Organizção dos Dados - Histograma

Note que, não haverá vantagem alguma se organizarmos estes dados numa tabela de

freqüências, uma vez que os dados pouco se repetem. Assim, torna-se útil o agrupamento

dos dados, que, de um modo geral, pode ser feito de acordo com os seguintes passos:

1. Organizar os dados num Rol.

2. Estabelecer o Número de Intervalos (categorias ou classes) para se dividir o con-

junto de dados.

A escolha do número de classes é arbitrária, a qual pode ser estabelecida de acordo

com o bom senso do pesquisador ou obtido por alguma fórmula matemática

construída para este m. Uma sugestão prática é a escolha entre 5 e 15 classes com

a mesma amplitude e duas fórmulas matemáticas que podem orientar na escolha do

número de classes, são:

(a) k = √n

(b) k = 1 + 3, 3 × log(n)

Onde k é o número de classes e n é o número total de observações.

3. Calcular a Amplitude Total:

AT ot = xm´ax − x

m´in

Onde xm´ax e x

m´in é o valor máximo e mínimo observado no conjunto de dados.

4. Determinar a Amplitude de Classe:

h =

AT ot

k

5. A partir do menor valor observado no conjunto de dados, ou de algum valor imediata-

mente inferior e adequadamente escolhido, delimitar as classes, ou seja, determinar

os limites inferiores e superiores de cada classe.

Neste momento, os seguintes símbolos são úteis:

(a) li − |Li - para indicar que o valor extremo inferior (li) não pertence à i − sima

classe, enquanto que o valor extremo superior (Li) pertence.

(b) li|−Li - para indicar que o valor extremo inferior (li) pertence à i−sima classe,

enquanto que o valor extremo superior (Li) não pertence.

6. Após todos estes passos, só resta proceder a contagem do número de observações

pertencentes à cada uma das classes e organizar estas informações numa tabela de

freqüências para dados agrupados.

De acordo com estes passos, o conjunto de dados anterior pode ser organizado como:

9

(Construir a tabela de freqüências para dados agrupados)

Representação Gráca de uma Variável Quantitativa Contínua - Histograma

Para a representação gráca de variáveis quantitativas contínuas é necessário alguma

adaptação do gráco de colunas, uma vez que, em geral, é necessário agrupar os dados em

classes e conseqüentemente há perda de informações.

Histograma - é um gráco indicado para representar dados agrupados em classes.

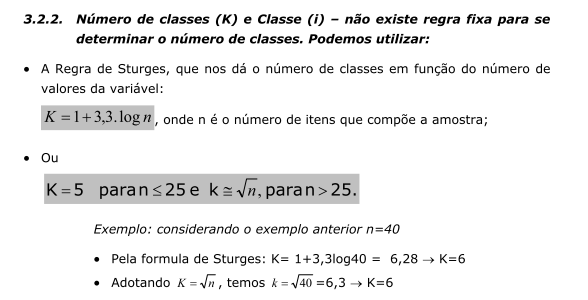
Este gráco é uma adaptação do gráco de colunas, onde as bases correspondem aos

intervalos de classe e as alturas são proporcionais às freqüências de classe. Veja como ca

o histograma para a distribuição das notas:

(Construir o histograma para a distribuição de freqüências em classes)

Xxxxxxxxxxxxxxxxxxxxxx



3.2.2. Número de classes (K) e Classe (i) – não existe regra fixa para se

determinar o número de classes. Podemos utilizar:

• A Regra de Sturges, que nos dá o número de classes em função do número de

valores da variável:

n

K log .3,3 1+ = , onde n é o número de itens que compõe a amostra;

• Ou

25. n para k e 25 n para 5 K

> ≅ ≤ =

,n

Exemplo: considerando o exemplo anterior n=40

• Pela formula de Sturges: K= 1+3,3log40 = 6,28 → K=6

• Adotando n K = , temos 40 = k =6,3 → K=6

2. Estabelecer o Número de Intervalos (categorias ou classes) para se dividir o con-

junto de dados.

=====================================

A escolha do número de classes é arbitrária, a qual pode ser estabelecida de acordo

com o bom senso do pesquisador ou obtido por alguma fórmula matemática

construída para este m. Uma sugestão prática é a escolha entre 5 e 15 classes com

a mesma amplitude e duas fórmulas matemáticas que podem orientar na escolha do

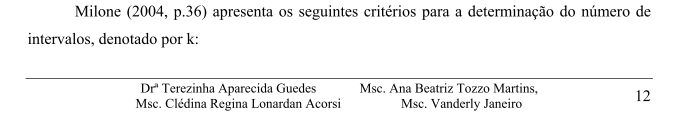
número de classes, são:

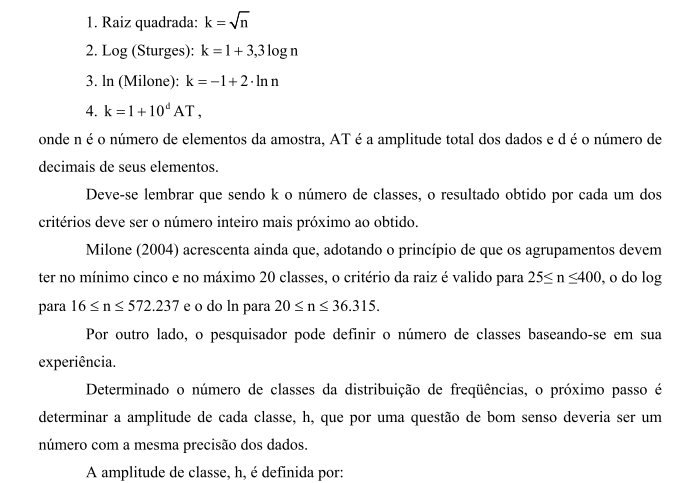
(a) k = √n

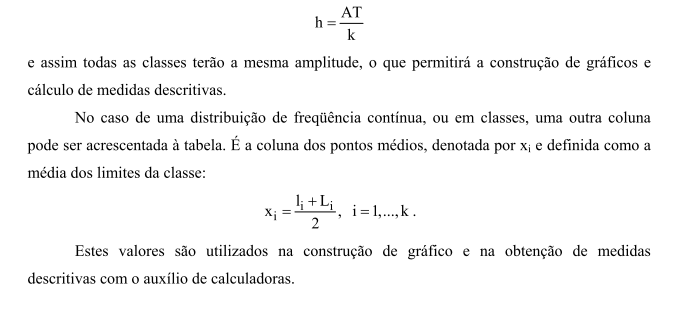
(b) k = 1 + 3, 3 × log(n)

Onde k é o número de classes e n é o número total de observações.

===================================







Milone (2004, p.36) apresenta os seguintes critérios para a determinação do número de

intervalos, denotado por k:

1. Raiz quadrada: k n =

2. Log (Sturges):

n log 3,3 1 k + =

3. ln (Milone): k 1 2 ln n = − + ⋅

4.

AT 10 1 k d + = ,

onde n é o número de elementos da amostra, AT é a amplitude total dos dados e d é o número de

decimais de seus elementos.

Deve-se lembrar que sendo k o número de classes, o resultado obtido por cada um dos

critérios deve ser o número inteiro mais próximo ao obtido.

Milone (2004) acrescenta ainda que, adotando o princípio de que os agrupamentos devem

ter no mínimo cinco e no máximo 20 classes, o critério da raiz é valido para 25≤ n ≤400, o do log

para 16 ≤ n ≤ 572.237 e o do ln para 20 ≤ n ≤ 36.315.

Por outro lado, o pesquisador pode definir o número de classes baseando-se em sua

experiência.

Determinado o número de classes da distribuição de freqüências, o próximo passo é

determinar a amplitude de cada classe, h, que por uma questão de bom senso deveria ser um

número com a mesma precisão dos dados.

A amplitude de classe, h, é definida por:

AT

h

k

=

e assim todas as classes terão a mesma amplitude, o que permitirá a construção de gráficos e

cálculo de medidas descritivas.

No caso de uma distribuição de freqüência contínua, ou em classes, uma outra coluna

pode ser acrescentada à tabela. É a coluna dos pontos médios, denotada por xi e definida como a

média dos limites da classe:

i i

i

l L

x , i 1,..., k

2

+

= = .

Estes valores são utilizados na construção de gráfico e na obtenção de medidas

descritivas com o auxílio de calculadoras.