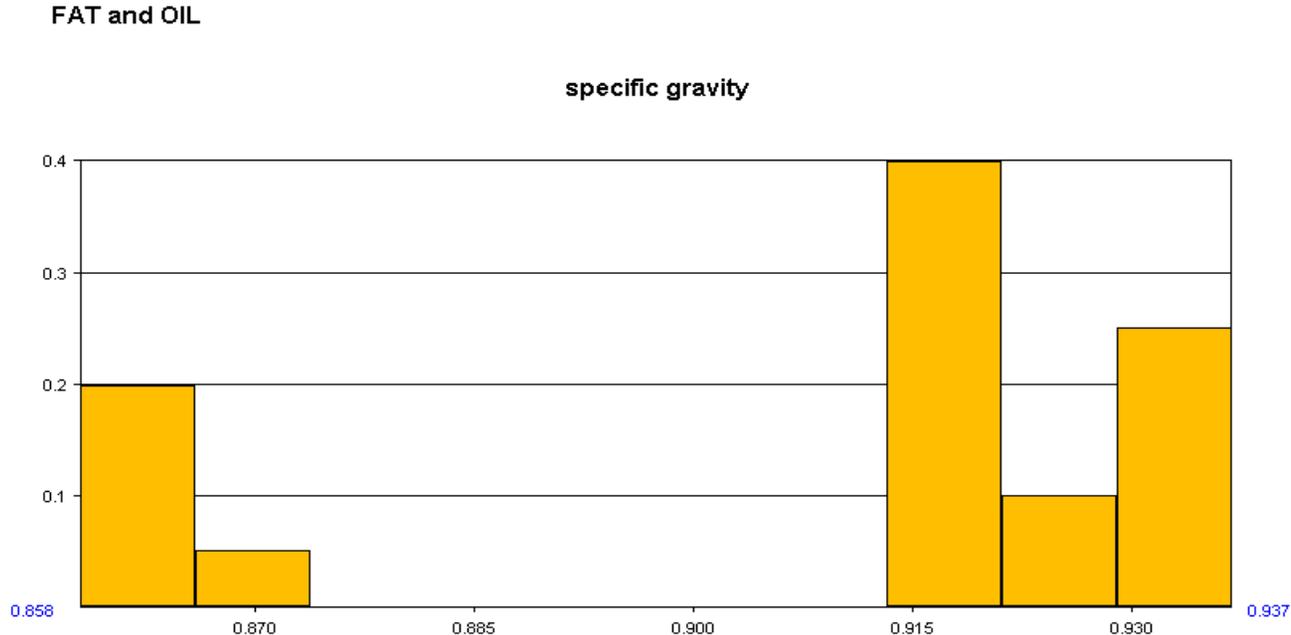
Estatísticas Descritivas: Dados Simbólicos 14/

👉 Comando STAT: Saída



Estatísticas Descritivas: Dados Simbólicos 15/

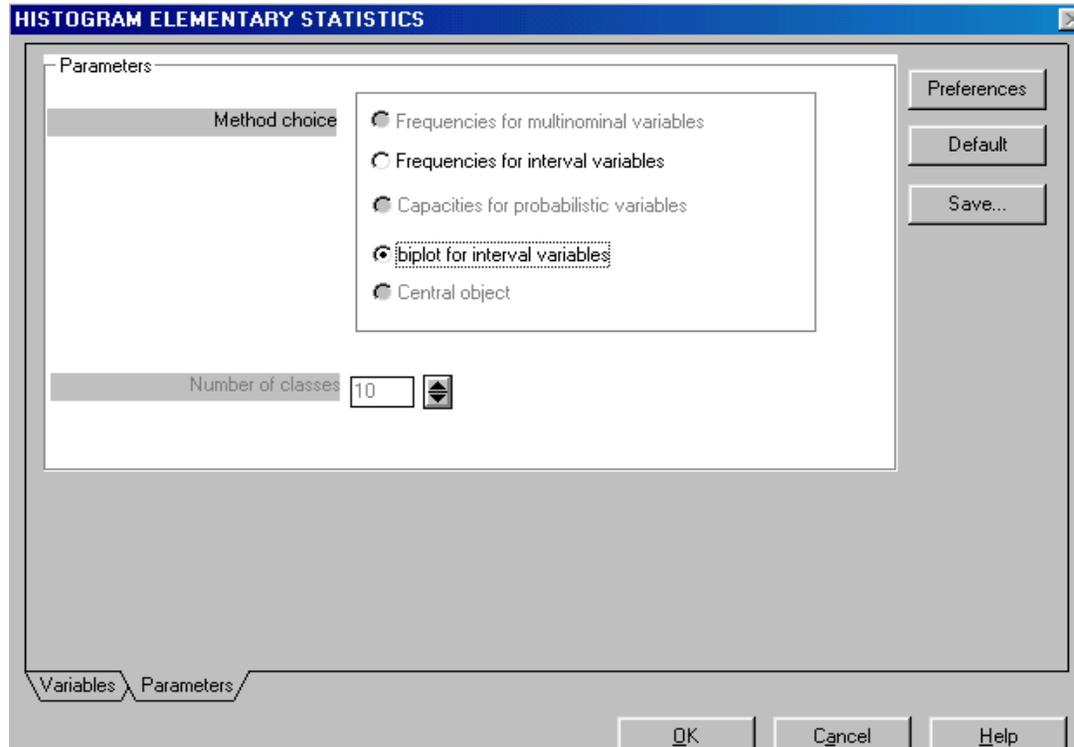
👉 Comando STAT: Saída



Relative frequencies (interval) - file OLEOS.SDS - variable 'specific gravity' - 10 classes - class width: 0.0079

Estatísticas Descritivas: Dados Simbólicos 16/

👉 Comando STAT: Parâmetros



Estatísticas Descritivas: Dados Simbólicos 17/

Comando STAT: Saída

```
-----  
SODAS - STAT  BIPLLOT
```

```
File:  OLEOS.SDS  
Title: FAT and OIL  
-----
```

```
Variables:
```

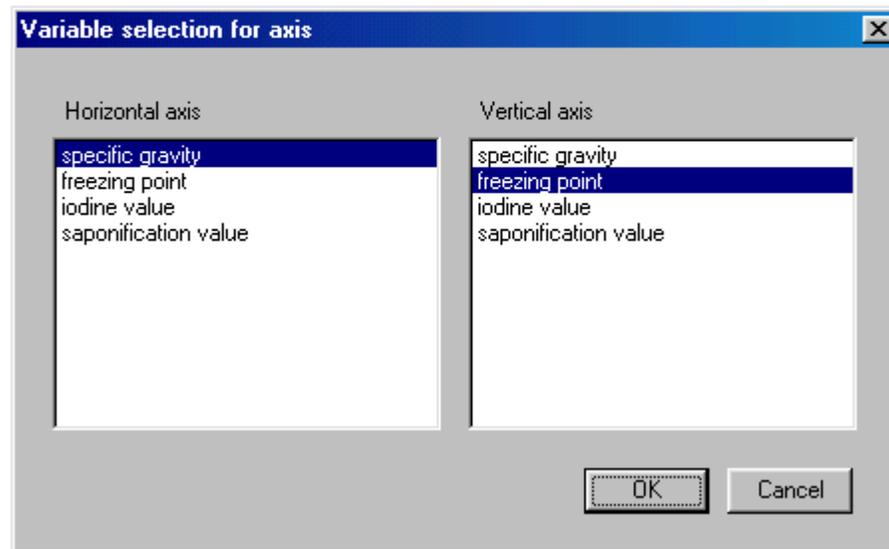
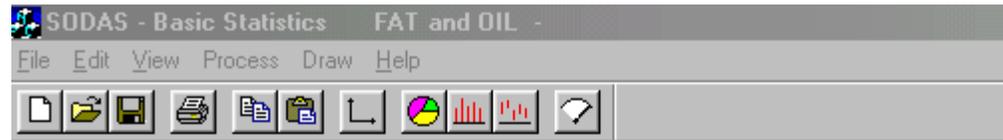
```
specific gravity (0.858 - 0.937)  
freezing point (-27.0 - 38.0)  
iodine value (40.0 - 208.0)  
saponification value (118.0 - 202.0)
```

```
Individuals:
```

```
linsee -  
perill -  
cotton -  
sesame -  
camell -  
olive -  
beef t -  
hog fa -
```

Estatísticas Descritivas: Dados Simbólicos 18/

👉 Comando STAT: Saída



Estatísticas Descritivas: Dados Simbólicos 19/

👉 Comando STAT: Saída

