

# Como analisar de forma simples um grande número de dados?

Uma multinacional planeja instalar sua próxima fábrica em Portugal. A escolha da localização desta fábrica depende de informações. Os seus executivos desejam saber, entre outras coisas, se há ou não mão-de-obra especializada na região; qual é a renda média dos moradores para desenvolver uma política de salários; e qual a capacidade de consumo do mercado português, pois disso dependerá o preço final de suas mercadorias. Da mesma maneira, um pesquisador que reúne muitos dados sobre uma determinada população tem, em seguida, de agrupá-los e classificá-los, ou trabalhar esses dados. Esses dados trabalhados podem ser usados nos estudos de determinadas situações e acontecimentos. Trata-se de um método de análise e elaboração científica dos dados, que chamamos de Estatística.

A Estatística é um ramo da Matemática que estuda os fenômenos do acaso, agrupando, classificando e ordenando experiências e observações sobre as manifestações de um fenômeno para extrair suas consequências. A Estatística tem seu desenvolvimento marcado pelas mais recentes descobertas científicas, como a Teoria da Relatividade, do físico Albert Einstein, e a Física Quântica. Além disso, tem-se configurado como um importante instrumento de apoio para variados campos do conhecimento, como os da Psicologia, da Economia, da Sociologia e da Medicina.

---

## 1. Estatística descritiva

O estudo estatístico não é feito tendo como referência uma única pessoa ou um fato isolado, e sim um conjunto de elementos. Esse conjunto é chamado de colectivo ou de população.

- Se o colectivo ou população é formado por muitos elementos, usa-se uma amostra muito selectiva que seja representativa do total. Cada elemento desse coletivo é chamado de indivíduo ou unidade estatística.
- Chamamos variável estatística uma característica da população. Ela pode ser **quantitativa** ou **qualitativa**, caso seja ou não numérica.

As pesquisas realizadas para estudar nível de instrução, religião ou preferência musical são exemplos de variáveis qualitativas. Já as pesquisas que envolvem estatura, número de habitantes e idade são exemplos de variáveis quantitativas.

## 2. Frequências absolutas

Para obter uma informação clara e precisa de uma série de dados estatísticos numéricos, devemos primeiro ordená-los. A essa ordenação chamamos de rol.

Exemplo:

4, 5, 1, 6, 8, 5, 6, 5, 7, 1,

4, 6, 9, 6, 7, 5, 4, 2, 5, 4,

6, 5, 6, 4, 5, 6, 6, 3, 4, 8,

5, 6, 4, 5, 7, 4, 6, 5, 10, 6

Perguntamos, numa turma de 40 alunos, as notas que eles obtiveram numa prova de Matemática. Recebemos como resposta as notas do quadro ao lado.

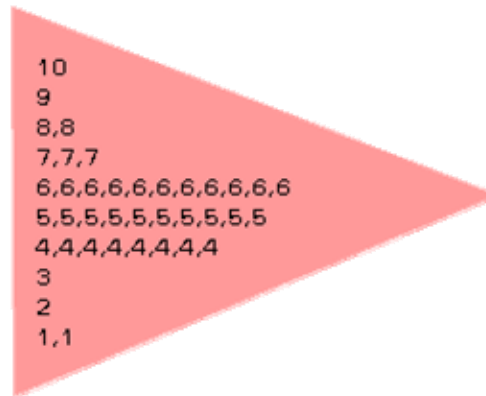
O conjunto das notas obtidas na prova de Matemática é o que chamamos de dados estatísticos

numéricos.

- Com esses dados, podemos deduzir algumas características:

Há apenas uma nota 10, nenhum 0 e muitas notas 5 e 6.

Se ordenarmos as notas da maior para a menor, como na Figura 1, abaixo, podemos obter outras informações:



- A nota mais alta foi um 10 e somente uma das pessoas da classe a obteve.
- A nota mais baixa foi 1 e dois alunos a receberam.
- Dez pessoas receberam 5 e 11 obtiveram 6.

Observamos que 10 alunos obtiveram nota 5 e que um recebeu um 10. De outra forma, podemos dizer que a nota 5 aparece com frequência 10' e a nota 10, com frequência 1'. Portanto:

A frequência absoluta de um resultado ou um dado é o número de vezes que esse dado aparece.

Podemos, agora, fazer uma tabela de duas colunas mostrando os dados e sua frequência. À esquerda, colocamos as notas e, à direita, as frequências absolutas respectivas:

Dados	Frequência absoluta	Interpretação estatística
1	2	a frequência de 1 é 2
2	1	a frequência de 2 é 1
3	1	a frequência de 3 é 1
4	8	a frequência de 4 é 8
5	10	a frequência de 5 é 10
6	11	a frequência de 6 é 11
7	3	a frequência de 7 é 3
8	2	a frequência de 8 é 2
9	1	a frequência de 9 é 1
10	1	a frequência de 10 é 1

A soma das frequências da tabela acima é 40. Observe como esta cifra coincide com o número total de dados.

A soma de todas as frequências absolutas é o número total de elementos da população, que é indicado pela letra **n**.

### 3. Frequência acumulada

A frequência acumulada é um dado igual à soma das frequências deste e a de todos os dados anteriores.

Exemplo:

Quantos alunos obtiveram nota abaixo de 5?

Contamos o número de alunos que tiraram esta nota:

$$2 + 1 + 1 + 8 = 12$$

Agora nos perguntamos quantos alunos obtiveram uma nota menor ou igual a 7.

$$2 + 1 + 1 + 8 + 10 + 11 + 3 = 36$$

Para responder de maneira mais simples a esse tipo de pergunta (quantos alunos obtiveram uma nota menor ou igual a 7), vamos criar uma nova coluna correspondente à da frequência acumulada na tabela anterior de frequências absolutas (Figura 1, abaixo).

Dados	Frequência absoluta	Frequência acumulada
1	2	2
2	1	3
3	1	4
4	8	12
5	10	22
6	11	33
7	3	36
8	2	38
9	1	39
10	1	40

Figura 1

Utilizemos agora esta nova tabela nos seguintes exemplos:

Quantos alunos obtiveram uma nota abaixo de 7?

$$2 + 1 + 1 + 8 + 10 + 11 + 3 = 36$$

Quantos alunos mereceram uma nota menor do que 8?

$$36$$

### 4. Frequências relativas

Chamamos de frequência relativa de um dado ao quociente entre a frequência com que o dado aparece (ou frequência absoluta) e o número total de dados:

$$\text{Frequência relativa} = \frac{\text{Frequência absoluta}}{\text{Número de indivíduos da população}}$$

Como indica a Tabela de frequências absoluta e acumulada:

- A soma das frequências absolutas é 40:

$$2 + 1 + 1 + 8 + 10 + 11 + 3 + 2 + 1 + 1 = 40$$

- A soma das frequências relativas é  $40/40 = 1$ :

$$2/40 + 1/40 + 1/40 + 8/40 + 10/40 + 11/40 + 3/40 + 2/40 + 1/40 + 1/40 = 40/40$$

Dados	Frequência absoluta	Frequência relativa
1	2	$2/40 = 0,050$
2	1	$1/40 = 0,025$
3	1	$1/40 = 0,025$
4	8	$8/40 = 0,200$
5	10	$10/40 = 0,250$
6	11	$11/40 = 0,275$
7	3	$3/40 = 0,075$
8	2	$2/40 = 0,050$
9	1	$1/40 = 0,025$
10	1	$1/40 = 0,025$

- Para lembrar:

A soma das frequências relativas é igual a 1, qualquer que seja o número total de observações, pois a soma dos numeradores é igual ao denominador comum.

## 5. Gráficos

Comprovamos que as tabelas de frequências dos dados estatísticos nos dão uma informação boa e ordenada do exemplo que estudamos.

Muitas vezes, no entanto, queremos ter uma visão generalizada e rápida. Por isso, os gráficos estatísticos são muito úteis para entender e comparar várias tabelas de frequências.

Continuando com o mesmo exemplo empregado até aqui, vamos agora apresentar graficamente os dados. Podemos fazê-lo de várias formas. As mais comuns são: o diagrama de barras, o histograma, o pictograma e o gráfico de sectores.

- Diagrama de barras ou colunas

O diagrama de barras ou colunas é utilizado na apresentação de variáveis qualitativas. Ele é composto por rectângulos dispostos verticalmente (em colunas) ou horizontalmente (em barras).

Caracteriza-se por ter todos os rectângulos de mesma largura e os comprimentos proporcionais às frequências (Figura 2, abaixo).

Exemplo:

seguinte gráfico de colunas:

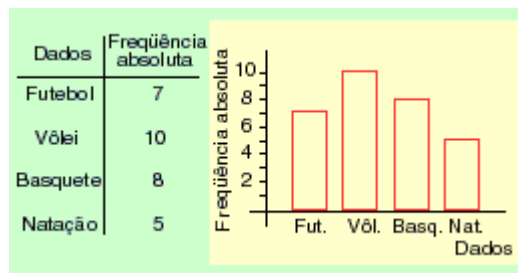


Figura 2

### ◆ Histograma

Os histogramas são diagramas de barras utilizados para variáveis quantitativas. São formados por retângulos justapostos.

A frequência correspondente a cada classe ou intervalo é representada pela superfície de um retângulo, cuja base, situada no eixo horizontal, é limitada pelos valores extremos.

A altura de cada retângulo é proporcional à frequência de cada classe ou intervalo.

### Exemplo:

Medimos a estatura dos 40 meninos e meninas da classe de nosso exemplo anterior. As medidas obtidas, em centímetros, como mostra a tabela abaixo são:

Altura dos alunos em centímetros

168 160 168 175 175 160 165 154 163 165  
 168 168 158 149 160 161 162 166 163 159  
 178 169 158 163 171 170 165 150 167 164  
 162 165 163 156 174 165 173 172 168 168

Como as alturas são muito diferentes e quase não coincidem umas com as outras, vamos agrupá-las em intervalos. É o que mostra a tabela abaixo:

Tabela de frequências (I)	
Intervalo	Frequência
de 148,5 a 153,5	2
de 153,5 a 158,5	4
de 158,5 a 163,5	11
de 163,5 a 168,5	14
de 168,5 a 173,5	5
de 173,5 a 178,5	4

Em seguida, representamos esses dados na Tabela de frequências II, abaixo, e no histograma (Figura 3, abaixo).

Tabela de frequências (II)	
Intervalo	Frequência
Altos de 168,5 a 178,5	9
Médios de 158,5 a 168,5	25
Baixos de 148,5 a 158,5	6

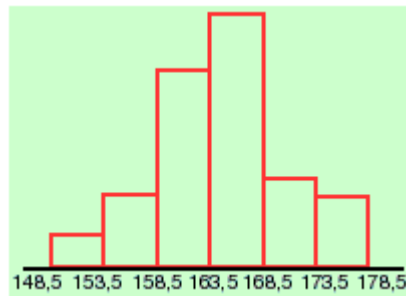


Figura 3

Dado um histograma, se unirmos os pontos médios das extremidades superiores de cada um dos rectângulos, obteremos uma linha poligonal chamada polígono de frequências, como a Figura 4, abaixo:

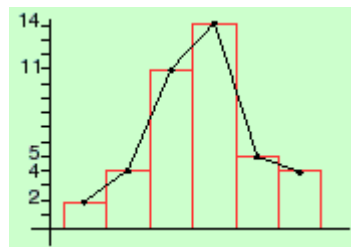


Figura 4

#### ➤ Pictograma

Outra forma de representar dados estatísticos é o pictograma.

- Trata-se de um gráfico em que são usados desenhos que têm relação directa com a área que está sendo pesquisada. Na Figura 5, por exemplo, o pictograma indica a fabricação de veículos, em três anos.
- Os pictogramas não são muito precisos e, por isso, são pouco utilizados pelos especialistas. Mas eles têm a vantagem de serem fáceis de visualizar e muito simples de interpretar.



Figura 5

➔ Gráfico de áreas ou gráfico de sectores

São gráfico circulares, como o da Figura 6, abaixo, utilizados para representar determinada população:

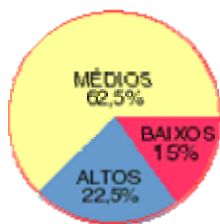


Figura 6

- Cada característica dessa população ocupa um sector circular do gráfico, proporcional à sua frequência. Os gráficos de sectores representam o conjunto de todos os dados expressos pela área de um círculo.
- Esses gráficos são bastante adequados quando os dados são classificados em poucas categorias. Costumam ser utilizados quando há poucos intervalos e são especialmente úteis para se estabelecerem comparações.

Exemplo:

Vamos comparar as estruturas do grupo de alunos (apresentadas anteriormente, no exemplo do item Histograma) distribuídas em alunos altos, médios e baixos.

Dividimos os 360 graus do ângulo central do círculo proporcionalmente às frequências. Como a soma das frequências é 40, faremos corresponder este número a 360° (Figura 6).

À frequência dos altos corresponderão:

$$360 \times 9/40 = 81 \text{ (22,5\%)}$$

À frequência dos médios corresponderão:

$$360 \times 25/40 = 225 \text{ (62,5\%)}$$

À frequência dos baixos corresponderão:

$$360 \times 6/40 = 54 \text{ (15\%)}$$

## 6. Estatísticas ajustadas e confiáveis

Muita gente se pergunta se é possível que as estatísticas, mesmo ajustadas e apresentadas de acordo com um determinado interesse particular, continuem sendo confiáveis.

A resposta para essa questão é positiva: as estatísticas podem continuar confiáveis.

Vamos comprová-la com o seguinte caso:

Exemplo:

Um gerente de vendas de uma livraria resolve impressionar seu chefe para obter um aumento de salário. Para tanto, elabora um gráfico estatístico das vendas realizadas no ano anterior, como mostra a tabela abaixo.

Vendas de livros no ano	
Meses do ano	Volumes vendidos
Janeiro	704 363
Fevereiro	707 450
Março	710 300
Abril	714 250
Maio	722 600
Junho	725 230
Julho	730 750
Agosto	736 125
Setembro	740 875
Outubro	743 500
Novembro	747 248
Dezembro	749 100

O resultado é apresentado no gráfico da Figura 7, abaixo:

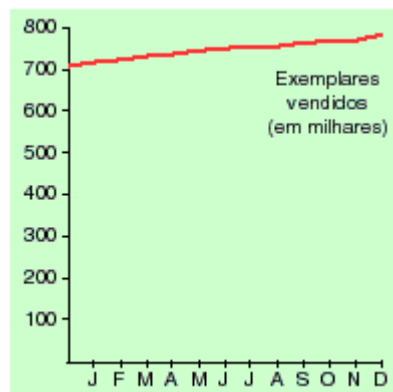


Figura 7

O efeito visual desse gráfico, com certeza, não tem muito impacto. Ele indica, ao contrário do desejado, que as vendas permaneceram praticamente estáveis durante todo o ano. Com ele, qualquer pretensão de aumento salarial não se justificaria.

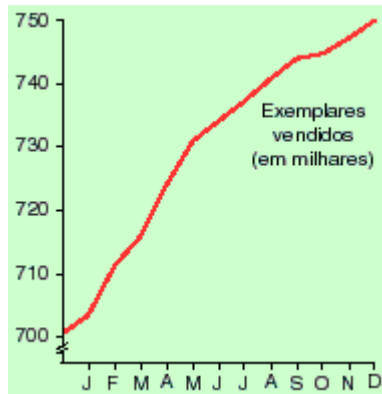


Figura 8

Mas o gerente de vendas não se dá por vencido. Ele faz um segundo gráfico, usando os mesmos dados, que mostra uma imagem completamente diferente (Figura 8, acima). Esse gráfico, sem dúvida, é muito mais favorável aos seus interesses do que o anterior.

E, embora ele apresente os dados de outra maneira (ampliando e focalizando apenas o espaço de vendas entre 700 mil e 750 mil exemplares), não é menos fiel à realidade do que o gráfico da Figura 6.

## 7. Medidas de posição central e de dispersão

Uma das principais tarefas da Estatística é resumir, classificar e extrair o máximo de informações de uma grande quantidade de dados. Os cálculos que estudaremos a seguir são importantes instrumentos matemáticos que permitem obter um conhecimento maior dos dados.

### ➔ Média aritmética

A média aritmética de vários números é obtida somando-se todos e dividindo a soma entre o total de dados. Ela é representada por  $\bar{X}$ .

$$\bar{X} = \frac{n_1 + n_2 + n_3 + n_4 + \dots + n_k}{K}$$

Exemplo:

As idades dos alunos de uma turma são:

Idades	Alunos
10	4
11	5
12	3
14	4
15	1

Calcule a idade média dos alunos dessa turma.

$$4 \times 10 + 5 \times 11 + 3 \times 12 + 4 \times 14 + 1 \times 15 =$$

$$40 + 55 + 36 + 56 + 15 = 202$$
$$202/17 = 11,88 \text{ idade média} = 11,88 \text{ anos}$$

Portanto, a idade média dos alunos é:

11,88 anos

#### ➤ Desvio de uma média

Quando calculamos a média de vários números, pode acontecer de alguns cálculos terem grandes desvios da média e outros apresentarem desvios mínimos.

- Como podemos medir esse grau de dispersão? Uma das maneiras é calcular o valor médio e, outra, assinalar as diferenças entre cada um dos valores observados, em relação ao valor médio. Quanto maior for essa diferença ou desvio da média, maior será a dispersão.

#### Exemplo:

As idades de 7 alunos de uma turma são:  
15, 15, 14, 14, 14, 13 e 13 anos.

A idade média desse grupo é:

$$2 \times 15 + 3 \times 14 + 2 \times 13 = 30 + 42 + 26 = 98$$
$$98/7 = 14 \text{ anos}$$
$$\text{Idade média} = 14 \text{ anos}$$

As idades de um segundo grupo de alunos são:

16, 16, 15, 14, 13, 12 e 12 anos

A idade média do grupo é:

$$2 \times 16 + 1 \times 15 + 1 \times 14 + 1 \times 13 + 2 \times 12 = 32 + 15 + 14 + 13 + 24 = 98$$
$$98/7 = 14 \text{ anos}$$
$$\text{Idade média} = 14 \text{ anos}$$

Quanto se desviam da média as idades de um e de outro grupo?

As diferenças entre as idades do primeiro grupo e a média são: 1, 1, 0, 0, 0, 1 e 1.

As diferenças entre as idades do segundo grupo e a média são: 2, 2, 1, 0, 1, 2 e 2.

Como no segundo grupo, as diferenças entre as idades em seus valores absolutos são maiores, as idades correspondentes apresentam um desvio maior em relação à média. Portanto, no segundo grupo a dispersão de idades é maior. Por outro lado, as idades dos alunos do primeiro grupo estão mais concentradas nas proximidades da média.

➔ Para lembrar:

Quando conhecemos a média de uma distribuição e os desvios de cada um dos valores em relação a essa média, a soma desses desvios é 0 (zero).

O cálculo da dispersão é importante, já que nos permite estabelecer qual grupo é mais ou menos homogêneo.

➔ Mediana

No caso da média, os valores extremos influem muito e podem produzir uma informação falsa. Portanto, torna-se necessário introduzir outra medida: a mediana.

➔ Para lembrar:

Quando uma série de números está ordenada em forma crescente:  $n_1, n_2, n_3, \dots, n_k$ , a mediana é o valor que ocupa a posição central da série disposta em ordem crescente ou decrescente.

Exemplo:

As notas dos alunos de uma classe são:

6, 5, 5, 7, 4, 6, 9, 0, 4, 6, 8, 7, 3, 5 e 8

Organizando essas notas em ordem crescente, temos:

0, 3, 4, 4, 5, 5, 5

7 notas

6

mediana

6, 6, 7, 7, 8, 8, 9

7 notas

Como achamos a mediana nas séries de dados de números pares?

0, 4, 5, 5, 5, 6, 7, 8, 8, 9

Consideram-se os dois valores centrais e a mediana será o valor médio dos dois números centrais:

$$\text{Mediana} = (5 + 6)/2 = 11/2 = 5,5$$

### ➤ Moda

Corresponde ao termo que tem maior frequência numa série de valores.

#### Exemplo:

Na série de valores 0, 2, 5, 3, 3, 1, 2, 3, 0, 7, 3, 5 a moda é 3, já que o número 3 tem maior frequência, pois aparece 4 vezes.

### ➤ Variância e desvio padrão

Como vimos anteriormente, a soma dos desvios de uma série sempre será igual a zero. Procurando contornar isso, os matemáticos decidiram elevar ao quadrado cada um desses desvios.

Para tanto, fizeram o cálculo do valor médio, acharam as diferenças entre os valores observados e o valor médio (desvio). Finalmente, elevaram ao quadrado as diferenças obtidas e as somaram.

### ➤ Variância

Chama-se variância ao quociente entre o valor desta soma (das diferenças obtidas entre os valores observados e o valor médio, ou desvio, elevadas ao quadrado) e o número **n** de elementos da série, que é indicada pela letra **V** ou pelo símbolo  $\sigma^2$ .  
Portanto, a fórmula é:

$$V = \frac{(\text{soma dos desvios})^2}{n}$$

A variância tem um inconveniente: seu valor fica muito grande ou, ao contrário, muito pequeno. Além disso, a dispersão é dada no quadrado da unidade dos dados.

Assim, para obter um dado na mesma unidade de medida, definiu-se uma outra medida de dispersão: o desvio padrão.

### ➤ Desvio padrão

O desvio padrão é a raiz quadrada do valor da variância. Ele é indicado por **DP** ou pela letra grega  $\sigma$ .

$$DP = \sqrt{V}$$

Portanto:

$$DP = \sqrt{\frac{(\text{soma dos desvios})^2}{n}}$$

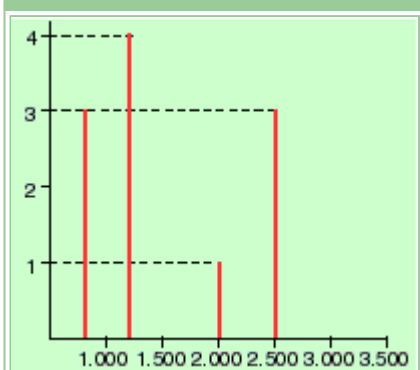
- O desvio padrão (**DP**) é sempre um número positivo ou 0 (zero). Quando o **DP** = 0, não há dispersão.
- Todos os valores estão concentrados no valor médio. Quanto maior for o **DP**, maior será a dispersão.

## EXERCÍCIOS

1. Numa caixa, há 25 lápis de cor. O número de lápis de cada cor é: 4 vermelhos, 3 azuis, 6 verdes, 7 amarelos e 5 castanhos. Assinalar o conjunto estatístico e completar a seguinte tabela:

Variável estatística (cor)	Frequência absoluta	Frequência relativa
1)		
2)		
3)		
4)		
5)		

2. O diagrama seguinte representa o que um grupo de 10 pessoas gasta mensalmente. Observe-o atentamente. Que tipo de diagrama é este?



- Quantas pessoas gastam 2500 euros por mês?
- Quantas pessoas gastam 2000 euros por mês?
- Quantas pessoas gastam 1200 euros por mês?
- Quantas pessoas gastam 700 euros por mês?

3. Os pesos, em quilogramas, de 13 jovens são: 51, 53, 49, 52, 50, 48, 54, 51, 49, 52, 53, 51 e 52. Calcular a média dos pesos e sua moda.