**8. Distribuição de Freqüência**

**8.1. Introdução**

Na Estatística trabalhamos, habitualmente, com grande número de dados, resultados de pesquisas realizadas.

Assim, é impossível examiná-los, mesmo que arrolados em ordem crescente ou decrescente. Daí os estatísticos condensarem tais dados em tabelas de fácil manejo. Os dados são assim apresentados em tabelas de freqüências.

**8.2. Conceitos Preliminares**

**a)** POPULAÇÃO OU UNIVERSO: conjunto de objetos ou indivíduos que apresentam “certa característica” pode ser finita ou infinita.

**b)** AMOSTRA: parte escolhida da população, com as mesmas características dela.

**c)** VARIÁVEL: símbolo capaz de assumir valores dentro de um conjunto. Pode ser:

1) Discreta - assume valores em pontos exatos. Ex.: pontos de um dado (1, 2, 3, 4, 5, 6);

2) Contínua - pode teoricamente, assumir qualquer valor entre dois valores dados. Ex.: altura entre 1,70 m e 1,80 m.

**d)** DADOS BRUTOS: dados não organizados.

**e)** ROL: dados brutos, obedecendo a uma ordem.

**f)** CLASSES DE FREQÜÊNCIA: agrupamento de dados brutos.

**g)** FREQÜÊNCIAS DE CLASSE (f): no de elementos de cada classe.

**Obs.**: para representar a soma ou total das freqüências de uma classe usamos a nomenclatura Σf.

**h)** DISTRIBUIÇÃO DE FREQÜÊNCIA: é uma série com local, época e espécie fixos, mas o fato é agrupado. É o conjunto formado pelas classes de freqüência e as freqüências de classe.

**i)** ARGUMENTO: fato ou espécie estudado (X, Y,...).

**j)** LIMITES: nos extremos de cada classe de freqüência

6 ⎯ 8 [6, 8)

8 ⎯ 10 (8, 10)

10 ⎯ 12 [10, 12], onde os 1os números são os limites inferiores e os 2os números são os limites superiores.

**l)** INTERVALO (h): é a diferença entre dois limites inferiores (ou superiores) consecutivos.

Os intervalos devem ser constantes.

**m)** PONTO MÉDIO (Xi): é o valor obtido pela soma entre o limite inferior da classe e a metade do intervalo (h/2).

Xi = limite inferior + h/2

Ou

É o valor obtido pelo quociente entre a soma do lim. superior com o lim. inferior de uma classe e 2.

Xi = limite superior + limite inferior

2

**n)** CLASSE MODAL: classe de maior freqüência (XMo)

**Obs.**: maior freqüência = f máximo.

**o)** AMPLITUDE TOTAL (At): diferença entre o maior e o menor valor observado.

AMPLITUDE DE CLASSE (Ai): diferença entre o maior valor e o menor valor da classe.

**p)** FREQÜÊNCIA ABSOLUTA: freqüência simples de cada classe (f). Indica o número de vezes em que aparece um determinado resultado.

**q)** FREQÜÊNCIA RELATIVA: razão entre a freqüência absoluta (simples) e a freqüência absoluta total, ou seja: fr = . É a freqüência que representa o percentual do que está sendo analisado. Indica o tamanho relativo de um conjunto de valores em termos de porcentagem.

**r)** FREQÜÊNCIA ACUMULADA (fA)

a) Freqüência Absoluta Acumulada “Abaixo de” (fA↓): é constituída, somando-se cada freqüência com a freqüência acumulada anterior. A expressão “abaixo de”, refere-se ao limite superior da respectiva classe de interesse.

b) Freqüência Absoluta Acumulada “Acima de” (fA↑): constitui-se pelo mesmo procedimento anterior, porém, partindo da última casa. A expressão “acima de”, refere-se ao limite inferior da respectiva classe de interesse.

**Obs.**: as freqüências acumuladas podem ser calculadas, utilizando-se as freqüências absolutas (f) ou relativas (fr). É utilizada especialmente quando tentamos situar um valor, uma classe, em função do desenvolvimento total da distribuição.

**Exemplo**

Sejam as verbas (R$) recebidas do governo estadual para a educação, de 25 cidades catarinenses em determinado ano (dados fictícios):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 52.500,00 | 18.300,00 | 35.700,00 | 43.800,00 | 22.150,00 |
| 6.830,00 | 3.250,00 | 17.603,00 | 35.600,00 | 7.800,00 |
| 16.323,00 | 42.130,00 | 27.606,00 | 18.350,00 | 12.521,00 |
| 25.300,00 | 31.452,00 | 39.610,00 | 22.450,00 | 7.380,00 |
| 28.000,00 | 21.000,00 | 14.751,00 | 39.512,00 | 17.319,00 |

Os dados assim estabelecidos são chamados de DADOS BRUTOS, que dificultam uma análise estatística rápida.

Para que esta análise seja possível organizamos um ROL, ou seja, os dados são colocados em ordem (crescente ou decrescente).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 3.250,00 | 14.751,00 | 18.350,00 | 27.606,00 | 39.512,00 |
| 6.830,00 | 16.323,00 | 21.000,00 | 28.000,00 | 39.610,00 |
| 7.380,00 | 17.319,00 | 22.150,00 | 31.452,00 | 42.130,00 |
| 7.800,00 | 17.603,00 | 22.450,00 | 35.600,00 | 43.800,00 |
| 12.521,00 | 18.300,00 | 25.300,00 | 35.700,00 | 52.500,00 |

Estes dados podem ser mais bem apresentados, ou seja, apresentar uma estética melhor. Para isso, aproximaremos os dados numa DISTRIBUIÇÃO DE FREQÜÊNCIA. Faremos assim, agrupamentos com as verbas.

**8.4. Como determinar o Número de Classes e os Intervalos de Classes?**

Uma das primeiras preocupações que temos, na construção de uma Distribuição de Freqüência, é a determinação do número de classes e, conseqüentemente, da amplitude e dos limites dos intervalos de classe.

O número de classes a ser utilizado depende muito da experiência do pesquisador e das questões que ele pretende responder com a variável contínua. Apesar da particularidade da pesquisa, podemos lançar mão de alguns critérios para a determinação do número de classes de uma distribuição.

**1o) Critério da Raiz** (para no máximo 20 dados) → se a sequencia estatística contém *n* elementos e se indicarmos por *i* o número de classes a ser utilizado, então temos:

i = 

Como o número *i* de classes deve ser necessariamente um número inteiro, e como dificilmente  é um número inteiro, deixaremos como opção para o valor de *i* o valor inteiro mais próximo de , uma unidade a menos ou a mais que este valor.

**2o) Regra de Sturges** → dá o número de classes em função do número de valores da variável. Indicado para valores de *n* muito grandes.

i = 1 + 3,3 ⋅ log n onde i = número de classes

n = número total de dados

**Obs. Para obtermos log n na calculadora financeira, usaremos ln n/ ln 10.**

Decidido o número de classes que deve ter a distribuição, resta-nos resolver o problema da determinação da Amplitude do Intervalo de Classe, o que conseguimos dividindo a amplitude total (At) pelo número de classes (i).

h = 

Determinação do número de classes do exemplo dado:

\* Critério da Raiz

i =  = 5 (podendo estar entre 4 e 6)

\* Regra de Sturges

i = 1 + 3,3 ⋅ log n

i = 1 + 3,3 ⋅ log 25

i = 1 + 3,3 ⋅ 1,3979

i = 1 + 4,6132

i = 5,6 ≅ 6

Escolhendo o número de 6 classes, podemos determinar a amplitude dos intervalos de classe do exemplo dado: h =  = 8.208,33 ≅ R$ 8.208,00

Com as verbas de R$ 3.250,00 a R$52.500,00 formamos então, intervalos de R$ 8.208,00 em R$8.208,00, obtendo classes de freqüência (teremos assim, 5 linhas com as classes):

Verba (R$) recebida do governo estadual para a educação, em 25 cidades catarinenses no ano “X”

|  |  |  |
| --- | --- | --- |
| **Ordem** | **Verba (R$)** | **f (freqüência)** |
| 1 | 3.250,00 11.458,00 | 4 |
| 2 | 11.458,00 19.666,00 | 7 |
| 3 | 19.666,00 27.874,00 | 5 |
| 4 | 27.874,00 36.082,00 | 4 |
| 5 | 36.082,00 44.290,00 | 4 |
| 6 | 44.290,00 52.498,00 | - |
| 7 | 52.498,00 60.706,00 | 1 |
| **Σ** | **-** | **25** |

Fonte: Dados Fictícios

Esta tabela, pode ser complementada com a FREQÜÊNCIA RELATIVA (fr), a FREQÜÊNCIA ABSOLUTA ACUMULADA “abaixo de” (fA↓) ­e “acima de” (fA↑), bem como com o PONTO MÉDIO (Xi) de cada classe, ou seja:

Verba (R$) recebida do governo estadual para a educação, em 25 cidades catarinenses no ano “X”

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Ordem** | **Verba (R$)** | **f** | **fr (%)** | **fA↓** | **frA↓(%)** | **fA­↑** | **frA­↑(%)** | **Xi** |
| 1 | 3.250,00 ├─ 11.458,00 | 4 | 16,0 | 4 | 16,0 | 25 | 100,0 | 7.354,00 |
| 2 | 11.458,00 ├─ 19.666,00 | 7 | 28,0 | 11 | 44,0 | 21 | 84,0 | 15.562,00 |
| 3 | 19.666,00 ├─ 27.874,00 | 5 | 20,0 | 16 | 64,0 | 14 | 56,0 | 23.770,00 |
| 4 | 27.874,00 ├─ 36.082,00 | 4 | 16,0 | 20 | 80,0 | 9 | 36,0 | 31.978,00 |
| 5 | 36.082,00 ├─ 44.290,00 | 4 | 16,0 | 24 | 96,0 | 5 | 20,0 | 40.186,00 |
| 6 | 44.290,00 ├─ 52.498,00 | - | - | 24 | 96,0 | 1 | 4,0 | 48.394,00 |
| 7 | 52.498,00 ├─ 60.706,00 | 1 | 4,0 | 25 | 100,0 | 1 | 4,0 | 56.602,00 |
| **Σ** | **-** | **25** | **100,0** | **-** | **-** | **-** | **-** | **-** |

Fonte: Dados Fictícios

Com esses dados podemos determinar:

\* Argumento: X = verbas

\* Intervalo: h = R$8.208,00

\* Frequência de classe: f = (4, 7, 5, 4, 4, -, 1) cidades

\* Classes de frequência: X = (3.250,00 11.458,00, 11.458,00 19.666,00, 19.666,00 27.874,00, 27.874,00 36.082,00, 36.082,00 44.290,00, 44.290,00 52.498,00, 52.498,00 60.706,00) reais

\* Limites inferiores: X1 = (3.250,00, 11.458,00, 19.666,00, 27.874,00, 36.082,00, 44.290,00, 52.498,00) reais

\* Limites superiores: X2 = (11.458,00, 19.666,00, 27.874,00, 36.082,00, 44.290,00, 52.498,00, 60.706,00) reais

\* Maior frequência: fmáx. = 7 cidades

\* Classe Modal: XMo = (11.458,00 19.666,00) reais

\* Amplitude total da tabela: At = 60.706,00 – 3.250,00 = 57.456,00 reais

**8.5. Representação Gráfica**

A distribuição de frequência pode ser representada graficamente através de duas formas:

**a)** HISTOGRAMAS: consistem numa área que abrange **todos** os elementos, formada por retângulos justapostos. Na base (eixo x) colocamos as classes de freqüência e na altura (eixo y), as freqüências absolutas (f).

Exemplo:

Verba (R$) recebida do governo estadual para a educação, em 25 cidades catarinenses no ano “X”

|  |  |  |
| --- | --- | --- |
| **Ordem** | **Verba (R$)** | **f** |
| 1 | 3.250,00 11.458,00 | 4 |
| 2 | 11.458,00 19.666,00 | 7 |
| 3 | 19.666,00 27.874,00 | 5 |
| 4 | 27.874,00 36.082,00 | 4 |
| 5 | 36.082,00 44.290,00 | 4 |
| 6 | 44.290,00 52.498,00 | - |
| 7 | 52.498,00 60.706,00 | 1 |
| **Σ** | **-** | **25** |

Fonte: Dados Fictícios

Fonte: Dados Fictícios

**b)** POLÍGONOS DE FREQÜÊNCIA: consistem num gráfico em curvas, que representa somente um elemento de cada classe (amostra). Na base colocamos os pontos médios (Xi) e na altura, as freqüências absolutas (f).

Exemplo:

Verba (R$) recebida do governo estadual para a educação, em 25 cidades catarinenses no ano “X”

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Ordem** | **Verba (R$)** | **f** | **fr (%)** | **fA↓** | **frA↓ (%)** | **fA↑** | **frA↑ (%)** | **Xi** |
| 1 | 3.250,00 11.458,00 | 4 | 16 | 4 |  | 25 |  | 7.354,00 |
| 2 | 11.458,00 19.666,00 | 7 | 28 | 11 |  | 21 |  | 15.562,00 |
| 3 | 19.666,00 27.874,00 | 5 | 20 | 16 |  | 14 |  | 23.770,00 |
| 4 | 27.874,00 36.082,00 | 4 | 16 | 20 |  | 9 |  | 31.978,00 |
| 5 | 36.082,00 44.290,00 | 4 | 16 | 24 |  | 5 |  | 40.186,00 |
| 6 | 44.290,00 52.498,00 | - | - | 24 |  | 1 |  | 48.394,00 |
| 7 | 52.498,00 60.706,00 | 1 | 4 | 25 |  | 1 |  | 56.602,00 |
| **Σ** | **-** | **25** | **100** | **-** |  | **-** | **-** | **-** |

Fonte: Dados Fictícios

Fonte: Dados Fictícios

Exercício!

**Observação:**

**\* Na tabela devem constar somente os resultados, mas, a parte apresente todos os desenvolvimentos (essa observação é válida para as duas questões).**

**\* Não esqueçam que a tabela deve ser construída dentro dos padrões das séries estatísticas;**

1) Abaixo estão apresentados às taxas de ocupação de determinado hotel, num período de 50 dias:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 12,0 | 25,0 | 44,5 | 49,3 | 69,0 |
| 12,5 | 25,4 | 45,8 | 58,6 | 69,3 |
| 14,0 | 25,8 | 45,8 | 58,9 | 69,5 |
| 14,3 | 27,9 | 45,8 | 58,9 | 73,2 |
| 16,8 | 28,6 | 47,0 | 62,0 | 78,2 |
| 18,2 | 28,9 | 47,8 | 62,3 | 78,5 |
| 18,7 | 32,6 | 48,5 | 63,0 | 88,4 |
| 24,0 | 36,0 | 48,5 | 65,0 | 95,0 |
| 24,4 | 36,8 | 48,7 | 66,8 | 97,0 |
| 25,0 | 42,3 | 48,9 | 68,2 | 98,9 |

Agrupe esses dados, tendo como limite inferior a menor medida da relação acima. A partir das classes estabelecer:

a) freqüência absoluta;

b) freqüência relativa;

c) freqüência absoluta acumulada “abaixo de” e “acima de”;

d) freqüência relativa acumulada “abaixo de” e “acima de”

e) ponto médio de cada classe;

f) maior freqüência;

g) classe modal;

h) argumento;

i) intervalo.

2)Considere os dados relativos ao gasto (R$) com deslocamento mensal de 45 acadêmicos

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 24,8 | 48,5 | 88,4 | 145,0 | 186,2 |
| 25,0 | 48,9 | 89,5 | 145,3 | 236,5 |
| 29,8 | 49,0 | 89,7 | 145,9 | 245,3 |
| 32,8 | 54,0 | 112,3 | 154,0 | 258,0 |
| 36,0 | 55,0 | 125,7 | 158,9 | 258,7 |
| 39,4 | 69,0 | 125,8 | 159,2 |  |
| 45,2 | 69,0 | 136,0 | 169,3 |  |
| 46,3 | 76,5 | 136,8 | 173,6 |  |
| 47,6 | 84,3 | 142,5 | 178,0 |  |
| 48,0 | 87,6 | 145,0 | 180,0 |  |

A partir desses dados, montem uma tabela agrupando os dados, sendo o limite inferior da 1a classe igual a 24,8. Apresente também:

a) freqüência absoluta;

b) freqüência absoluta acumulada “abaixo de” e “acima de”;

c) freqüência relativa;

d) freqüência relativa acumulada “abaixo de” e “acima de”;

e) ponto médio de cada classe.

**Observações: \* é necessário que a tabela esteja com todos os dados necessários para a resolução das questões;**

3) Uma amostra aleatória de 103 residências de famílias, classe média, com dois filhos, revelou a seguinte distribuição do consumo mensal de energia elétrica:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 53 | 142 | 167 | 195 | 225 | 241 | 258 | 274 | 287 | 298 | 349 |
| 84 | 147 | 168 | 201 | 226 | 242 | 259 | 275 | 288 | 299 | 350 |
| 98 | 149 | 168 | 201 | 226 | 245 | 263 | 276 | 288 | 300 | 354 |
| 100 | 150 | 168 | 202 | 237 | 245 | 263 | 277 | 289 | 321 |  |
| 113 | 151 | 173 | 211 | 237 | 250 | 264 | 277 | 290 | 325 |  |
| 132 | 151 | 173 | 215 | 238 | 250 | 268 | 278 | 291 | 327 |  |
| 134 | 152 | 173 | 215 | 238 | 251 | 269 | 280 | 295 | 338 |  |
| 138 | 154 | 179 | 216 | 239 | 251 | 269 | 280 | 296 | 342 |  |
| 138 | 156 | 184 | 217 | 239 | 251 | 270 | 283 | 296 | 345 |  |
| 139 | 166 | 188 | 223 | 239 | 254 | 270 | 284 | 298 | 349 |  |

Pede-se:

1. distribuição de freqüência;

b) freqüência absoluta acumulada “abaixo de” e “acima de”;

c) freqüência relativa;

d) freqüência relativa acumulada “abaixo de” e “acima de”;

e) ponto médio de cada classe;

f) leitura da 4a classe, com exceção do ponto médio.