



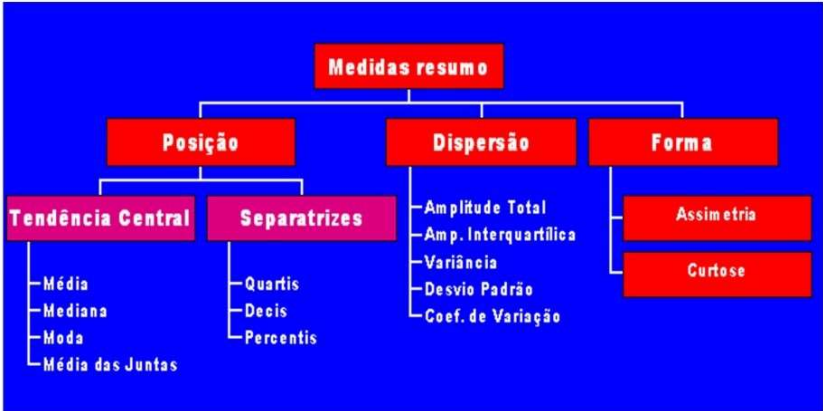
Probabilidade e Estatística aplicada à Engenharia

Profa. Marília Miranda

- Medidas de posição e dispersão para dados brutos



MEDIDAS RESUMO



NOTAÇÃO PADRÃO

Medida	Amostra (Estatística)	População (Parâmetro)
Média	\bar{X}	μ
Desvio padrão	S	σ
Variância	S^2	σ^2
Tamanho	n	N



Probabilidade e Estatística aplicada à Engenharia

Profa. Marília Miranda

- Distribuição de frequência
- Histograma



DISTRIBUIÇÕES DE FREQUÊNCIA

- A **distribuição de frequência** é uma tabela que mostra classes ou intervalos das entradas de dados de uma contagem do número de entradas em cada classe.
- As **classes** são pequenos intervalos mutuamente exclusivos, tais que, quando reunidos, abrangem todo o conjunto de dados.
- A **frequência $f(i)$** de uma classe é o número de entrada de dados em uma classe.



DISTRIBUIÇÕES DE FREQUÊNCIA

- Dados brutos: 24, 26, 24, 21, 27, 27, 30, 41, 32, 38

Classe	Frequência
15 — 25	3
25 — 35	5
35 — 45	2



Probabilidade e Estatística aplicada à Engenharia

Profa. Marília Miranda

- Medidas de posição e dispersão para dados agrupados (Parte I)



RELEMBRANDO...

- Dados brutos: 24, 26, 24, 21, 27, 27, 30, 41, 32, 38

Classe	Frequência
15 — 25	3
25 — 35	5
35 — 45	2



MEDIDAS PARA TABELAS DE FREQUÊNCIA

- Média e Desvio Padrão ponderados pelas frequências:

$$\bar{X} = \frac{\sum (X \cdot f)}{n}$$

$$S = \sqrt{\frac{\sum (X^2 \cdot f) - n \cdot \bar{X}^2}{n - 1}}$$



MÉDIA

- Medida de tendência central;
- Medida mais comum;
- Funciona como um “ponto de equilíbrio”;
- Afetada por valores extremos ('outliers');
- Definição: soma dos valores, dividida pelo número de valores observados.

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$



MÉDIA

Exemplo:

- Notas finais dos alunos de três turmas

Turma	Notas dos alunos	Média da turma
A	4 5 5 6 6 7 7 8	6,00
B	1 2 4 6 6 9 10 10	6,00
C	0 6 7 7 7 7,5 7,5	6,00

PEDRO A. BARBETTA – Estatística Aplicada às Ciências Sociais 6ed. Editora da UFSC, 2006.



DISTRIBUIÇÕES DE FREQUÊNCIA

- Conceitos importantes:
 - (1) *Limite inferior de classe*: é o menor número que pode pertencer à classe;
 - (2) *Limite superior de classe*: é o maior que pode pertencer à classe;
 - (3) *Largura (ou amplitude) da classe* - h : distância entre os limites inferiores (ou superiores) de suas consecutivas classes;



DISTRIBUIÇÕES DE FREQUÊNCIA

- Conceitos importantes:
 - (4) *Amplitude Total* - A : diferença entre o maior e o menor valor observado no conjunto de dados.



COEFICIENTE DE VARIAÇÃO

- Medida de dispersão relativa;
- Pode ser expresso como uma %;
- Mostra a variação relativa a média;
- Usado para comparar 2 ou mais grupos.
- Fórmula:

$$CV = \left(\frac{s}{\bar{X}} \right) \cdot 100\%$$



EXEMPLO

- Encontre a idade média das crianças com diarreia crônica da tabela abaixo

Idade (Meses)	Número de crianças
0 — 20	69
20 — 40	26
40 — 60	13
60 — 80	9
80 — 100	6
100 — 140	7
Total	130

Fonte: Dr. Paulo Costa



VARIÂNCIA E DESVIO PADRÃO

- São medidas de dispersão;
- Considera como os dados estão distribuídos;
- Mostra a variação em torno da média;
- Para calcular estas medidas, deve-se considerar os desvios de cada valor em relação à média aritmética. Depois obtém-se uma média desses desvios.



DISTRIBUIÇÕES DE FREQUÊNCIA

- Características adicionais que podem ser incluídas em uma tabela de frequências:
 - (1) *Ponto médio de uma classe* - $x(i)$: soma dos limites inferiores e superiores da classe dividida por dois.
 - exemplo: classe = 0 | - 10
 - ponto médio = $(0+10)/2 = 5$



EXEMPLO

$$\begin{aligned} x_1 &= \frac{0+20}{2} = 10 \quad e \quad f_1 = 69 & x_4 &= \frac{60+80}{2} = 70 \quad ; \quad f_4 = 9 \\ x_2 &= \frac{20+40}{2} = 30 \quad ; \quad f_2 = 26 & x_5 &= \frac{80+100}{2} = 90 \quad ; \quad f_5 = 6 \\ x_3 &= \frac{40+60}{2} = 50 \quad ; \quad f_3 = 13 & x_6 &= \frac{100+140}{2} = 120 \quad ; \quad f_6 = 7 \end{aligned}$$

A média é:

$$\bar{x} = \frac{69 \cdot 10 + 26 \cdot 30 + 13 \cdot 50 + 9 \cdot 70 + 6 \cdot 90 + 7 \cdot 120}{130} = \frac{4130}{130} = 31,77$$



VARIÂNCIA E DESVIO PADRÃO

Descrição	Notação	Resultados numéricos	Soma
Valores (notas dos alunos)	\bar{X}	4 5 5 6 7 7 8	48
Média	\bar{X}	6	
Desvios	$X - \bar{X}$	-2 -1 -1 0 0 1 1 2	0
Desvios quadráticos	$(X - \bar{X})^2$	4 1 1 0 0 1 1 4	12

$$S^2 = \frac{\sum (X - \bar{X})^2}{n - 1}$$



$$S^2 = (4 + 1 + 1 + 0 + 0 + 1 + 1 + 4) / 7 = 12 / 7 = 1,71$$

O desvio padrão (**S**) é a raiz quadrada da variância. $S = \sqrt{1,71} = 1,31$

PEDRO A. BARBETTA - Estatística Aplicada às Ciências Sociais 6ed. Editora da UFSC, 2006.



MODA

- Medida de tendência central;
- Valor que ocorre com **maior frequência** em uma frequência ordenada;
- Não é afetado por valores extremos;
- Pode não existir moda (série amodal) como pode existir várias modas (bimodal ou multimodal);
- Pode ser usada para dados quantitativos e qualitativos.



MEDIANA

- Medida de tendência central;
- Divide a distribuição ao meio, deixando os 50% menores valores de um lado e os 50% maiores valores do outro
- Não é afetada por valores extremos.



QUARTIS



DISTRIBUIÇÕES DE FREQUÊNCIA

(2) *Frequência relativa de uma classe – fr(i):* porção ou porcentagem de dados que está em determinada classe.

- Para calculá-la, divide a frequência **f(i)** pelo tamanho **n** da amostra.

- exemplo: 0 |- 10 → f(i) = 20

$$n = 160$$

$$\text{freq. relativa} = 20/160$$

$$= 0,125$$



DISTRIBUIÇÕES DE FREQUÊNCIA

(3) *Frequência acumulada de uma classe – F(i):* é a soma da frequência para aquela classe e todas as anteriores.

- a frequência acumulada da última classe é igual ao tamanho n da amostra.



DISTRIBUIÇÕES DE FREQUÊNCIA

Tabela 1.

Notas da primeira prova de Estatística Aplicada
Universidade Fictícia, 2010

Classes de Notas	fi	xi	fri	fri%	Fi
0 -- 10	5	5	0,01	1	5
10 -- 20	15	15	0,03	3	20
20 -- 30	20	25	0,04	4	40
30 -- 40	45	35	0,09	9	85
40 -- 50	100	45	0,20	20	185
50 -- 60	130	55	0,26	26	315
60 -- 70	100	65	0,20	20	415
70 -- 80	60	75	0,12	12	475
80 -- 90	15	85	0,03	3	490
90 -- 100	10	95	0,02	2	500
-	$\sum_{i=1}^{10} fi = 500$	-	-	-	-

Fonte: www.supercatalogo.com.br/unit/Distribuicao_de_Frequencia.pdf



HISTOGRAMA



EXEMPLO

- E qual seria o valor do desvio padrão?

$$\sum f_i \cdot x_i^2 = 69 \cdot 10^2 + 26 \cdot 30^2 + 13 \cdot 50^2 + 9 \cdot 50^2 + 6 \cdot 90^2 + 7 \cdot 120^2$$

$$\sum f_i \cdot x_i^2 = 256 \ 300$$

- Substituindo na fórmula abaixo, encontramos $S = 31,1$ (CV = 97,8%)

$$S = \sqrt{\frac{\sum (X^2 \cdot f) - n \cdot \bar{X}^2}{n - 1}}$$

- Medida de tendência *não central*;
- Divide os dados ordenados em 4 quartos



Q_1 : separa os 25% inferiores dos 75% superiores;

Q_2 : é a mediana;

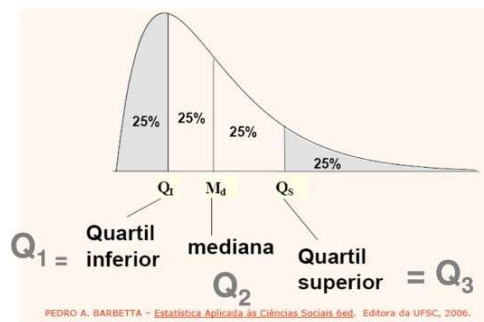
Q_3 : separa os 75% inferiores dos 25% superiores



QUARTIS

- Posição do i -ésimo quartil:

$$\text{Posição } Q_i = \frac{i \cdot (n+1)}{4}$$



QUARTIS

- Exemplo:

Dados Brutos: 10.3 4.9 8.9 11.7 6.3 7.7

Ordenado: 4.9 6.3 7.7 8.9 10.3 11.7

Posição: 1 2 3 4 5 6

$$Q_1 \text{ Posição} = \frac{1 \cdot (n+1)}{4} = \frac{1 \cdot (6+1)}{4} = 1.75 \approx 2$$

$$Q_1 = 6.3$$



QUARTIS

- Exemplo:

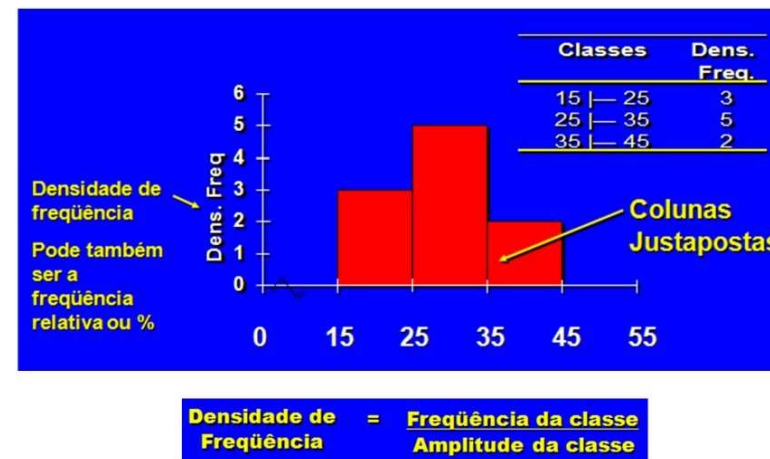
- Um **histograma** é um diagrama de barras que representa a distribuição de frequência de um conjunto de dados.

- Propriedades:

- a escala horizontal é quantitativa e mede os valores dos dados;
- a escala vertical mede as frequências das classes;
- as barras consecutivas devem estar encostadas umas nas outras (colunas justapostas).



HISTOGRAMA



Dados Brutos: 10.3 4.9 8.9 11.7 6.3 7.7
 Ordenado: 4.9 6.3 7.7 8.9 10.3 11.7
 Posição: 1 2 3 4 5 6

$$Q_2 \text{ Posição} = \frac{2 \cdot (n+1)}{4} = \frac{2 \cdot (6+1)}{4} = 3.5$$

$$Q_2 = \frac{7.7 + 8.9}{2} = 8.3$$



QUARTIS

- Exemplo:

Dados Brutos: 10.3 4.9 8.9 11.7 6.3 7.7
 Ordenado: 4.9 6.3 7.7 8.9 10.3 11.7
 Posição: 1 2 3 4 5 6

$$Q_3 \text{ Posição} = \frac{3 \cdot (n+1)}{4} = \frac{3 \cdot (6+1)}{4} = 5.25 \cong 5$$

$$Q_3 = 10.3$$



AMPLITUDE TOTAL

- É uma medida que mede variabilidade (dispersão);
- Diferença entre o maior e o menor valor observado:

$$\text{Amplitude Total} = X_{\text{máx}} - X_{\text{mín}}$$

- Para encontrar a amplitude, os dados devem ser quantitativos;
- Não considera a distribuição dos dados.



AMPLITUDE INTERQUARTÍLICA

- É uma medida de dispersão que pode ser utilizada para comparar grupos de dados;
- Também conhecida como dispersão central;

central;

- Dispersão dos 50% centrais;
- Não afetado por valores extremos.
- Fórmula:

$$\textit{Amplitude Interquartílica} = Q_3 - Q_1$$



COEFICIENTE DE VARIAÇÃO

- Medida de dispersão relativa;
- Pode ser expresso como uma %;
- Mostra a variação relativa a média;
- Usado para comparar 2 ou mais grupos.
- Fórmula:

$$CV = \left(\frac{S}{\overline{X}} \right) \cdot 100\%$$



