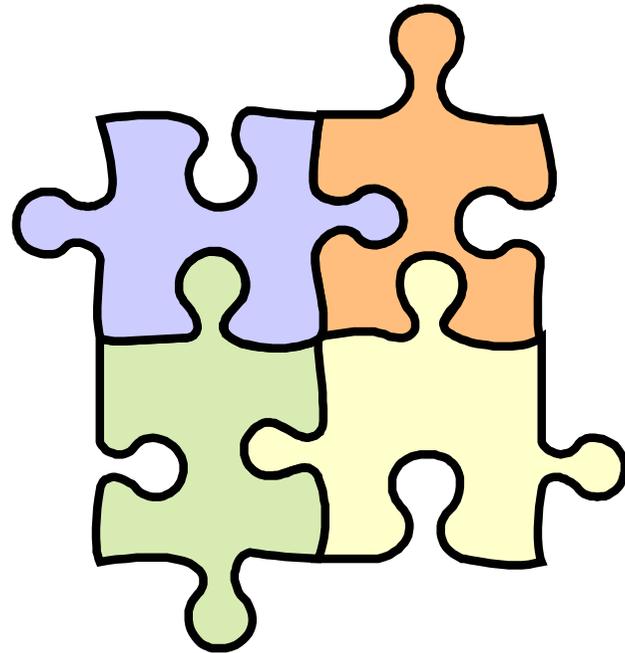


Estatística Básica



Armando Oscar Cavanha Filho

1- INTRODUÇÃO

A Estatística tem ampliado a sua participação na linguagem das atividades profissionais da atualidade, já que os números e seus significados traduzem, de forma objetiva, as questões do cotidiano, propiciando análises com base em fatos e DADOS. A seguir são apresentados alguns conceitos e definições úteis sobre Estatística:

- A Estatística compreende a coleta, a apresentação e a caracterização da informação, visando assistir a análise de dados e o processo de decisão.
- A Estatística Descritiva envolve a coleta, a análise e a apresentação de conjuntos de dados, para descrever as diversas características destes conjunto de dados.
- A Estatística Inferencial consiste nos métodos de estimativas de uma população com base nos estudos sobre amostras (por vezes é impossível trabalhar com a população inteira).
- A População (Universo) é a totalidade dos itens que estão sendo considerados.
- A Amostra é a parte da população que é selecionada para análise.
- População finita é aquela que possui um limite quantitativo (exemplo: a produção de veículos no país, ou no mundo, a cada ano), enquanto a infinita se

refere de quantitativos sem limite (exemplo: todos os resultados, cara ou coroa, dos lances de uma moeda qualquer).

- O Parâmetro é uma medida sintética que descreve um estado da população.
- Os Dados podem ser do tipo Qualitativos ou Quantitativos (Discretos ou Contínuos).
- Os Dados Quantitativos Discretos são aqueles que podem ser contados (exemplo: número de peças de roupa).
- Os Dados Quantitativos Contínuos são os que podem ser medidos. Estão limitados pela precisão do sistema de medição (exemplo: altura ou peso de um indivíduo).
- Números randômicos ou aleatórios são valores tomados sem nenhuma lei de formação, normalmente obtidos de uma tabela apropriada ou gerados eletronicamente por microprocessadores.
- Estudos Enumerativos envolvem a tomada de decisão, com base nas características de uma população sob análise (ex. Votações políticas).
- Estudos Analíticos envolvem a tomada de uma ação sobre um processo visando o aumento de performance no futuro (ex. Processo de fabricação de peças de automóveis).

2 – MEDIDAS DE TENDÊNCIA CENTRAL

I - MÉDIA ARITMÉTICA

A média aritmética de um conjunto é o resultado da divisão entre a soma de todos os valores pela quantidade de valores.

$$Xa = \frac{\sum Xi}{N}$$

Exemplo: Dados os valores 5,4,8,9,2,1, qual a média aritmética do conjunto dado ?

X1	5
X2	4
X3	8

X4	9
X5	2
X6	1
Soma	29
Soma / n	4,833333

A média aritmética deste conjunto é 4,83.

II - MÉDIA GEOMÉTRICA

A média geométrica é o resultado da raiz enésima do produto de todos os valores dados.

$$X_g = \sqrt[n]{X_1 \cdot X_2 \dots X_n}$$

Exemplo: Dados os valores 5,4,8,9,2,1, qual a média geométrica do conjunto dado ?

		produto
		1
X1	5	5
X2	4	20
X3	8	160
X4	9	1440
X5	2	2880
X6	1	2880
	produto	7385
	Produto $^{(1/n)}$	4,41291

A média geométrica deste conjunto é 4,41.

III - MÉDIA HARMÔNICA

A média harmônica é o recíproco da média aritmética dos recíprocos de todos os valores dados.

$$X_h = \frac{1}{\frac{1/X_1 + 1/X_2 + \dots + 1/X_N}{N}}$$

Exemplo: Dados os valores 5,4,8,9,2,1, qual a média harmônica do conjunto dado ?

		Inverso
X1	5	0,2
X2	4	0,25
X3	8	0,125
X4	9	0,111111
X5	2	0,5
X6	1	1

	Soma	2,186111
	Soma/n	0,364352
	1/soma/n	2,7446

A média harmônica deste conjunto é 2,74.

IV - MÉDIA PONDERADA

A média ponderada é a média (aritmética, harmônica ou geométrica) que leva em conta o quanto cada valor é mais confiável ou mais frequente no conjunto de dados.

Média aritmética ponderada:

$$X_w = \frac{X_1.f_1 + X_2.f_2 + \dots + X_N.f_n}{f_1+f_2+\dots+f_n}$$

Exemplo: Dados:

Valor	Frequência ou peso
5	3
4	4
8	1
9	2
2	2
1	8

qual a média aritmética ponderada do conjunto dado ?

X1	5	f1	3	x1f1	15
X2	4	f2	4	x2f2	16
X3	8	f3	1	x3f3	8
X4	9	f4	2	x4f4	18
X5	2	f5	2	x5f5	4
x6	1	f6	8	x6f6	8

		soma f	20	soma xf	69
				soma xf / soma f	3,45

A média aritmética ponderada deste conjunto é de 3,45.

Média geométrica ponderada:

$$X_{gw} = \left\{ (X_1^{f_1} \cdot X_2^{f_2} \cdot X_3^{f_3} \cdot X_n^{f_n}) \right\}^{(1 / (f_1+f_2+\dots+f_n))}$$

Exemplo: Dados:

Valor	Frequência ou peso
5	3
4	4
8	1
9	2
2	2

1	8
---	---

qual a média geométrica ponderada do conjunto dado ?

						1
x1	5	f1	3	$x1^{f1}$	125	125
x2	4	f2	4	$x2^{f2}$	256	32000
x3	8	f3	1	$x3^{f3}$	8	256000
x4	9	f4	2	$x4^{f4}$	81	20736000
x5	2	f5	2	$x5^{f5}$	4	82944000
x6	1	f6	8	$x6^{f6}$	1	82944000
		som a f	20		Produto	82944000
					Prod ^ (1/ som f)	2,488509

A média geométrica ponderada deste conjunto é 2,48.

Média harmônica ponderada:

$$X_{hw} = \frac{1}{\frac{1}{X_1 \cdot f_1} + \frac{1}{X_2 \cdot f_2} + \dots + \frac{1}{X_N \cdot f_N}}$$
$$\frac{1}{f_1 + f_2 + \dots + f_n}$$

Exemplo: Dados:

Valor	Frequência ou peso
5	3
4	4
8	1
9	2
2	2
1	8

qual a média harmônica ponderada do conjunto dado ?

x1	5	f1	3	x1f1	15	1/x1f1	0,07
x2	4	f2	4	x2f2	16	1/x2f2	0,06
x3	8	f3	1	x3f3	8	1/x3f3	0,13
x4	9	f4	2	x4f4	18	1/x4f4	0,06
x5	2	f5	2	x5f5	4	1/x5f5	0,25
x6	1	f6	8	x6f6	8	1/x6f6	0,13
		soma f	20			soma 1/xf	0,68
						Som/n	0,03
						1/so/n	29,21

A média harmônica ponderada deste conjunto é 29,21.

V - PONTO MÉDIO DO INTERVALO

É a média aritmética dos extremos.

Exemplo: Dados:

Valor
5
4
8
9
2
1

Qual o ponto médio do intervalo ?

Valores em ordem crescente
1
2

4
5
8
9

$$Pmi = (1 + 9) / 2 = 5$$

O ponto médio do intervalo deste conjunto é 5.

VI - MEDIANA

É o valor que se encontra na posição média da série ordenada de dados. Para quantidade de dados par, é a média dos dois valores de posição média da série ordenada.

Exemplo: Dada a série:

Valor
5
4
8
9
2
1

Ordena-se a série:

valores em ordem crescente
1
2
4
5
8
9

A mediana deste conjunto é $(4+5) / 2 = 4,5$.

Se a série dada for:

Valor
5
4
8
9
2
1
7

Então:

valores em ordem crescente
1
2
4
5
7

8
9

A mediana deste conjunto é 5.

VII - MODA

É o valor de mais alta frequência de aparecimento na série dada.

Exemplo: Dada a série:

Valor	Frequência ou peso
5	3
4	4
8	1
9	2
2	2
1	8

A moda deste conjunto é 1 (aparece oito vezes).

Caso os dados fossem:

Valor	Frequência ou peso
5	3
4	4
8	1
9	8
2	2
1	8

Existiriam duas modas para este conjunto de dados: 9 e 1, cada uma com 8 aparecimentos (série bi-modal).

3- MEDIDAS DE DISPERSÃO

Trata-se da medida de variação ou distribuição dos dados. Uma série de dados pode ter os valores mais “aproximados” ou mais “espalhados” que outra.

I - FAIXA

É a diferença entre o maior e o menor valor da série.

Exemplo: Dada a série:

Valor
5
4
8
9
2
1

Então:

valores em ordem crescente
1
2
4
5
8
9

$$\text{FAIXA} = 9 - 1 = 8$$

A faixa deste conjunto de dados é 8.

II - VARIÂNCIA e DESVIO PADRÃO

Para a amostra

É aproximadamente a média aritmética dos quadrados dos desvios de todos os valores sobre a média aritmética da população.

$$\sigma^2(x) = \frac{\sum(x-m)^2}{n-1}$$

onde x é cada valor da série, m é a média aritmética da série.

Exemplo: Dada a série seguinte (amostra), qual a variância ?

Valor
5
4
8
9
2
1

Então:

	Valor	va-me	(va-me)^2
	5	0,17	0,03
	4	-0,83	0,69
	8	3,17	10,03
	9	4,17	17,36
	2	-2,83	8,03
	1	-3,83	14,69
Soma	29		50,83
Média arit	4,83	varian	10,17

A variância deste conjunto é 10,17.

O Desvio Padrão é a raiz quadrada da variância

No exemplo = $10,17^{(1/2)} = 3,19$.

Para a população

$$s^2(x) = \frac{\sum(x-m)^2}{n}$$

Exemplo: Dada a série seguinte (população), qual a variância ?

Valor
5
4
8
9
2
1

	Valor	va-me	(va-me)^2
	5	0,17	0,03

	4	-0,83	0,69
	8	3,17	10,03
	9	4,17	17,36
	2	-2,83	8,03
	1	-3,83	14,69
soma	29		50,83
média arit	4,83	varian	8,47

A variância deste conjunto é 8,47.

O Desvio Padrão é a raiz quadrada da variância.

No exemplo, $8,47^{(1/2)} = 2,91$

III - COEFICIENTE DE VARIAÇÃO

É a razão entre o desvio padrão e a média aritmética, apresentada na forma percentual:

$$CV = \frac{100 \times s(x)}{m}$$

Exemplo: Dada a série seguinte, qual o coeficiente de variação ?

Valor
5
4
8
9
2
1

	Valor	dif	dif^2
	5,00	0,17	0,03
	4,00	-0,83	0,69
	8,00	3,17	10,03
	9,00	4,17	17,36
	2,00	-2,83	8,03
	1,00	-3,83	14,69
soma	29,00		50,83
media	4,83	varian	10,17
		Des pad	3,19

$$CV = \frac{100 \times 3,19}{4,83} = 66,04\%$$

IV-DESVIO MÉDIO

É a média dos desvios absolutos da amostra com relação à média.

Exemplo: Dada a série seguinte, qual o desvio médio ?

Valor
5
4
8
9
2
1

	Valor	Dif	mod dif
	5,00	0,17	0,17
	4,00	-0,83	0,83
	8,00	3,17	3,17
	9,00	4,17	4,17

	2,00	-2,83	2,83
	1,00	-3,83	3,83
Soma	29,00		15,00
Media	4,83		2,50

O desvio médio deste conjunto é 2,50.

4- HISTOGRAMAS

São gráficos em barras que apresentam a dispersão dos dados ao longo da faixa de existência.

Após ordenar a série dada, divide-se a série em grupos em intervalos iguais, quantificando-os, criando-se então o gráfico. É normal ter-se a quantidade de classes entre 4 e 12. Esta escolha é subjetiva e toma-se este número em função do gráfico que se deseja obter.

Exemplo: Dada a série seguinte, traçar o histograma para 4 classes.

Valor
5
4
8
9
2
1

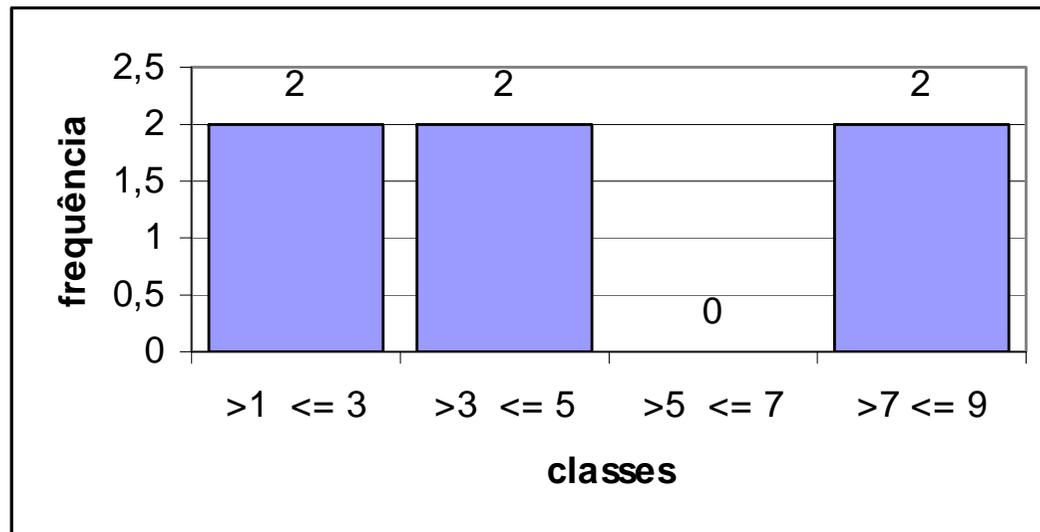
Série ordenada:

Valor
1
2
4
5
8
9

A faixa é $9-1=8$, e o intervalo, para 4 classes, $8/4=2$

Portanto:

Classe	Valores
$>1 \leq 3$	1,2
$>3 \leq 5$	4,5
$>5 \leq 7$	
$>7 \leq 9$	8,9



5 - FREQUÊNCIA RELATIVA

Frequência relativa de uma classe é a frequência dessa classe dividida pela de todas elas, em percentagem.

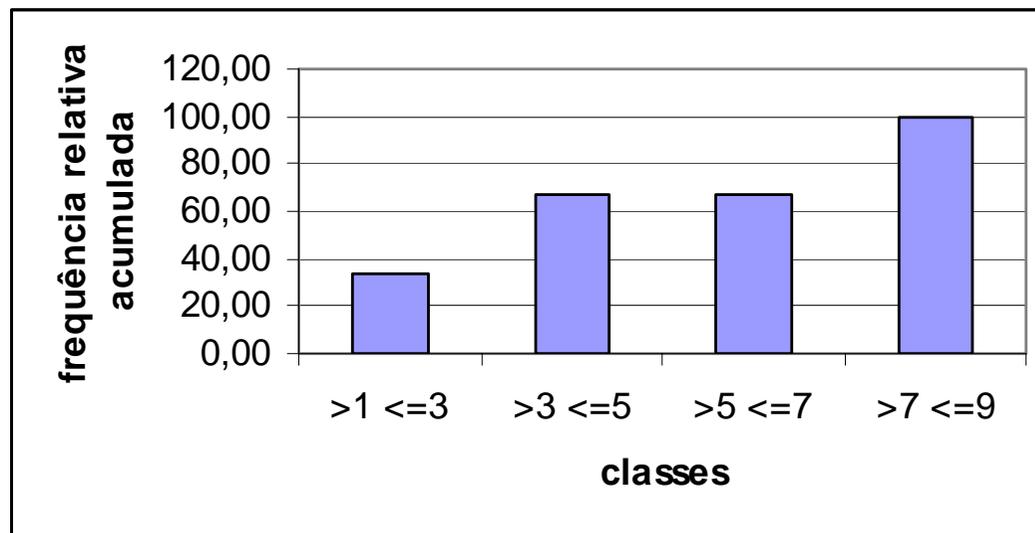
classe	valores	freq relat %
>1 <=3	1,2	33,33333
>3 <=5	4,5	33,33333
>5 <=7		
>7 <=9	8,9	33,33333
		100

FREQUÊNCIA ACUMULADA

Frequência acumulada é o gráfico de frequência de ocorrências, por classe, apresentado de forma cumulativa.

classe	valores	freq relat	freq relat
--------	---------	------------	------------

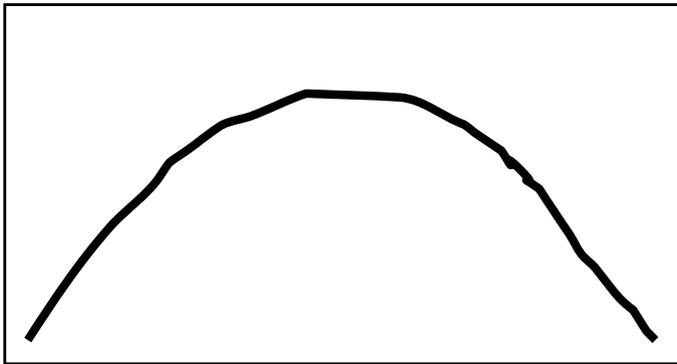
			acumul
>1 <=3	1,2	33,33333	33,33333
>3 <=5	4,5	33,33333	66,66667
>5 <=7		0	66,66667
>7 <=9	8,9	33,33333	100



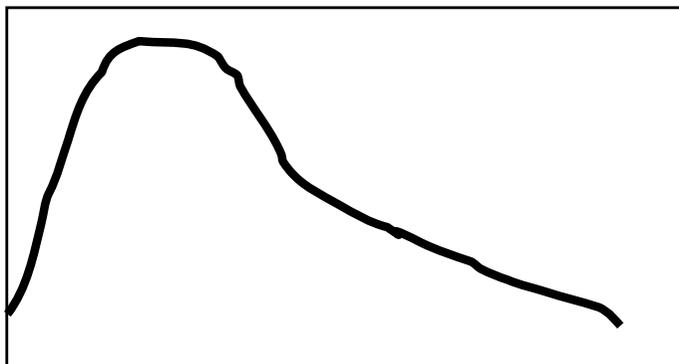
6 - TIPOS DE CURVAS DE FREQUÊNCIA

Segundo a forma resultante do traçado da distribuição de freqüências, podem aparecer diversos tipos de curvas, como:

Curvas simétricas, em forma de sino, normais:



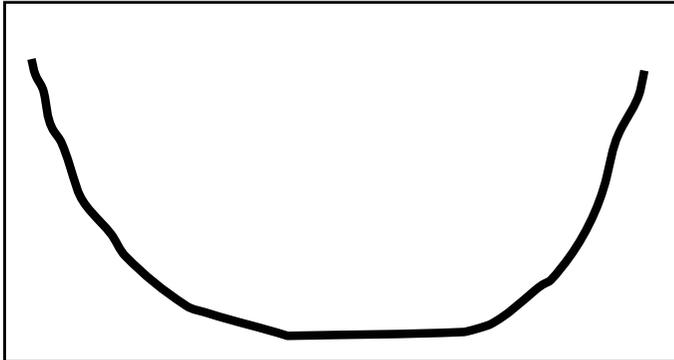
Curvas desviadas para a esquerda:



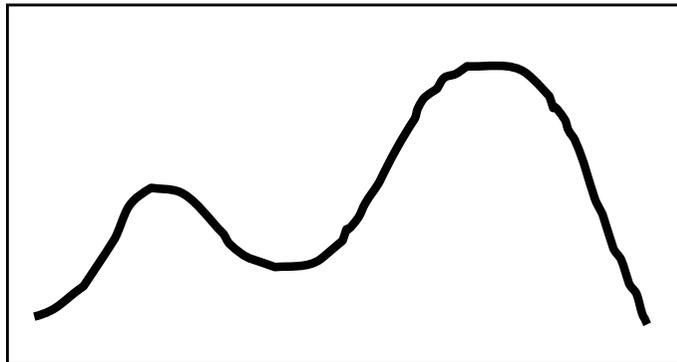
Curvas desviadas para a direita:



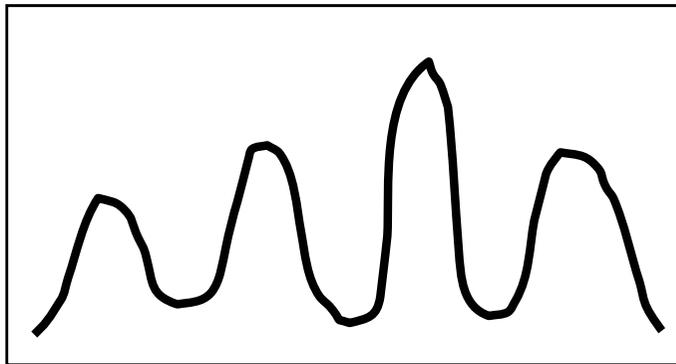
Curvas em forma de U:



Curvas com mais de uma moda, bimodal:



Curvas com mais de uma moda, multimodal:



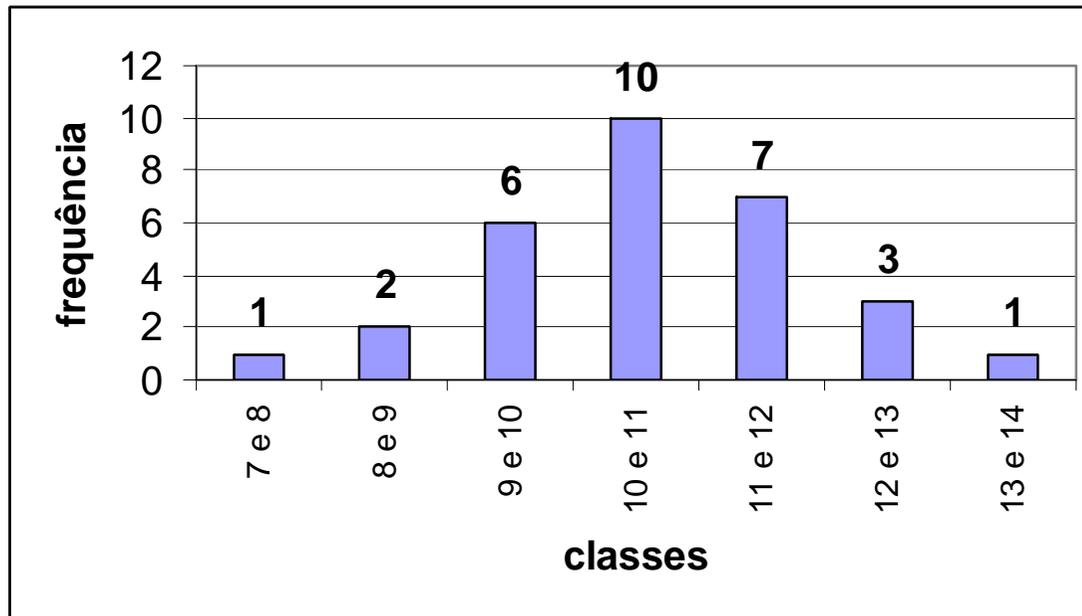
Exemplo para curva normal: Dados os valores seguintes, obter o histograma.

10	9	9
11	11	7
11	10	13
11	10	10
9	8	11
9	12	9

10	8	10
10	12	10
10	10	11
9	11	12

Resultados:

media 10,1
des pad 1,322224
classes 6
intervalo de classes 1,00



A curva resultante se assemelha a uma normal, pois o seu formato parece um sino.

O cálculo das diversas medidas de tendência central e medidas de dispersão pode auxiliar na análise dos dados, mostrando que o posicionamento da média, moda e mediana se alterna em função do tipo de curva:

