

Estatística básica

@CursoDS_ProfDanilo

Prof. Dr. Danilo Morales Teixeira

Sites e contatos

- Slack: pythonprofdanilo.slack.com
- GitHub: https://github.com/danmorales/CursoDS_ProfDanilo
 - E-mail: cursods.profdanilo@gmail.com
- Facebook: <https://www.facebook.com/profdanilods/>

Conteúdo

- Média
- Mediana
- Desvio padrão
- Erro padrão da média

Média e mediana

Média e mediana são medidas de tendência central muito utilizadas em estatística

Média - definição

- Média é definida como sendo a soma de todos os elementos dividida pelo número de elementos da amostra
- Mais indicado quando os dados estão igualmente espaçados

$$Média = \frac{X_1 + X_2 + \cdots + X_N}{N} = \frac{\sum_{i=1}^N X_i}{N}$$

Média - exemplo

- Sejam as notas das provas:
- notas=[5.7,3.4,6.8,7.3,3.2,7.9]
 - Qual a média das notas?
- $M = \frac{5.7+3.4+6.8+7.3+3.2+7.9}{6} = \frac{34.3}{6} = 5.7$

Mediana - definição

- Representa o valor central do conjunto de dados
- Valores devem estar ordenados em ordem crescente ou decrescente
- Quando o número de elementos é par, a mediana é calculada como sendo a média entre os dois valores centrais

Mediana – exemplo 1

- Sejam diferentes pessoas com idades:
- Idades=[34,23,45,78,43,12,39,41,56]
 - Qual a mediana?
 - Ordenando os valores:
- Idades = [12,23,34,39,41,43,45,56,78]
 - Mediana = 41

Mediana – exemplo 2

- Sejam diferentes pessoas com idades:
- Idades=[34,23,45,78,43,12,39,41,56,37]
 - Qual a mediana?
 - Ordenando os valores:
- Idades = [12,23,34,37,39,41,43,45,56,78]
 - Mediana = $(39+41)/2 = 80/2 = 40$

Desvio padrão - definição

- Medida de dispersão em torno da média
- Desvio padrão baixo indica que os valores estão próximos da média
 - Utilizado para expressar o coeficiente de correlação
 - Usado para medir a confiança em cálculos estatísticos
 - Definido como:

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

Desvio padrão - definição

Desigualdade de Chebyshev nos diz que a população mínima é dada por:

Distância da média	População mínima
$\sqrt{2}\sigma$	50%
2σ	75%
3σ	89%
4σ	94%
5σ	96%
6σ	97%
$k\sigma$	$\left(1 - \frac{1}{k^2}\right) \times 100\%$

Desvio padrão - definição

Para o caso em que a distribuição é praticamente normal, temos:

Distância da média	População mínima
0.68449σ	50%
0.994458σ	68%
1σ	68.27%
2σ	95.45%
3σ	99.73%
4σ	99.993%
7σ	99.9999999997%

Desvio padrão - exemplo

- Sejam as notas das provas:
- notas=[5.7,3.4,6.8,7.3,3.2,7.9]
- Qual o desvio padrão das notas?
 - Média = 5.7

Desvio padrão - exemplo

Nota	Nota-Média	(Nota-Média) ²
5.7	5.7 - 5.7 = 0.0	0.0
3.4	3.4 - 5.7 = -2.3	5.29
6.8	6.8 - 5.7 = 1.1	1.21
7.3	7.3 - 5.7 = 1.6	1.96
3.2	3.2 - 5.7 = -2.5	6.25
7.9	7.9 - 5.7 = 2.2	4.84

$$\sigma = \sqrt{\frac{1}{6} \sum (Nota - Media)^2} = \sqrt{\frac{19.55}{6}} = \sqrt{3.26} = 1.80$$

Erro padrão da média - definição

- Imagine que temos uma população com N pessoas, com média de altura H e desvio padrão σ
- Imagine uma pequena amostra desta população com n pessoas e com altura média h
 - Qual o erro padrão desta pequena amostra?
- O erro padrão da média nos diz o quanto esta pequena população está afastada da média da população toda

Erro padrão da média - definição

O erro padrão da média é definido como:

$$SE = \frac{\sigma}{\sqrt{n}}$$

onde σ é o desvio padrão da população

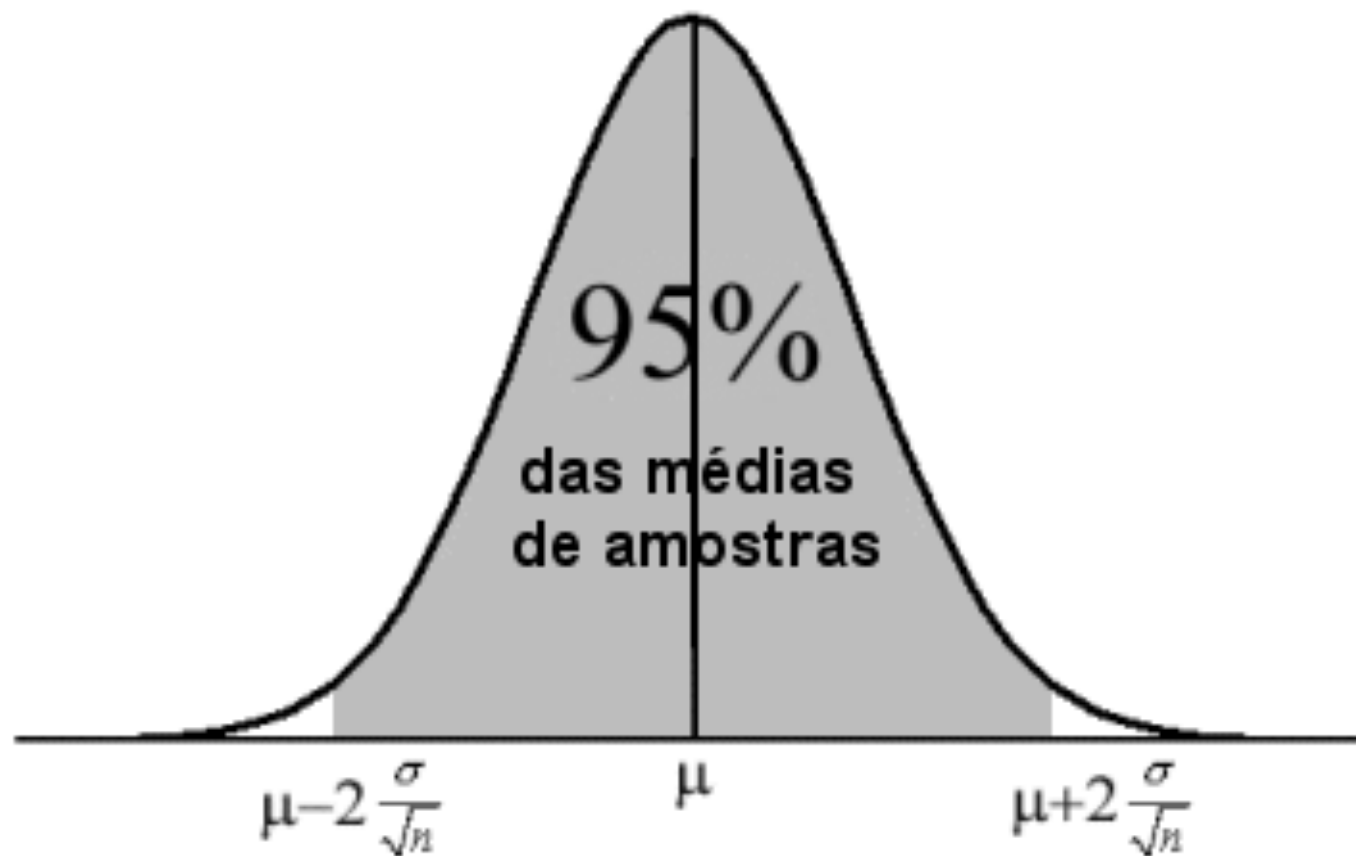
Erro padrão da média - exemplo

Seja uma população com média de altura $H=1.73m$ e desvio padrão $\sigma=0.35m$. Qual o erro padrão da média para uma população com 15 pessoas?

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{0.35}{\sqrt{15}} = 0.09m$$

Isto nos diz que 68% desta pequena população teria uma altura entre 1.64m e 1.82m

Erro padrão da média - exemplo



Aplicações

Vejam os vídeos da descrição com aplicações implementando as equações do zero e utilizando a biblioteca NumPy do Python

Obrigado

- Gostou? Deixe o seu like
 - Curta o canal para receber notificações de novos vídeos
- Dúvidas? Mande um e-mail ou deixe o seu comentário aqui no vídeo