Saudações Pythônicas!

Pat Simões

contato@sp.scipy.org.br

1. MÉDIAS, VARIÂNCIAS E OUTROS BICHOS

Dados estatísticos

O que são? De onde vêm?

Em estatísica descritiva
Dados são informações que resultam da:

- Observação ou percepção de algo
- Medida de algo

Exemplos:

*Medida do tempo de queda de um objeto, ou massa do mesmo;

*Intensidade do cheiro de um perfume;

*Número de visitantes em um website.

Dados podem ser...

Numéricos (quantitativos)

Discretos: número de filhos por família; número de casos de gripe em um ano

Contínuos: temperatura de uma sala; altura das pessoas em uma sala.

Categóricos (qualitativos)

Nominais: cor de cabelo (louro, preto, ruivo...); estado civil (casado, solteiro, . . .)

Ordinais:

Escolaridade (ensino fundamental, médio, etc.); faixas de karatê (branca, amarelo,...).

Mas espera aí!

Idade é um dado discreto ou contínuo? E a faixa etária como fica nessa história?

Em análise de dados,
Tudo depende do que se quer

Idade pode ser tratada como dado discreto ou contínuo ou até ser categorizada. Outros exemplos que podem ser categorizados: PH e pressão arterial.

Podemos sumarizar informações sobre os dados quantitativos através de valores médios e medidas de dispersão.

Exemplo: tempo de queda de um objeto

# medida	tempo (seg)
1	3.0
2	3.4
3	3.2
4	3.4

Se cada lançamento dá um valor diferente, qual afinal é o tempo de queda?

Podemos representar o tempo de queda por um valor que considera todos os dados obtidos pela medida. **Este valor é a média (ou valor esperado):**

$$média = \frac{(3.1 + 3.4 + 3.2 + 3.4)}{4} = 3.3\,s$$

Ou para 4 valores de uma observação qualquer x, podemos reescrever como:

$$\bar{x} = \frac{(x_1 + x_2 + x_3 + x_4)}{4} \qquad \bar{x} = \langle x \rangle = \mu = média$$

E se eu tiver N observações, como fica?

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_N}{N}$$

Ou seja, a média é a soma de todos os valores dividido pelo número total de valores.

Ou de maneira condensada:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\bar{x} = \langle x \rangle = \mu = média$$

$$x_i = valor\ da\ medida\ i$$

E quando tenho algum dado muito discrepante, será que a média representa bem meu conjunto de valores?

Notas	Nota prova
Joca	8.5
Tiquinho	8.3
Jajá	0
Binho	8.5
Tião	8.0

Embora a maioria dos alunos tenha nota na faixa de 8 a 8.5, a média deu 6.6

O que fazer?
Achar a mediana!

Para achar a mediana:

a. Ordenar os valores;
b. Se o número de valores for ímpar, a mediana é o número do meio; se for par, é a média entre os dois valores mais centrais:

Nomes	Nota prova
Joca	8,5
Tiquinho	8,3
Jajá	0
Binho	8.5
Tião	8.0

0	8.0	8.3	8.5	8.5

A mediana é:
8.3

Em meio a tantos dados, com valores diferentes, existe uma maneira de quantificar a dispersão dos valores da observação, calculando a VARIÂNCIA.

A variância é uma medida da dispersão dos dados em relação à média, ou seja, uma medida do quanto os dados estão próximos ou não da média.

$$s = \frac{1}{N}\left[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_N - \bar{x})^2\right]$$

$$s = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

Além da variância, outra grandeza usada que também reflete a dispersão dos valores é o DESVIO PADRÃO.

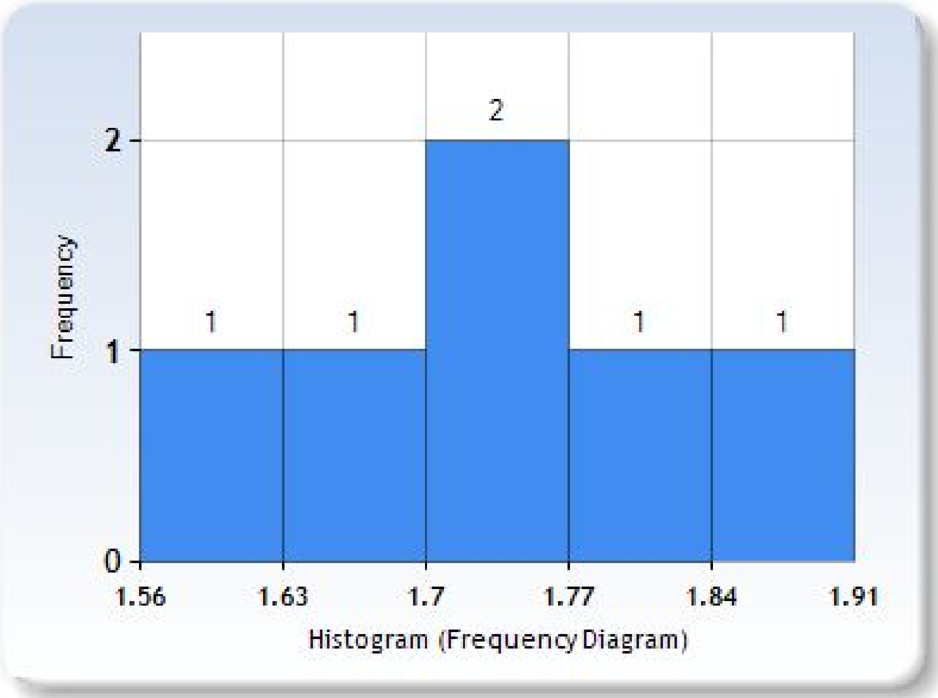
O desvio padrão nada mais é que a raiz quadrada da variância e também uma medida de como os valores se afastam da média.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

Distribuição gráfica dos dados

Nomes	**Altura**
Joca	1.80
Tiquinho	1.56
Jajá	1.70
Binho	1.73
Tuco	1.65
Tião	1.89

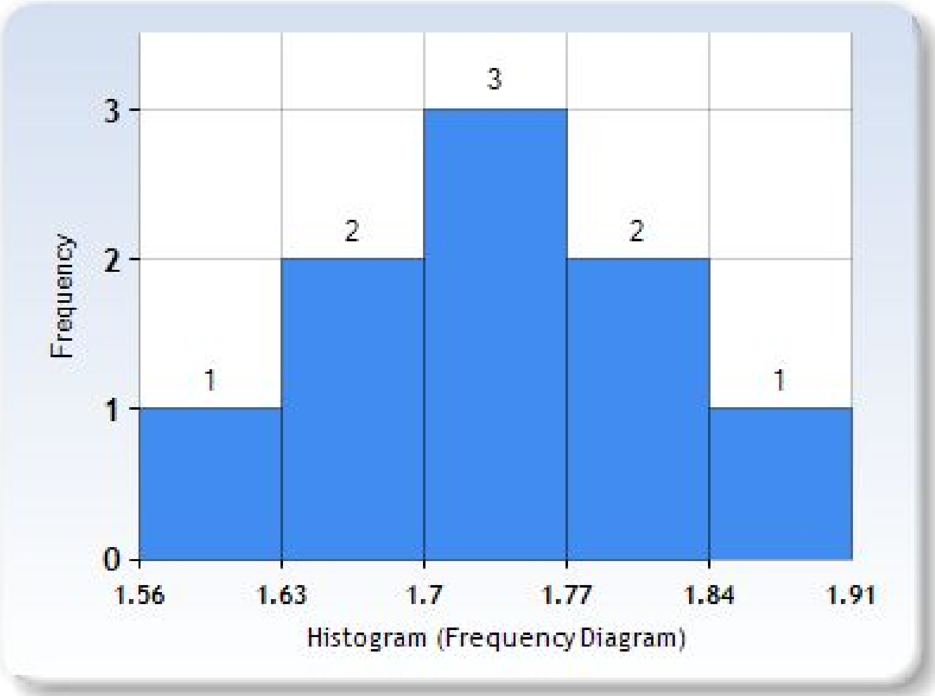
Histograma



Distribuição gráfica dos dados

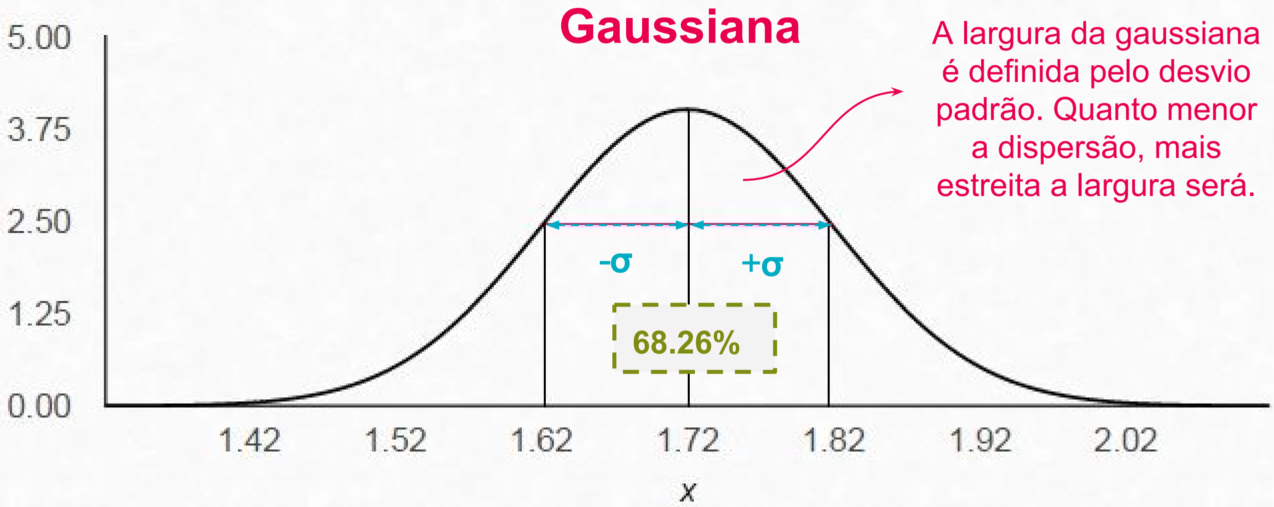
Nomes	Altura
Joca	1.80
Tiquinho	1.56
Jajá	1.70
Binho	1.73
Tuco	1.65
Tião	1.89
Azeitona	1.64
Jorjão	1.80
Pipoca	1.73

Histograma



Distribuição gráfica dos dados

Se tivéssemos infinitos valores, a cara da distribuição de dados seria a de uma distribuição gaussiana (normal)



$$\mu = E(X) = 1.72 \qquad \sigma = SD(X) = 0.099 \qquad \sigma^2 = Var(X) = 0.01$$

Thanks!

Sugestões de referências:

*Fundamentos da teoria dos erros - J.E. Vuolo. Ed. Blucher.

*Estatística Básica - W.O. Bussab, P.A. Morettin. Ed. Saraiva.

*Guia Mangá de Estatística - Shin Takahashi. Ed. Novatec.

Credits

- Presentation template by SlidesCarnival