

Estatística

Estatística é uma área da matemática que tem por objetivo levantar dados de determinado fenômeno ou população, assim, relacionando fatos e números.

A **estatística** é o campo da matemática que **relaciona fatos e números** em que há um conjunto de métodos que nos possibilita coletar dados e analisá-los, assim sendo possível realizar alguma interpretação deles. A estatística é dividida em duas partes: **descritiva** e **inferencial**. A estatística descritiva é caracterizada pela organização, análise e apresentação dos dados, enquanto a estatística inferencial tem como característica o estudo de uma amostra de determinada população e, com base nela, a realização de análises e a apresentação de dados.

Leia também: [O que é margem de erro de uma pesquisa?](#)

Princípios da estatística

Veremos, a seguir, os principais conceitos e princípios da estatística. Com base neles, será possível definir conceitos mais sofisticados.

- **População ou universo estatístico**

A população ou universo estatístico é o **conjunto formado por todos elementos** que participam de um determinado tema pesquisado.

Exemplos de universo estatístico

a) Em uma cidade, todos os habitantes pertencem ao universo estatístico.

b) Em um dado de seis faces, a população é dada pelo número de faces.

$\{1, 2, 3, 4, 5, 6\}$

- **Dado estatístico**

O dado estatístico é um **elemento que pertence ao conjunto da população**, obviamente esse dado deve estar envolvido com o tema da pesquisa.

População	Dado estatístico
Dado de seis faces	4
Campeões Brasileiros de Mountain Bike	Henrique Avancini

- **Amostra**

Chamamos de amostra o **subconjunto formado com base no universo estatístico**. Uma amostra é utilizada quando a população é muito grande ou infinita. Em casos em que coletar todas as informações do universo estatístico é inviável por motivos financeiros ou logísticos, também se faz necessário a utilização de amostras.

A escolha de uma amostra é de extrema importância para uma pesquisa, e ela deve representar de maneira fidedigna a população. Um exemplo clássico da utilização das amostras em uma pesquisa é na realização do [censo demográfico](#) do nosso país.

- **Variável**

Em estatística, a variável é o objeto de estudo, isto é, **o tema que a pesquisa pretende estudar**. Por exemplo, ao estudar-se as características de uma cidade, o número de habitantes pode ser uma variável, assim como o volume de chuva em

determinado período ou até mesmo a quantidade de ônibus para o transporte público. Note que o conceito de variável em estatística é dependente do contexto da pesquisa.

A organização dos dados em estatística dá-se em **etapas**, como em todo processo de organização. Inicialmente é escolhido o tema a ser pesquisado, em seguida, é pensado o método para a coleta dos dados da pesquisa, e o terceiro passo é a execução da coleta. Após o fim dessa última etapa, faz-se a análise do que foi coletado, e assim, com base na interpretação, busca-se resultados. Veremos, agora, alguns conceitos importantes e necessários para a organização dos dados.

• **Rol**

Em casos em que os dados podem ser representados por números, ou seja, quando a variável é quantitativa, utiliza-se o rol para **organização desses dados**. Um rol pode ser crescente ou decrescente. Caso uma variável não seja quantitativa, ou seja, caso seja qualitativa, não é possível utilizar-se o rol, por exemplo, se os dados são sentimentos sobre determinado produto.

Exemplo

Em uma sala de aula, foram coletadas as alturas dos alunos em metros. São elas: 1,70; 1,60; 1,65; 1,78; 1,71; 1,73; 1,72; 1,64.

Como o rol pode ser organizado de maneira crescente ou decrescente, segue que:

$$\text{rol: } \{1,60; 1,64; 1,65; 1,70; 1,71; 1,72; 1,73; 1,78\}$$

Observe que, com o rol já montado, é possível encontrar um dado com mais facilidade.

• **Tabela de distribuição de frequência**

Em casos nos quais há muitos elementos no rol e muitas repetições de dados, o rol torna-se obsoleto, pois a organização desses dados é inviável. Nesses casos, as tabelas e a [distribuição de frequências](#) servem como uma excelente ferramenta de organização.

Na tabela de distribuição de **frequência absoluta**, devemos colocar a frequência em que cada dado aparece, ou seja, a quantidade de vezes que ele aparece.

Vamos construir a tabela de distribuição de [frequência absoluta](#) das idades, em anos, dos alunos de uma determinada classe.

Distribuição de frequências absolutas	
Idade	Frequência (F)
8	2
9	12
10	12
11	14
12	1
Total (F_T)	41

Da tabela podemos obter as seguintes informações: na classe temos 2 alunos com a idade de 8 anos, 12 alunos com 9 anos, e mais 12 alunos com 10 anos, e assim sucessivamente, alcançando o total de 41 alunos. Na tabela de distribuição de **frequências acumuladas**, devemos somar a frequência da linha anterior (na tabela de distribuição de frequência absoluta).

Vamos construir a tabela de distribuição de frequência acumulada das idades da mesma classe do exemplo anterior, veja:

Distribuição de frequências acumuladas	
Idade	Frequência (F)
8	2
9	14
10	26
11	40
12	41
Total (F _T)	41

Na tabela de **distribuição de frequências relativas**, utiliza-se a porcentagem em que cada dado aparece. Novamente faremos os cálculos baseados na tabela de distribuição de frequência absoluta. Sabemos que 41 corresponde a 100% dos alunos da classe, logo, para determinar a [porcentagem](#) de cada idade, basta dividirmos a frequência da idade por 41 e multiplicarmos o resultado por 100, para, assim, escrevermos na forma de porcentagem.

$2 : 41 = 0,048 \cdot 100 \rightarrow 4,8\%$

$12 : 41 = 0,292 \cdot 100 \rightarrow 29,2\%$

$12 : 41 = 0,292 \cdot 100 \rightarrow 29,2\%$

$14 : 41 = 0,341 \cdot 100 \rightarrow 34,1\%$

$1 : 41 = 0,024 \cdot 100 \rightarrow 2,4\%$

Distribuição de frequências relativas	
Idade	Frequência (F)
8	4,8%
9	29,2%
10	29,2%
11	34,1%

12	2,4%
Total (F_T)	100%

Leia também: [Aplicação de estatística: frequência absoluta e frequência relativa](#)

• Classes

Em casos em que a variável é contínua, isto é, quando ela possui diversos valores, é necessário agrupá-los em [intervalos reais](#). Na estatística esses intervalos são chamados de **classes**.

Para construir a tabela de **distribuição de frequências em classes**, devemos colocar os intervalos na coluna da esquerda, com seu devido título, e na coluna da direita, devemos colocar a frequência absoluta de cada um dos intervalos, ou seja, quantos elementos pertencem a cada um deles.

Exemplo

Altura dos alunos da classe do 3º ano do Ensino Médio de uma escola.

Distribuição de frequência em classes	
Altura (metros)	Frequência absoluta (F)
[1,40; 1,50[1
[1,50; 1,60[4
[1,60; 1,70[8
[1,70; 1,80[2
[1,80; 1,90[1
Total (F_T)	16

Analisando a tabela de distribuição de frequência em classes, podemos ver que, na turma do terceiro ano, temos 1 estudante que possui altura entre 1,40 m e 1,50 m, assim como temos 4 estudantes com altura entre 1,50 e 1,60 m, e assim sucessivamente. Podemos observar também que os estudantes possuem altura entre 1,40 m e 1,90 m, a diferença entre essas medidas, ou seja, entre a maior altura e a menor altura da amostra, é chamada de **amplitude**.

A diferença entre o limite superior e o limite inferior de uma classe é chamada de **amplitude da classe**, assim, a segunda, que possui 4 alunos com alturas entre 1,50 metro (inclusos) e 1,60 metro (não inclusos), possui amplitude de:

$$1,60 - 1,50$$

0,10 metro

Veja também: [Medidas de dispersão: amplitude e desvio](#)

Medidas de posição

As medidas de posição são utilizadas em casos em que é possível construir-se um rol numérico com os dados ou uma tabela de frequência. Essas medidas indicam a posição dos elementos em relação ao rol. As três principais medidas de posição são:

- **Média**

Considere o rol com os elementos $(a_1, a_2, a_3, a_4, \dots, a_n)$, a média aritmética desses n elementos é dada por:

$$\bar{X} = \frac{a_1 + a_2 + a_3 + a_4 + \dots + a_n}{n}$$

Exemplo

Em um grupo de dança, as idades dos integrantes foram coletadas e representadas no rol a seguir:

(18, 20, 20, 21, 21, 21, 22, 22, 25, 30)

Vamos determinar a idade média dos integrantes desse grupo de dança.

De acordo com a fórmula, devemos somar todos os elementos e dividir esse resultado pela quantidade de elementos do rol, assim:

$$\bar{X} = \frac{18 + 20 + 20 + 21 + 21 + 21 + 22 + 22 + 25 + 30}{10}$$

$$\bar{X} = \frac{220}{10}$$

$$\bar{X} = 22$$

Portanto, a idade média dos integrantes é de 22 anos.

Para saber mais sobre essa medida de posição, leia nosso texto: [Média](#).

- **Mediana**

A mediana é dada pelo elemento central de um rol que possui uma quantidade ímpar de elementos. Caso o rol possua uma quantidade par de elementos, devemos considerar os dois elementos centrais e calcular a média aritmética entre eles.

Exemplo

Considere o rol a seguir.

(2, 2, 3, 3, 4, 5, 6, 7, 9)

Veja que o elemento 4 divide o rol em duas partes iguais, logo, ele é o elemento central.

Exemplo

Calcule a mediana das idades do grupo de dança.

Lembre-se de que o rol das idades desse grupo de dança é dado por:

(18, 20, 20, 21, 21, 21, 22, 22, 25, 30)

Veja que o número de elementos desse rol é igual a 10, logo, não é possível dividir o rol em duas partes iguais. Assim devemos tomar dois elementos centrais e realizar a média aritmética desses valores.

$$\frac{21 + 21}{2}$$

$$\frac{42}{2}$$

$$21$$

Veja mais detalhes dessa medida de posição em nosso texto: [Mediana](#).

- **Moda**

Chamaremos de moda o elemento do rol que possui maior frequência, ou seja, o elemento que mais aparece nele.

Exemplo

Vamos determinar a moda do rol das idades do grupo de dança.

(18, 20, 20, 21, 21, 21, 22, 22, 25, 30)

O elemento que mais aparece é o 21, portanto, a moda é igual a 21.

Medidas de dispersão

As medidas de dispersão são **utilizadas nos casos em que a média já não é suficiente**. Por exemplo, imagine que dois carros tenham percorrido uma média de 40.000 quilômetros. Somente com conhecimento sobre média podemos afirmar que os dois carros andaram determináveis quilômetros cada um, certo?

No entanto, imagine que um dos carros tenha percorrido 79.000 quilômetros, e o outro, 1.000 quilômetros, veja que somente com as informações sobre média não é possível realizar afirmações com precisão.

As **medidas de dispersão** nos indicarão o quanto os elementos de um rol numérico estão afastados da média aritmética. Temos duas importantes medidas de dispersão:

- **Variância (σ^2)**

Vamos chamar de variância a média aritmética dos quadrados da diferença entre cada elemento do rol e a média aritmética desse rol. A variância é representada por: σ^2 .

Considere o rol $(x_1, x_2, x_3, \dots, x_n)$ e que ele possua média aritmética \bar{x} . A variância é dada por:

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

- **Desvio-padrão (σ)**

O desvio-padrão é dado pela raiz da variância, ele nos indica o quanto um elemento está disperso em relação à média. O desvio padrão é denotado por σ .

Exemplo

Determine o desvio-padrão do conjunto de dados (4, 7, 10). Veja que, para isso, é necessário determinar-se primeiro a variância, e que, para tanto, é necessário antes o cálculo da média desses dados.

$$\bar{x} = \frac{4 + 7 + 10}{3}$$

$$\bar{x} = \frac{21}{3}$$

$$\bar{x} = 7$$

Substituindo esses dados na fórmula da variância, temos:

$$\sigma^2 = \frac{(4 - 7)^2 + (7 - 7)^2 + (10 - 7)^2}{3}$$

$$\sigma^2 = \frac{(-3)^2 + 0^2 + 3^2}{3}$$

$$\sigma^2 = \frac{9 + 0 + 9}{3}$$

$$\sigma^2 = \frac{18}{3}$$

$$\sigma^2 = 6$$

Para determinar o desvio-padrão, devemos extrair a raiz da variância.

$$\sigma^2 = 6$$

$$\sigma = \sqrt{6}$$

Leia mais: [Medidas de dispersão: variância e desvio-padrão](#)

Para que serve a estatística?

Vimos que a estatística está relacionada a **problemas de contagem ou organização de dados**. Além disso, ela tem um importante papel no desenvolvimento de ferramentas que possibilitam o processo de organização de dados, com em tabelas. A estatística está presente também em **diversos campos da ciência**, com base na coleta de dados e em seu tratamento, é possível trabalhar modelos matemáticos que permitem maior desenvolvimento na área estudada. Alguns campos em que a estatística é fundamental: economia, meteorologia, marketing, esportes, sociologia e geociências.

Na meteorologia, por exemplo, os dados são coletados em determinado período, depois de organizados, eles são tratados, e assim, com base neles, constrói-se um modelo matemático que nos permite afirmar sobre o clima de dias anteriores com maior grau de confiabilidade. A estatística é um ramo da ciência que nos permite fazer afirmações com certo grau de confiabilidade, mas nunca com 100% de certeza.

Divisões da estatística

A estatística é dividida em duas partes, descritiva e inferencial. A primeira está relacionada à contagem dos elementos envolvidos na pesquisa, esses elementos são contados um a um. Na **estatística descritiva**, temos como principais ferramentas as medidas de posição, como média, mediana e moda, assim como as medidas de dispersão, como variância e desvio-padrão, temos também tabelas de frequências e gráficos.

Ainda na estatística descritiva, temos uma metodologia muito bem definida para uma **apresentação de dados com grau considerável de confiabilidade** que passa por organização e coleta, resumo, interpretação e representação e, por fim,

análise de dados. Um exemplo clássico da utilização da estatística descritiva ocorre na realização do censo populacional (de 10 em 10 anos) pelo Instituto Brasileiro de Geografia e Estatística ([IBGE](#)).

A **estatística inferencial**, por sua vez, é caracterizada não por coletar dados dos elementos de uma população um a um, e sim por realizar a **análise de uma amostra dessa população, tirando conclusões** sobre ela. Na estatística inferencial, deve-se tomar cuidado na escolha da amostra, pois ela deve representar muito bem a população. Alguns resultados iniciais, como cálculo de média, na estatística inferencial chamado de esperança, são deduzidos com base nos conhecimentos da estatística descritiva.

A estatística inferencial é utilizada, por exemplo, nas pesquisas eleitorais. Escolhe-se uma amostra da população, de forma que a represente, e assim é realizada a pesquisa. Na escolha de uma amostra que não represente muito bem essa população, dizemos que a pesquisa é **tendenciosa** e, por consequência, não confiável.



Área da matemática responsável pela contagem e pela organização de dados.

Exercícios resolvidos

Questão 1 – (U. F. Juiz de Fora – MG) Um professor de física aplicou uma prova, valendo 100 pontos, em seus 22 alunos e obteve, como resultado, a distribuição das notas, vista no quadro seguinte:

40	20	10	20	70	60
90	80	30	50	50	70
50	20	50	50	10	40
30	20	60	60	–	–

Faça os seguintes tratamentos de dados:

- Escreva o rol dessas notas.
- Determine a frequência relativa da maior nota.

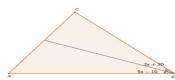
Resolução

- Para fazer o rol dessas notas, devemos escrevê-las de maneira crescente ou decrescente. Assim temos que:

10, 10, 20, 20, 20, 20, 30, 30, 40, 40, 50, 50, 50, 50, 50, 60, 60, 60, 80, 90

b) Observando o rol, podemos ver que a maior nota foi igual a 90 e que sua frequência absoluta é igual a 1, pois ela aparece apenas uma vez. Para determinar a frequência relativa, devemos dividir a frequência absoluta dessa nota pela frequência total, nesse caso, igual a 22. Assim:

Frequência relativa



Para passar esse número para porcentagem, devemos multiplicá-lo por 100.

$$0,045 \cdot 100$$

4,5%

Questão 2 – (Enem) Depois de jogar um dado em forma de cubo e de faces numeradas de 1 a 6, por 10 vezes consecutivas, e anotar o número obtido em cada jogada, constituiu-se a seguinte tabela de distribuição de frequências.

Número obtido	Frequência
1	4
2	1
4	2
5	2
6	1

A média, a mediana e a moda dessa distribuição de frequências são, respectivamente:

- a) 3, 2 e 1
- b) 3, 3 e 1
- c) 3, 4 e 2
- d) 5, 4 e 2
- e) 6, 2 e 4

Resolução

Alternativa B.

Para determinarmos a média, observe que existe repetição dos números obtidos, assim, utilizaremos a média aritmética ponderada.

$$\bar{x} = \frac{1 \cdot 4 + 2 \cdot 1 + 4 \cdot 2 + 5 \cdot 2 + 6 \cdot 1}{4 + 1 + 2 + 2 + 1}$$

$$\bar{x} = \frac{4 + 2 + 8 + 10 + 6}{10}$$

$$\bar{x} = \frac{30}{10}$$

$$\bar{x} = 3$$

Para determinar a mediana, devemos organizar o rol de maneira crescente ou decrescente. Lembre-se de que a frequência é a quantidade de vezes que a face aparece.

1, 1, 1, 1, **2**, **4**, 4, 5, 5, 6

Como o número de elementos do rol é par, devemos calcular a média aritmética dos elementos centrais que dividem o rol ao meio para determinar a mediana, assim:

$$\frac{2 + 4}{2}$$

$$\frac{6}{2}$$

$$3$$

A moda é dada pelo elemento que mais aparece, ou seja, que possui maior frequência, portanto, temos que a moda é igual 1.

Assim a média, a mediana e a moda são, respectivamente, iguais a:

3, 3 e 1

Por Robson Luiz
Professor de Matemática

Fonte: Brasil Escola - <https://brasilecola.uol.com.br/matematica/estatistica-2.htm>