
ESTATÍSTICA

Pelo Professor Geraldo Pacheco

A **Estatística** é uma parte da Matemática Aplicada que fornece métodos para coleta, organização, descrição e interpretação de dados e para utilização na tomada de decisões.

Fases do método estatístico:

1. Coleta de dados;
2. Crítica dos dados;
3. Apuração dos dados;
4. Exposição ou apresentação dos dados;
5. Análise dos dados.

1. Noções

Variável é, convencionalmente, o conjunto de resultados possíveis de um fenômeno.

População Estatística ou **Universo Estatístico** é formada por todos os elementos de um conjunto que têm pelo menos uma característica em comum – característica esta que representa interesse em determinado estudo estatístico.

Amostra é um subconjunto de elementos extraídos de uma população e pode ser

Finita: Quando apresenta um número finito de elementos e

Infinita: Quando apresenta um número infinito de elementos.

Variável é a característica estudada de uma população e podem ser:

Qualitativa: Expressa por atributos ou palavras, como raça, área de estudos;

Quantitativa discreta: Resultante de contagem, como número de sócios de um certo clube;

Quantitativa contínua: Proveniente de medida – número real, como determinar a altura dos adolescentes de uma escola.

2. Elementos de uma distribuição de freqüência

Classe de freqüência ou simplesmente, **classe** são intervalos de variação da variável.

Limite de classe: São os extremos de cada classe.

A tabela, com dados agrupados por intervalos ou não, que mostra a relação entre a variável e a freqüência é chamada de **tabela de freqüências** ou de **distribuição de freqüências**.

Freqüência Absoluta (f_i)

Freqüência Absoluta de cada variável é o número ou quantidade de vezes que cada valor é observado.

f_i lemos freqüência absoluta da classe i

Freqüência Relativa (f_r)

Freqüência Relativa de cada variável é a razão entre a freqüência absoluta e o número total de elementos.

$$f_r = \frac{f_i}{\sum f_i} = \frac{f_i}{n}, \text{ onde } n = \sum f_i \text{ e } n = \text{número total de dados}$$

Porcentagem ($f_r\%$)

$$f_r\% = (100f_r)\%$$

$$\text{Propriedades: } \sum f_r = 1 \text{ e } \sum f_r\% = 100\%$$

Freqüência absoluta acumulada (f_{ia}) corresponde à soma de cada freqüência absoluta com as freqüências absolutas anteriores.

Freqüência relativa acumulada (F_{ra}) corresponde à soma de cada freqüência relativa com as freqüências relativas anteriores.

3. Tipos de gráficos

Nos **Gráficos de barras** colocamos as frequências num eixo horizontal, enquanto nos **Gráficos de colunas** as frequências são colocadas num eixo vertical. Em ambos são usados retângulos, todos com a mesma largura, cujos comprimentos são proporcionais às frequências.

Nos **Gráficos de setores** desenhamos um círculo e o dividimos em setores que tenham áreas proporcionais às percentagens (ou frequências) ou seja 360° equivalem a **100%**.

A distribuição de frequências por intervalos é feita através dos **Histogramas de frequências absolutas** e dos **Polígonos de frequências**.

4. Distribuição de frequências com dados agrupados

Vamos exemplificar para entendermos melhor, o que desejo falar:

Se um radar instalado em rodovia federal, registrou as velocidades de 50 veículos. As velocidades, em Km. Estão indicadas neste quadro;

62	123	95	123	81	123	60	72	86	108
109	84	121	60	128	77	91	51	100	63
104	107	63	117	116	69	116	82	95	72
94	84	123	52	90	100	79	101	98	110
79	92	73	83	74	125	56	86	98	76

E se tentássemos elaborar o quadro de frequências utilizando esse dados, pouco ou nada poderíamos concluir, pois eles são muitos diferentes. Nesses casos, é interessante agrupa-los em **classes** ou **intervalos**, escolhendo-se, convenientemente, a **amplitude dos intervalos**.

No exemplo, podemos agrupar as velocidades em intervalos de qual amplitude?

Vamos utilizar algumas regras conhecidas (**Ver observação sobre classes e amplitude logo abaixo**), então teremos como solução:

$i = \sqrt{ht}$, **i = Número de Classes** e **ht = Amplitude total**, neste caso teremos:

ht = 123 – 51 = 72, portanto teremos um total de classes igual a **i = $\sqrt{72} = 8,485\dots$** , portanto **8 classes**, com **amplitude** $h = \frac{ht}{i}$ e iremos considerar a menor velocidade 50 km e a maior

velocidade 130 km então teremos com **amplitude** $h = \frac{130 - 50}{8} = \frac{80}{8} = 10$.

Obtemos, assim a seguinte **tabela de freqüências**:

CLASSES	VELOCIDADE (km)	f_i	f_{ia}	f_r	$f_r\%$	$f_{ra}\%$
1	[50;60[3	3	$\frac{3}{50} = 0,06$	6%	6%
2	[60;70[6	9	$\frac{6}{50} = 0,12$	12%	18%
3	[70;80[8	17	$\frac{8}{50} = 0,16$	16%	34%
4	[80;90[7	24	$\frac{7}{50} = 0,14$	14%	48%
5	[90;100[8	32	$\frac{8}{50} = 0,16$	16%	64%
6	[100;110[7	39	$\frac{7}{50} = 0,14$	14%	78%
7	[110;120[4	43	$\frac{4}{50} = 0,08$	8%	86%
8	[120;130[7	50	$\frac{7}{50} = 0,14$	14%	100%
Total		50		1	100%	

Observação:

- O ponto que divide o intervalo de classe em duas partes iguais é denominado **Ponto Médio do Intervalo**. Por exemplo, a velocidade dos veículos na quinta classe [90;100[pode ser representada por: $x_5 = \frac{90+100}{2} = 95km$.
- O intervalo real $[a;b[$ também é representado, em **Estatística**, pela notação $a |— b$
- **Amplitude (h) de um Intervalo de Classe** ou, simplesmente **Intervalo de Classe** é a medida do intervalo que define a classe ou seja ela é obtida pela diferença entre os limites superior e inferior dessa classe, no exemplo, a amplitude da 1ª classe [50;60[do exemplo anterior é $h = 60 - 50 = 10$.

$$h = \frac{ht}{i}$$

ht = Amplitude Total da Distribuição é a diferença entre o limite superior da última classe (**Limite Superior Máximo**) e o limite inferior da primeira classe (**Limite Inferior Mínimo**)

- **Números de Classes (i)**

a) Regras de **STURGES**

$$i = 1 + 3,3 \log n$$

i = Número de Classes e n = Número total de dados

b) O Número de Classes **há quem prefira** : $i = \sqrt{ht}$ onde ht = Amplitude total

“Os sonhos de DEUS jamais vão morrer”

www.geraldopacheco.blogspot.com

Considerações Finais:

1. Sempre que possível, evitar classe com frequência nula ou frequência relativa muito exagerada.
2. Quando o resultado não é exato, devemos arredondá-lo para mais.

5. Arredondamentos de dados

Muitas vezes, é necessário ou conveniente suprimir unidades inferiores às de determinada ordem, esta técnica denominada **Arredondamentos de Dados**.

A norma **NBR 5891** da **Associação Brasileira de Normas Técnicas (ABNT)** estabelece as regras de arredondamento na numeração decimal, em uso na atualidade.

Essas regras estão de acordo com a resolução **886/66** do **IBGE**, são elas:

a) Quando o primeiro algarismo a ser abandonado no arredondamento é **0, 1, 2, 3** ou **4**, fica inalterado o último algarismo a permanecer.

Ex: a) 25,32 → 25,3 b) 3,201 → 3,20

b) Quando o primeiro algarismo a ser abandonado no arredondamento é **6, 7, 8** ou **9**, aumenta-se uma unidade ao último algarismo a permanecer.

Ex: 19,417 → 19,42 b) 2,09 → 2,1

c) Quando o primeiro algarismo a ser abandonado no arredondamento é 5, há dois procedimentos:

- Se após o algarismo 5 seguir em qualquer casa um número diferente de 0, aumenta-se em uma unidade o algarismo que antecede o 5;

Ex: 237,85001 → 237,9 b) 5,5256 → 5,53

- Se após o algarismo 5 não seguir (em qualquer casa) um número diferente de 0, ao algarismo que antecede o 5 será acrescentada uma unidade, se for ímpar, e permanecerá como está, se for par;

Ex: a) 246,35 → 246,4 b) 246,85 → 246,8

Observação:

Nos *softwares* de computadores (como o **Excel**) e calculadoras científicas, porém, não é aplicado o critério indicado no item c. Nesses casos, se o primeiro algarismo a ser abandonado for ao algarismo 5, o arredondamento será feito com o aumento de uma unidade ao algarismo que antecede o 5.

Ex: 246,35 → 246,4 b) 246,85 → 246,9

6. Medidas de Tendências central

São medidas estatísticas que descrevem a tendência que os dados têm de se agruparem em torno de certos valores centrais.

Médias

Média (M) de uma distribuição é a média aritmética dos valores dados.

Podem ser:

$$\text{Média Aritmética Simples} = \mathbf{M} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum_{i=1}^n X_i}{n}$$

$$\text{Média Aritmética Ponderada} = \mathbf{M} = \frac{X_1 P_1 + X_2 P_2 + \dots + X_n P_n}{P_1 + P_2 + \dots + P_n} = \frac{\sum_{i=1}^n X_i P_i}{\sum_{i=1}^n P_i}$$

Moda

Moda (M_o) de uma distribuição é o valor que nela aparece com maior frequência.

Cálculo da **Moda** para dados agrupados com intervalos de classe:

Fórmula CZUBER

$$M_o = L_i + \left(\frac{d_1}{d_1 + d_2} \right) \cdot h$$

M_o = Valor Modal ou Moda de distribuição de frequências

L_i = Limite inferior da classe modal

d_1 = Frequência simples da classe Modal – Frequência simples anterior à classe modal

d_2 = Frequência simples da classe Modal – Frequência Posterior à classe modal

h = Amplitude da classe modal

“Os sonhos de DEUS jamais vão morrer”

www.geraldopacheco.blogspot.com

Mediana

Mediana (M_d) de uma distribuição é o valor que ocupa a posição central, quando todos os valores são colocados em ordem

A **Mediana** é designada, muitas vezes, por **Valor Mediano**

Cálculo da **Mediana** para dados agrupados com intervalos de classe:

$$M_d = L_i + \left(\frac{\frac{\sum f_i}{2} - f_{anterior}}{f_{M_d}} \right) \cdot h$$

$$\text{Classe Mediana} = \frac{\sum f_i}{2}$$

M_d = Valor da Mediana

L_i = Limite inferior da classe mediana

$f_{anterior}$ = Frequência acumulada da classe anterior à classe mediana

f_{M_d} = Frequência simples da classe mediana

h = Amplitude do intervalo da classe mediana

7. Medidas de Dispersão ou de variabilidade

$$\text{Desvio médio} = D_m = \frac{|X_1 - M| + |X_2 - M| + \dots + |X_n - M|}{n} = \frac{\sum_{i=1}^n |X_i - M|}{n}$$

$$\text{Variância} = V_a = \frac{(X_1 - M)^2 + (X_2 - M)^2 + \dots + (X_n - M)^2}{n} = \frac{\sum_{i=1}^n (X_i - M)^2}{n}$$

$$\text{Desvio Padrão} = D_p = \sqrt{V_a}$$

Importante:

Tanto o **desvio padrão** com a **variância** são usados como medidas de dispersão ou variabilidade. O uso de uma ou de outra dependerá da finalidade que se tenha em vista. A **variância** é uma medida que tem pouca utilidade como **estatística descritiva**, porém é extremamente importante na **inferência estatística** e em **combinações de amostras**.
