**DISTRIBUIÇÃO DE FREQUÊNCIA**

 É um tipo de tabela **que condensa uma coleção de dados conforme as freqüências** (repetições de seus valores).

***Tabela primitiva ou dados brutos:*** É uma **tabela ou relação de elementos que não foram numericamente organizados**. É difícil formarmos uma idéia exata do comportamento do grupo como um todo, a partir de dados não ordenados.

*Ex : 45, 41, 42, 41, 42 43, 44, 41 ,50, 46, 50, 46, 60, 54, 52, 58, 57, 58, 60, 51*

***ROL:*** É a **tabela obtida após a ordenação dos dados** (crescente ou decrescente).

*Ex : 41, 41, 41, 42, 42 43, 44, 45 ,46, 46, 50, 50, 51, 52, 54, 57, 58, 58, 60, 60*

***Distribuição de frequência sem intervalos de classe:*** É a simples condensação dos dados conforme as repetições de seus valores. Para um **ROL** de tamanho razoável esta distribuição de frequência é inconveniente, já que exige muito espaço. Veja exemplo abaixo:

|  |  |
| --- | --- |
| Dados | Frequência |
| 41 | 3 |
| 42 | 2 |
| 43 | 1 |
| 44 | 1 |
| 45 | 1 |
| 46 | 2 |
| 50 | 2 |
| 51 | 1 |
| 52 | 1 |
| 54 | 1 |
| 57 | 1 |
| 58 | 2 |
| 60 | 2 |
| Total | 20 |

***Distribuição de frequência com intervalos de classe:*** Quando o tamanho da amostra é elevado, é mais racional efetuar o agrupamento dos valores em vários intervalos de classe.

60-41 = 19 $\sqrt{n}= \sqrt{20}=4,47$ (**linhas da tabela**)

h= (de quanto em quanto varia o intervalo) h = $\frac{19}{5}=3,8 $arredonda 4

**Ou Regra de Sturges = k = 1 +3,3. log (n) ou** $\sqrt{n}=$

K = 1 + 3,3 . log 20

K = 1+ 3,3 . 1,3

K= 1 + 4,29 = 5,3 arredonda **5 linhas**

|  |  |
| --- | --- |
| Classes | Frequências |
| 41 |------- 45 | 7 |
| 45 |------- 49 | 3 |
| 49 |------- 53 | 4 |
| 53 |------- 57 | 1 |
| 57 |------- 61 | 5 |
| Total | 20 |

1 número : inclui e o último : exclui

**ELEMENTOS DE UMA DISTRIBUIÇÃO DE FREQÜÊNCIA (com intervalos de classe) **

**CLASSE:** são os intervalos de variação da variável e é simbolizada por ***i*** e o número total de classes simbolizada por ***k***. Ex: na tabela anterior ***k*** = 5 e **49 |------- 53** é a **3ª classe**, onde ***i*** = 3.

**LIMITES DE CLASSE:** são os extremos de cada classe. O menor número é o limite inferior de classe ( **li** ) e o maior número, limite superior de classe ( **Li** ). *Ex: em* ***49 |------- 53,... l3*** *= 49 e* ***L3*** *= 53. O símbolo* ***|-------*** *representa* ***um intervalo fechado à esquerda e aberto à direita.*** *O dado 53 do* ***ROL*** *não pertence a classe 3 e sim a classe 4 representada por* ***53 |------- 57****.*

**AMPLITUDE DO INTERVALO DE CLASSE:** é obtida através da diferença entre o limite superior e inferior da classe e é simbolizada por ***hi = Li - li***. *Ex: na tabela anterior* ***hi*** *= 53 - 49 = 4.* ***Obs****: Na distribuição de freqüência c/ classe o* ***hi será igual em todas as classes.***

**AMPLITUDE TOTAL DA DISTRIBUIÇÃO:**  é a diferença entre o limite superior da última classe e o limite inferior da primeira classe. ***AT = L(max) - l(min)***. *Ex: na tabela anterior* ***AT*** *= 61 - 41= 20.*

**AMPLITUDE TOTAL DA AMOSTRA (ROL):** é a diferença entre o valor máximo e o valor mínimo da amostra (ROL). Onde ***AA = Xmax - Xmin***. Em nosso exemplo ***AA*** = 60 - 41 = 19.

**Obs**: ***AT*** sempre será maior que ***AA***.

**PONTO MÉDIO DE CLASSE**: é o ponto que divide o intervalo de classe em duas partes iguais. *.......Ex: em* ***49 |------- 53*** *o ponto médio* ***x3*** *= (53+49)/2 = 51, ou seja* ***x3=( l3 + L3 )/2.***

**Método prático para construção de uma Distribuição de Freqüências c/ Classe **

**1º -** Organize os dados brutos em um **ROL**.

**2º -** Calcule a **amplitude amostral** ***AA***.

* + - No nosso exemplo: ***AA*** = 60 - 41 = 19

**3º -** Calcule o número de classes através da "**Regra de Sturges": ou**

 Nesta expressão:

– k é o número de classes.

– N é o número total de observações na amostra.

– Log é o logaritmo comum da base 10.

**Obs:** Qualquer regra para determinação do nº de classes da tabela não nos levam a uma decisão final; esta vai depender, na realidade de um julgamento pessoal, que deve estar ligado à natureza dos dados.

No nosso exemplo: ***n*** = 20 dados, então ,a princípio, a regra sugere a adoção de 5 classes.

**4º -** Decidido o nº de classes, calcule então a amplitude do intervalo de classe**h > AA / i**.

No nosso exemplo: **AA/i** = 19/5 = 3,8 . Obs: Como **h > AA/i** um valor ligeiramente superior para haver folga na última classe. Utilizaremos então ***h*** = 4

**5º -** Temos então o menor nº da amostra, o nº de classes e a amplitude do intervalo. Podemos montar a tabela, com o cuidado para não aparecer classes com freqüência = 0 (zero).

No nosso exemplo: o menor nº da amostra = 41 + ***h*** = 45, logo a primeira classe será representada por ...... **41 |------- 45**. As classes seguintes respeitarão o mesmo procedimento.

O primeiro elemento das classes seguintes sempre serão formadas pelo último elemento da classe anterior.

**TIPOS DE FREQUÊNCIAS**

***Freqüências simples ou absoluta:*** são os valores que realmente representam o número de dados de cada classe. A soma das freqüências simples é igual ao número total dos dados da distribuição.

***Freqüências relativas:*** são os valores das razões entre as freqüência absolutas de cada classe e a freqüência total da distribuição. A soma das **freqüências relativas é igual a 1 (100 %).**

.

**Polígono de freqüência:** é um gráfico em linha, sendo as freqüências marcadas sobre perpendiculares ao eixo horizontal, levantadas pelos pontos médios dos intervalos de classe. Para realmente obtermos um polígono (linha fechada), devemos completar a figura, ligando os extremos da linha obtida aos pontos médios da classe anterior à primeira e da posterior à última, da distribuição.

.

**Polígono de freqüência acumulada:** é traçado marcando-se as freqüências acumuladas sobre perpendiculares ao eixo horizontal, levantadas nos pontos correspondentes aos limites superiores dos intervalos de classe.

***Freqüência simples acumulada de uma classe:*** é o total das freqüências de todos os valores inferiores ao limite superior do intervalo de uma determinada classe.

***Freqüência relativa acumulada de uma classe:*** é a freqüência acumulada da classe, dividida pela frequência total da distribuição.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **CLASSE** | **f freq. simples** | **xi Ponto Médio** | **fr freq. relativa** | **Fr Percentual** | **FA** | **FAr** |
| 50 |-------- 54  | 4 |  $\frac{50+54}{2}$ 52 | $\frac{4}{40}= $0,100 | 10% | 4 | $\frac{4}{40}=$0,100 |
| 54 |-------- 58 | 9 | 56 | $\frac{9}{40}= $0,225 | 22,5% | 13 | $\frac{13}{40}=$0,325 |
| 58 |-------- 62 | 11 | 60 | $\frac{11}{40}= $0,275 | 27,5% | 24 | 0,600 |
| 62 |-------- 66 | 8 | 64 | $\frac{8}{40}= $0,200 | 20% | 32 | 0,800 |
| 66 |-------- 70 | 5 | 68 | $\frac{5}{40}= $0,125 | 12,5% | 37 | 0,925 |
| 70 |-------- 74 | 3 | 72 | $\frac{3}{40}= $0,075 | 7,5% | 40 | 1,000 |
| Total | 40 |  | 1,000 | 100% |  |  |

**f** = frequência simples; **xi** = ponto médio de classe; **fr** = frequência relativa

**FA**= frequência simples acumulada; e **FAr** = frequência relativa acumulada.

* **Obs:** uma distribuição de freqüência sem intervalos de classe é representada graficamente por um diagrama onde cada valor da variável é representado por um segmento de reta vertical e de comprimento proporcional à respectiva freqüência.

.

**EXERCÍCIO**

Os dados abaixo referem-se ao número de pessoas que residem em uma amostra **de 35 domicílios** do bairro Vila Nova

9-1 = 8 $\sqrt{n}= \sqrt{35}=5,9$ (**linhas da tabela**)

h= (de quanto em quanto varia o intervalo) h = $\frac{8}{5,9}=1,36 $arredonda 2

**Ou Regra de Sturges = k = 1 +3,3. log (n) ou** $\sqrt{n}=$

K = 1 + 3,3 . log 35

K = 1+ 3,3 . 1,54

K= 1 + 5,1 = 6 arredonda **6linhas= 5 linhas Pesquisador**

2 3 4 4 5 3 4

5 6 5 3 1 5 5

1 3 4 5 5 5 3

2 2 5 4 4 2 3

5 4 5 4 2 4 9

a) **Construa uma distribuição de frequência em classes**.

b) Determine as frequências simples acumuladas de cada classe.

c) Determine as frequências relativas de cada classe.

d) Qual o percentual de pessoas que residem **mais do que** 4 em um domicílio? **37%**

e) Qual o percentual de pessoas que residem até duas pessoas em um domicílio? **20%**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **CLASSE** | **f freq. simples** | **xi Ponto Médio** | **fr freq. relativa** | **Fr Percentual** | **FA** |  |
| 1 |-------- 3  | 7 | (1+3)/2 = 2 | $\frac{7}{35}=0,2$ x100 | 20% | 7 |  |
| 3 |-------- 5 | 15 | 4 | $$\frac{15}{35 }=0,43$$ | 43% | 22 |  |
| 5 |-------- 7 | 12 | 6 | $$\frac{12}{35}=0,34$$ | 34% | 34 |  |
| 7 |-------- 9 | 0 | 8 | 0 | 0% | 34 |  |
| 9 |-------- 11 | 1 | 10 | $$\frac{1}{35}=0,03$$ | 3% | 35 |  |
| Total | 35 |  |  | 100% |  |  |

.

**CÁLCULO DA MEDIDAS DESCRITIVAS**

**MÉDIA, MEDIANA, MODA, DESVIO PADRÃO E COEFICIENTE DE VARIAÇÃO**

EXEMPLO 1: DADOS SEM AGRUPAMENTO – Exemplo: número de irmãos da nossa turma

**1 – 1-0-3-6-3-2-1-5-5**

**Rol : 0-1-1-1-2-3-3-5-5-6**

**x-x-x-x-x-x-x qtos alunos (amostra n=10)**

**Média:** Número que representa uma série ou sequência de dados

MÉDIA X = $\frac{\sum\_{}^{}x.}{n}$ = $\frac{27}{10}=2,7$

**Moda:** número que mais se repete em uma série de dados

MODA: Mo = 1

**Mediana:** divide uma sequência de dados em duas partes iguais

**0-1-1-1-2-3-3-5-5-6**

**Md = 2,5**

**Par (10): (2+3)/2 = 2,5**

**Exemplo adaptando o anterior deixando 9 alunos Impar: 1-1-1-2-3-3-5-5-6 = Md =3**

MEDIDAS DE DISPERSÃO

**Desvio Padrão:** Mede a variabilidade dos dados (homogeneidade)

Desvio Padrão

$s= \sqrt{\frac{\sum\_{}^{}x^{2}-n.(x^{2})}{n-1}}$ = $= \sqrt{\frac{111-10 . (2,7^{2})}{10-1}}= \sqrt{\frac{111-10 . 7,29}{10-1}}$ = $\sqrt{\frac{111-72,9}{9}}$ = $\sqrt{\frac{38,1}{9}}$ = $\sqrt{4,23}$ = 2,05 = 2,1

**(x)0-1-1-1-2-3-3-5-5-6 = 02 +12+12+12+22+32+32+52+52+62 = 0+1+1+1+4+9+9+25+25+36=111**

**Coeficiente de variação:** mostra a variabilidade dos dados em porcentagem

Coeficiente de variação: (desvio padrão div. Pela média)

$CV= \frac{S}{x} . 100$ = $CV= \frac{2,1}{2,7}.100=0,7778 . 100=77,78\%$ grande variação

**Exemplo 2 : Dados agrupados SEM intervalo de classes**

 A distribuição abaixo indica o **número de acidentes** ocorridos com 70 motoristas de uma empresa de ônibus:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Nº de acidentes x | No motoristas (f) frequencia | fr frequencia relativa percentual | **FA** | **x.f** | **x2.f** |
| **0** | **20** | $\frac{20}{70}= $**28,57%** | **20** | **0.20=0** | **02.20= 0** |
| **1** | **10** | $\frac{10}{70}$**14,29%** | **30** | **1.10=10** | **12. 10 =10** |
| **2** | **16** | $\frac{16}{70}$**22,86%** | **46**  | **2.16=32** | **22 .16 = 64** |
| **3** | **9** | $\frac{9}{70}$**12,86%** | **55** | **3.9=27** | **32.9=81** |
| **4** | **6** | $\frac{6}{70}$**8,57%** | **61** | **4.6=24** | **42.6=96** |
| **5** | **5** | $\frac{5}{70}$**7,14%** | **66** | **5.5=25** | **52.5=125** |
| **6** | **3** | $\frac{3}{70}$**4,29%** | **69** | **6.3=18** | **62.3=108** |
| **7** | **1** | $\frac{1}{70}$**1,43%** | **70** | **7.1=7** | **72.1=49** |
| **Total ∑** | **70** | **100,00%** |  | **143 acidentes** | **533** |

Determine:

a) o númerode motoristas que não sofreram nenhum acidente. 20

b) o númerode motoristas que sofreram **pelo menos 4** acidentes. 15

c) o númerode motoristas que sofreram **menos** de 3 acidentes. 46

d) o númerode motoristas que sofreram no mínimo 3 e no máximo 5 acidentes.

e) a percentagem dos motoristas que sofreram no máximo 2 acidentes.

CALCULE AS MEDIDAS DESCRITIVAS

MÉDIA X = $\frac{\sum\_{}^{}x.f}{n}$ = $\frac{143}{70}=2,04=2,0$ =

MODA: (maior frequencia) = 0 acidentes

**Mediana:** Md= $\frac{\sum\_{}^{}f}{2}=\frac{70}{2}=35$ procura valor igual ou que ultrapasse esse valor na coluna FA (frequencia acumulada).

Se ultrapassar: é o valor da primeira coluna direto

Se igual: Soma o valor da linha mediana+ próxima linha e divide por 2.

Nosso caso ultrapassou então a Mediana é Md = 2

**Desvio Padrão:**

Desvio Padrão

$s= \sqrt{\frac{\sum\_{}^{}x^{2}.f-n.(x^{2})}{n-1}}$ = $= \sqrt{\frac{533-70. (2^{2})}{70-1}}$ = $\sqrt{\frac{533-70.4}{69}}$ .

$\sqrt{\frac{533-280}{69}}$ =$\sqrt{\frac{253}{69}}= \sqrt{3,66= } $1,9

**Coeficiente de variação:** mostra a variabilidade dos dados em porcentagem

Coeficiente de variação:

$CV= \frac{S}{x} . 100$ = $CV= \frac{1,9}{2} . 100=$95%

**CÁLCULO DAS MEDIDAS DESCRITIVAS EM DADOS COM INTERVALO DE CLASSE**

EXEMPLO COM INTERVALO DE CLASSES.

EXEMPLOS: 41, 41, 41, 42, 42, 43, 44, 45, 46, 46, 50, 50, 51, 52, 54, 57, 58, 58, 60,60

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Classes | Frequencias ( f) | FA | Freq percentual | ponto médio (x) | x. f média | x2.f desvio padrao |
| 41 - 45  | 7 | 7 | 7/20 = 0,35 ou 35% | (41+45)/2 =43 | 43.7=301 | 432.7= 12943 |
| 45 - 49 | 3 | 10 | 0,15 ou 15% | 47 | 47.3=141 | 472.3=6627 |
| 49 - 53 | 4 | 14 | 0,20 ou 20% | 51 | 51.4=204 | 512.4=10404 |
| 53 - 57 | 1 | 15 | 0,05 ou 5% | 55 | 55.1=55 | 552.1=3025 |
| 57 - 61 | 5 | 20 | 0,25 ou 25% | 59 | 59.5=295 | 592.5=17405 |
| Total ∑ | 20 |  | 100% |  | 996 | **50404** |

MÉDIA X = $\frac{\sum\_{}^{}x.f}{n}$ = $\frac{996}{20}$ = 49,8

**Arredondamento 8/85 = 0,094 menor q 5, deixa do jeito q está 0,09 x 100 = 9%**

 **9/85 = 0,0105 igual ou maior q 5, arredonda p mais = 0,11 x 100 = 11%**

MODA: 43

MEDIANA EM DADOS AGRUPADOS

1º Passo: $\sum\_{}^{}\frac{f}{2}$ = $\frac{20}{2}=10$ Este procura da FA da tabela, valor **igual ou que supere**

Med = $ l+ \frac{\left⌊\frac{\sum\_{}^{}f}{2}-FAA\right⌋.h}{f}$ = = $ 45+ \frac{\left⌊10-7\right⌋.4}{3}$ = 45 + 4 = 49

l= limite inferior da classe mediana

FAA – Frequencia acumulada anterior a classe mediana

h= amplitude da classe mediana

f= frequência simples da classe mediana

Desvio Padrão

$s= \sqrt{\frac{\sum\_{}^{}x^{2}.f-n.(x^{2})}{n-1}}$ = $= \sqrt{\frac{50404-20 .(49,8)^{2}}{20-1}}$ = $\sqrt{\frac{50404-20 .2480,04}{20-1}}$ =

$\sqrt{\frac{50404-49600,8}{19}}$ =$\sqrt{42,27=} 6,5$

**Coeficiente de variação:**

$CV= \frac{S}{x} . 100$ = $CV= \frac{6,5}{49,8} . 100= 13,05\%$

 **EXERCÍCIOS DE FIXAÇÃO E PREPARAÇÃO PARA AVALIAÇÃO**

**Exercícios de Fixação: cálculo de medidas descritivas**

1. **Seja a sequência de números que representam o número de animais (aves) de estimação que os acadêmicos na turma possuem em casa: 0- 2 – 3 – 3 -4- 4 – 4 – 5**

Calcule a média, mediana, moda e desvio padrão e coeficiente de variação:

MÉDIA X =

Moda:

Mediana:

Então sequencia par **:**

Desvio padrão:

Coeficiente de variação:

$CV= \frac{S}{x} . 100$ = $CV=$

Exemplo com dados agrupados com intervalo de classes:

Seja a distribuição das notas de uma turma de 44 alunos:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Notas | f | Freq relativa | FA | Ponto médio x | x .f | x2.f |
| 0 – 2 | 5 | 5/44 = 0,11 =11% | 5 | (0+2)/2 = 1 | 1.5 = 5 | 12 . 5 = 5 |
| 2 – 4 | 8 | 8/44= 0,18 = 18% | 13 | 3 | 3.8=24 | 32 . 8 = 72 |
| 4 - 6 | 14 | 14/44 = 0,318 = 32% | 27 | 5 | 5.14= 70 | 52 . 14 = 350 |
| 6 - 8 | 10 | 10/44 = 0,227 = 23% | 37 | 7 | 7.10 = 70 | 72.10 = 490 |
| 8 - 10 | 7 | 7/44 = 0,159 = 16% | 44 | 9 | 9.7= 63 | 92 .7=567 |
| **Total** | **44** | **100,00%** |  |  | **232 pontos** | **1484** |

Exercício: Calcule a média, mediana, moda e desvio padrão e coeficiente de variação:

MÉDIA X = $\frac{\sum\_{}^{}x.f}{n}$ = $\frac{232}{44}=5,27=5,3$

MODA: 5

MEDIANA EM DADOS AGRUPADOS (não cai na prova)

Desvio Padrão $\sqrt{\frac{∑x^{2}.f-n.(x)^{2}}{n-1}}$ = $\sqrt{\frac{1484-44.(5,3)^{2}}{44-1}}$ = $\sqrt{\frac{1484-44. 28,09}{44-1}}$ = $\sqrt{\frac{1484-1235,96}{44-1}}$ = $\sqrt{\frac{248,04}{43}}$ = $\sqrt{5,7}=2,38=2,4$

Coeficiente de variação:

$CV= \frac{S}{x} . 100$ = $CV$ = $\frac{2,4}{5,3}=0,452=0,45 . 100=45\% $

Cv = desvio padrão dividido pelo média, vezes 100. Tem o resultado em porcentagem