



Escola Politécnica de Pernambuco
Especialização em Ciência de Dados e Analytics

Estatística Computacional

Aula 1.1 – Estatística Descritiva – PARTE II

Prof. Dr. Rodrigo Lins Rodrigues

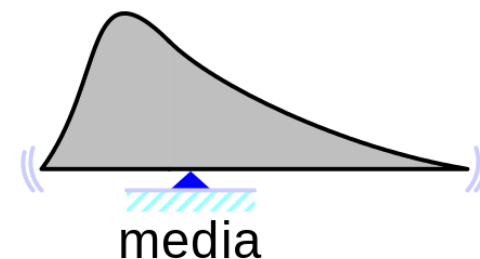
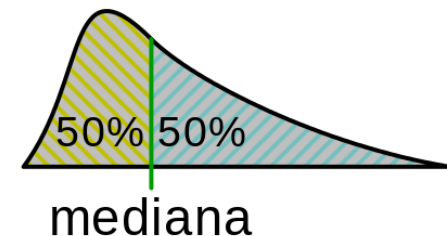
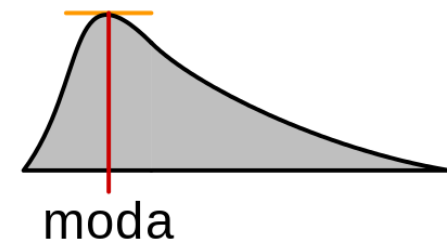
rodrigolins.rodrigues@ufrpe.br

Medidas de Tendência Central



Medidas de Tendência Central

- Localizam-se geralmente no **centro de uma distribuição**;
- Indica a posição dos dados em relação ao eixo dos valores assumidos pela variável;
- As principais são:
 - ✓ Média;
 - ✓ Moda;
 - ✓ Mediana;
- A média pode ser: **aritmética**, geométrica, quadrática, cúbica, etc.



Medidas de Tendência Central

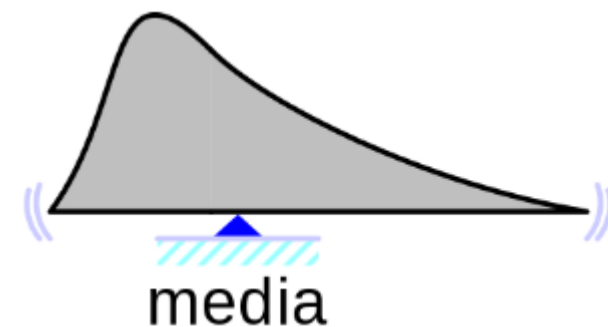
- **Média aritmética:**

- É a mais utilizada entre as medidas de tendência central;
- É a soma do total de valores dividida pelo número total de observações:

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

- **Observação:**

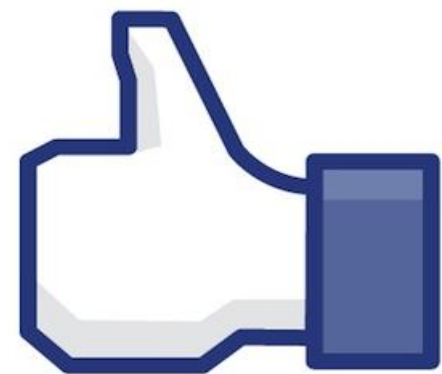
- Quando se refere a **população** a média é representada por μ ;
- Quando se refere a **amostra** a média é representada por \bar{X}



Medidas de Tendência Central

▪ Vantagens da média aritmética:

- ✓ É a medida de tendência central **mais conhecida e de maior uso**;
- ✓ Seus **cálculos são fáceis**;
- ✓ Serve para fazer comparações entre dois ou mais fenômenos;
- ✓ É sempre **possível encontrar a média** aritmética em dados numéricos;
- ✓ Seu **valor é único**;



Medidas de Tendência Central

▪ Desvantagens da média aritmética:

- ✓ É **fortemente influenciada por valores extremos**, nem sempre representando a verdade;
- ✓ Não pode ser aplicada para variáveis qualitativas;



Medidas de Tendência Central

▪ Exemplo 1:

- ✓ Calcular a média aritmética para os dados referente às notas dos alunos da especialização em Ciência dos Dados.

Notas									
5,7	6,5	6,9	8,3	8,0	4,2	6,3	7,4	5,8	6,9

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\bar{X} = \frac{5,7+6,5+...+6,9}{10} = 6,6$$

Medidas de Tendência Central

▪ Exemplo 2:

✓ Calcular a média salarial dos alunos desta turma.

Salário			
R\$ 2.000,00	R\$ 1.800,00	R\$ 2.100,00	R\$ 18.500,00

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\bar{X} = \frac{2000+1800+2100+18.500}{4} = 6.100,00$$

Medidas de Tendência Central

Qual o problema com o exemplo 2 ?



Medidas de Tendência Central

- **Média aritmética ponderada:**

- ✓ Na média ponderada as observações tem pesos diferenciados (p_i);

$$\bar{X} = \frac{\sum_{i=1}^n x_i p_i}{\sum_{i=1}^n p_i}$$

- ✓ É muito utilizada em provas de concurso público, onde as matérias tem pesos diferentes.

Medidas de Tendência Central

- **Exemplo 1:**

✓ Na tabela abaixo é mostrado as notas de um aluno e seus respectivos pesos por bimestre.

Período	Nota	Peso
1º Bimestre	4,5	1
2º Bimestre	7,0	2
3º Bimestre	5,5	3
4º Bimestre	6,5	4

$$\bar{X} = \frac{4,5*1+7,0*2+5,5*3+6,5*4}{1+2+3+4} = 6,1$$

Medidas de Tendência Central

- **Média aritmética para dados discretos agrupados:**

- ✓ Quando os valores discretos de x_i se repetem os dados são agrupados em uma tabela de frequência;
- ✓ Utiliza-se o mesmo critério da média ponderada, onde os pesos passam a ser representados por frequências absolutas (F_i);

$$\bar{X} = \frac{\sum_{i=1}^n x_i F_i}{\sum_{i=1}^n F_i}$$

Medidas de Tendência Central

- **Exemplo:**

- ✓ A tabela representa uma pesquisa de satisfação com 66 entrevistados, onde avaliou o desempenho de uma seguradora de saúde, por meio das notas atribuídas variando de 5 a 10.

Notas	Número de entrevistados
5	24
6	26
7	5
8	7
9	3
10	1

Medidas de Tendência Central

- **Exemplo:**

✓ A média aritmética é calculada a partir da equação:

$$\bar{X} = \frac{5 * 24 + 6 * 26 + 7 * 5 + 8 * 7 + 9 * 3 + 10 * 1}{66}$$

$$\bar{X} = 6,12$$

Medidas de Tendência Central

- **Média aritmética para dados contínuos agrupados:**
 - ✓ Cada classe não tem um único valor definido e sim um conjunto de valores pertencentes a classe;
 - ✓ Para que a média possa ser calculada assume-se que x_i é o ponto central da classe:

$$\bar{X} = \frac{\sum_{i=1}^k x_i F_i}{\sum_{i=1}^k F_i}$$

Medidas de Tendência Central

- **Exemplo:**

✓ A tabela apresenta as classes de salários pagos aos funcionários de uma empresa e suas respectivas frequências:

Classe	F_i	$F_{ri}(\%)$
1 -- 3	240	17,14
3 -- 5	480	34,29
5 -- 7	320	22,86
7 -- 9	150	10,71
9 -- 11	130	9,29
11 -- 13	80	5,71
Soma	1.400	100

Medidas de Tendência Central

- **Exemplo:**

✓ Considerando x_i o ponto médio da classe e aplicando a equação:

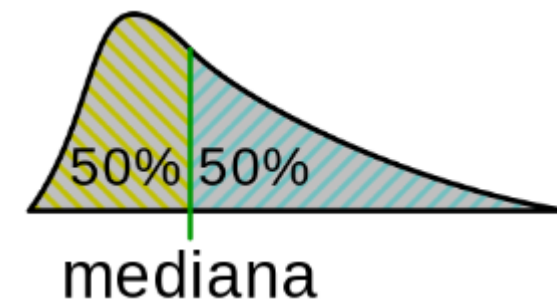
$$\bar{X} = \frac{2 * 240 + 4 * 480 + 6 * 320 + 8 * 150 + 10 * 130 + 12 * 80}{1.400}$$

$$\bar{X} = 5,557$$

Medidas de Tendência Central

- **Mediana:**

- ✓ É o valor médio para um conjunto de dados;
- ✓ **Não depende de todos os valores** da série, podendo, como a moda, não se alterar com a mudança de alguns elementos;
- ✓ É muito empregada em pesquisas onde não interessam valores extremos, pois **não é influenciada por esses valores**;



Medidas de Tendência Central

- **Mediana para dados não agrupados:**

- ✓ As observações devem ser colocadas em ordem crescente;
- ✓ Pode ser calculada da seguinte forma:

$$Md(x) = \begin{cases} \frac{x_{\frac{n}{2}} + x_{(\frac{n}{2})+1}}{2} & , \text{ se } n \text{ for par} \\ x_{\frac{(n+1)}{2}} & , \text{ se } n \text{ for ímpar} \end{cases}$$

em que n é o número total de observações.

Medidas de Tendência Central

- **Exemplo:**

- ✓ A tabela apresenta a produção mensal de livros relacionados com Data Science no mundo.
- ✓ Para o cálculo da mediana as observações devem ser ordenadas:

Mês	Produção (unidades)
Jan	210
Fev	180
Mar	203
Abr	195
Mai	208
Jun	230
Jul	185
Ago	190
Set	200
Out	182
Nov	205
Dez	196

Medidas de Tendência Central

- Solução:

180 < 182 < 185 < 190 < 195 < 196 < 200 < 203 < 205 < 208 < 210 < 230
1º 2º 3º 4º 5º 6º 7º 8º 9º 10º 11º 12º

$$Md(x) = \frac{x_{\frac{n}{2}} + x_{(\frac{n}{2})+1}}{2} \Rightarrow Md(x) = \frac{x_{\frac{12}{2}} + x_{(\frac{12}{2})+1}}{2}$$

$$Md = \frac{196+200}{2} = 198$$

Mês	Produção (unidades)
Jan	210
Fev	180
Mar	203
Abr	195
Mai	208
Jun	230
Jul	185
Ago	190
Set	200
Out	182
Nov	205
Dez	196

Medidas de Tendência Central

- **Mediana para dados agrupados em classes:**

- ✓ **Passo 1:** Calcular a posição da mediana por meio da seguinte equação.

$$Pos(Md) = n/2$$

- ✓ **Passo 2:** Identificar a classe mediana a partir da coluna de frequência acumulada;

Medidas de Tendência Central

- **Mediana para dados agrupados em classes:**

- ✓ **Passo 3:** Calcular a mediana pela seguinte equação:

$$Md = LI_{Md} + \frac{\left(\frac{n}{2} - F_{ac(Md-1)}\right)}{F_{Md}} \times A_{Md}$$

Em que:

LI_{Md} = limite inferior da classe mediana

F_{Md} = frequência absoluta da classe mediana

$F_{ac(Md-1)}$ = frequência acumulada da classe anterior à classe mediana

A_{Md} = amplitude da classe mediana

n = número total de observações

Medidas de Tendência Central

- **Exemplo:**

Considere os dados da tabela referente a salários pagos a funcionários.

Classe	F_i	$F_{ri}(\%)$
1 -- 3	240	17,14
3 -- 5	480	34,29
5 -- 7	320	22,86
7 -- 9	150	10,71
9 -- 11	130	9,29
11 -- 13	80	5,71
Soma	1.400	100

Medidas de Tendência Central

- **Solução:**

- ✓ **Passo 1:** Calcular a posição da mediana.

$$Pos(Md) = \frac{n}{2} = \frac{1.400}{2} = 700$$

- ✓ **Passo 2:** Identificação da classe mediana;

3 -- 5	480	34,29
---------	-----	-------

Medidas de Tendência Central

- **Solução:**

✓ **Passo 3:** Calculando a mediana.

$$Md = LI_{Md} + \frac{\left(\frac{n}{2} - F_{ac(Md-1)}\right)}{F_{Md}} \times A_{Md}$$

$$\begin{aligned} LI_{Md} &= 3 \\ F_{Md} &= 480 \\ F_{ac(Md-1)} &= 240 \\ A_{Md} &= 2 \\ n &= 1.400 \end{aligned}$$

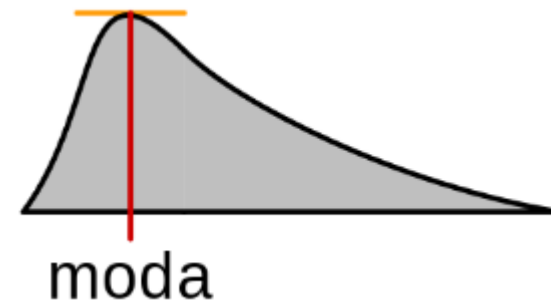
Logo

$$Md = 3 + \frac{(700 - 240)}{480} \times 2 = \boxed{4.916}$$

Medidas de Tendência Central

- **Moda:**

- ✓ A moda (M_0) é a observação que ocorre com maior frequência no conjunto de dados;
- ✓ É a única medida de posição que também pode ser **utilizada para variáveis qualitativas**;
- ✓ Em uma única série pode-se ter **mais de uma moda**.



Medidas de Tendência Central

- **Vantagens da Moda:**

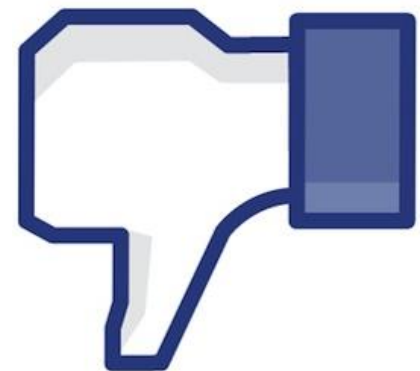
- ✓ A moda **não depende de todos os valores da série**, não se alterando com a modificação de alguns deles;
- ✓ A moda **não é influenciada** por valores extremos;
- ✓ A moda sempre tem existência real;



Medidas de Tendência Central

- **Desvantagem da Moda:**

- ✓ A moda não é aplicável a um número pequeno de dados observados, ao contrário da média, que pode ser aplicada a uma amostra de tamanho pequena;
- ✓ A moda nem sempre pode ser calculada, como no caso de se ter uma série amodal;



Medidas de Tendência Central

- **Moda para dados qualitativos:**

- ✓ Uma emissora de TV entrevistou 500 telespectadores buscando analisar suas preferências por categoria de interesse.

Categorias de interesse	Fi
Filmes	71
Novelas	46
Jornalismo	90
Humor	98
Esporte	120
Shows	35
Variedades	40

Medidas de Tendência Central

- **Moda para dados agrupados em classes:**

- ✓ **Passo 1:** Identificar a classe modal, que é a com maior frequência absoluta;
- ✓ **Passo 2:** Calcular a moda (M_o):

$$M_o = LI_{M_o} + \frac{F_{M_o} - F_{M_o-1}}{2F_{M_o} - (F_{M_o-1} + F_{M_o+1})} \times A_{M_o}$$

- ✓ LI_{M_o} = limite inferior da classe modal
- ✓ F_{M_o} = frequência absoluta da classe modal
- ✓ F_{M_o-1} = Frequência absoluta da classe anterior à classe modal;
- ✓ F_{M_o+1} = Frequência absoluta da classe posterior à classe modal;
- ✓ A_{M_o} = Frequência absoluta da classe anterior à classe modal;

Medidas de Tendência Central

- **Exemplo:**

- ✓ Conjunto de dados contínuos com 200 observações agrupadas em classes.

Classe	Fi
01 -- 10	21
10 -- 20	36
20 -- 30	58
30 -- 40	24
40 --50	19
Soma	200

Medidas de Tendência Central

- **Solução:**

- ✓ **Passo 1:** Verifica-se que a classe modal é a terceira (20|--30);

- ✓ **Passo 2:** Calculando a moda:

$$M_o = LI_{M_o} + \frac{F_{M_o} - F_{M_o-1}}{2F_{M_o} - (F_{M_o-1} + F_{M_o+1})} \times A_{M_o}$$

- ✓ $LI_{M_o} = 20, F_{M_o} = 58, F_{M_o-1} = 36, F_{M_o+1} = 24, A_{M_o} = 10$

$$M_o = 20 + \frac{58 - 36}{2 \times 58 - (36 + 24)} \times 10 = \boxed{23,9}$$

Agora é com vocês!

- Quais as diferenças entre as medidas: média, mediana e moda ?
- Quais as limitações em utilizar apenas medidas de tendência central no estudo de uma determinada variável ?
- Na tabela de dados que construímos, calcule a média, moda e mediana para a variável “Idade”.



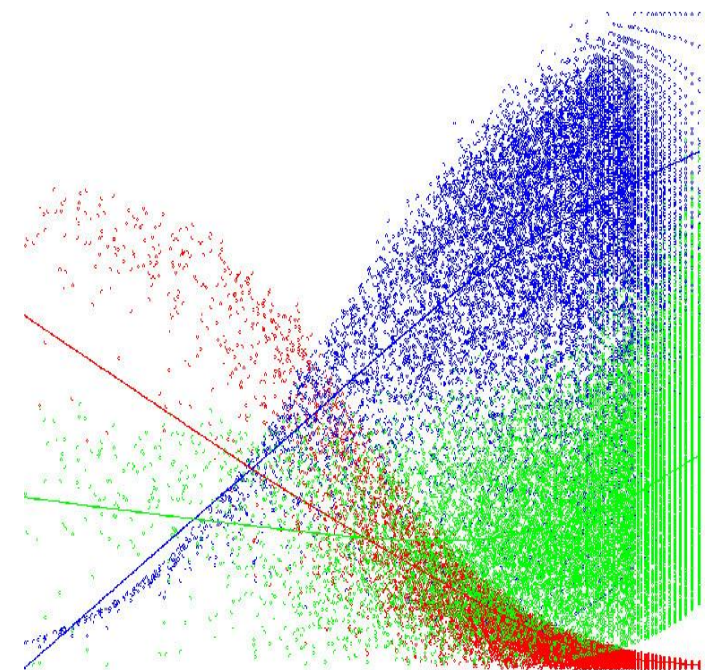


The background image shows a man with glasses and a dark shirt sitting at a desk with a laptop. Behind him is a wall covered in extensive handwritten notes, diagrams, and graphs. The notes include financial calculations, percentages, and various diagrams. A prominent graph on the left shows a line with an upward arrow, labeled with '2011' and '2012'. Another graph below it shows a wavy line. The wall is filled with various mathematical expressions and financial terms, creating a complex and detailed background.

Medida de dispersão

Medidas de dispersão

- São utilizadas para caracterizar a variabilidade dos dados;
- As medidas de dispersão mais comuns são:
 - ✓ **Amplitude;**
 - ✓ Desvio médio;
 - ✓ **Variância;**
 - ✓ **Desvio padrão;**
 - ✓ Erro padrão;
 - ✓ Coeficiente de Variação (CV).
- Quanto menor os valores menor a dispersão dos dados.



Medidas de dispersão

- **Amplitude total;**

- ✓ É a medida de dispersão ou variabilidade mais simples;
- ✓ É representada pela diferença entre o maior e o menor valor de um conjunto de observações:

$$A = x_{max} - x_{min}$$

Medidas de dispersão

- **Desvio Médio;**

- ✓ Considera a soma dos desvios absolutos de todas as observações dividido pelo tamanho da população (N) ou da amostra (n):

$$D_m = \frac{\sum_{i=1}^N |x_i - \mu|}{N} \text{ (para a população)}$$

$$D_m = \frac{\sum_{i=1}^N |x_i - \bar{x}|}{n} \text{ (para amostras)}$$

Medidas de dispersão

- **Exemplo do desvio médio;**

- ✓ A tabela apresenta as instâncias percorridas (em km) por um veículo para entrega de 10 encomendas ao longo do dia:

12,4	22,6	18,9	9,7	14,5	22,5	26,3	17,7	31,2	20,4
------	------	------	-----	------	------	------	------	------	------

$$D_m = \frac{|12,4 - 19,62| + |22,6 - 19,62| + \dots + |20,4 - 19,62|}{10}$$

$$D_m = 4,98$$

Medidas de dispersão

- **Variância;**

✓ É baseada na média dos desvios quadrados;

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

(Para a população)

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

(Para a amostra)

Medidas de dispersão

- **Exemplo da variância;**

✓ Considerando os dados do exemplo anterior...

12,4 22,6 18,9 9,7 14,5 22,5 26,3 17,7 31,2 20,4

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

$$S^2 = \frac{(12,4 - 19,62)^2 + (22,6 - 19,62)^2 + \dots + (20,4 - 19,62)^2}{9-1}$$

$$S^2 = 41,94$$

Medidas de dispersão

- **Desvio Padrão;**

- ✓ Como a variância considera a média dos desvios quadrados, seu valor tende a ser grande e de difícil interpretação;
- ✓ Para resolver este problema extrai-se a raiz quadrada para obter o desvio padrão:

$$\sigma = \sqrt{\sigma^2}$$

(Para a população)

$$S = \sqrt{S^2}$$

(Para a amostra)

Medidas de dispersão

- **Exemplo do desvio padrão:**

✓ Considere novamente os dados do exemplo anterior

12,4 22,6 18,9 9,7 14,5 22,5 26,3 17,7 31,2 20,4

$$S^2 = \frac{(12,4 - 19,62)^2 + (22,6 - 19,62)^2 + \dots + (20,4 - 19,62)^2}{9 - 1}$$

$$S^2 = 41,94$$

Logo:

$$\sqrt{41,94} = 6,476$$

Medidas de dispersão

- **Coeficiente de Variação (CV);**

- ✓ É uma medida de dispersão relativa que fornece variação dos dados em relação a média;
- ✓ Pode ser calculado como:

$$CV = \frac{\sigma}{\mu} \times 100 (\%)$$

(Para a população)

$$CV = \frac{s}{\bar{X}} \times 100 (\%)$$

(Para a amostra)

Medidas de dispersão

- **Exemplo - Coeficiente de Variação (CV);**

✓ Calculando o CV para os dados da tabela abaixo

124	111	132	142	108	127	133	144	148	105
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

$$CV = \frac{S}{\bar{X}} \times 100 (\%)$$

$$CV = \frac{15,364}{127,4} \times 100 (\%) = 12,06\%$$

Agora é com vocês!

- Qual a importância de utilizar as medidas de dispersão?
- Cite três tipos de medidas de dispersão e fale sobre cada uma delas?



[illegible][illegible][illegible][illegible][illegible][illegible]

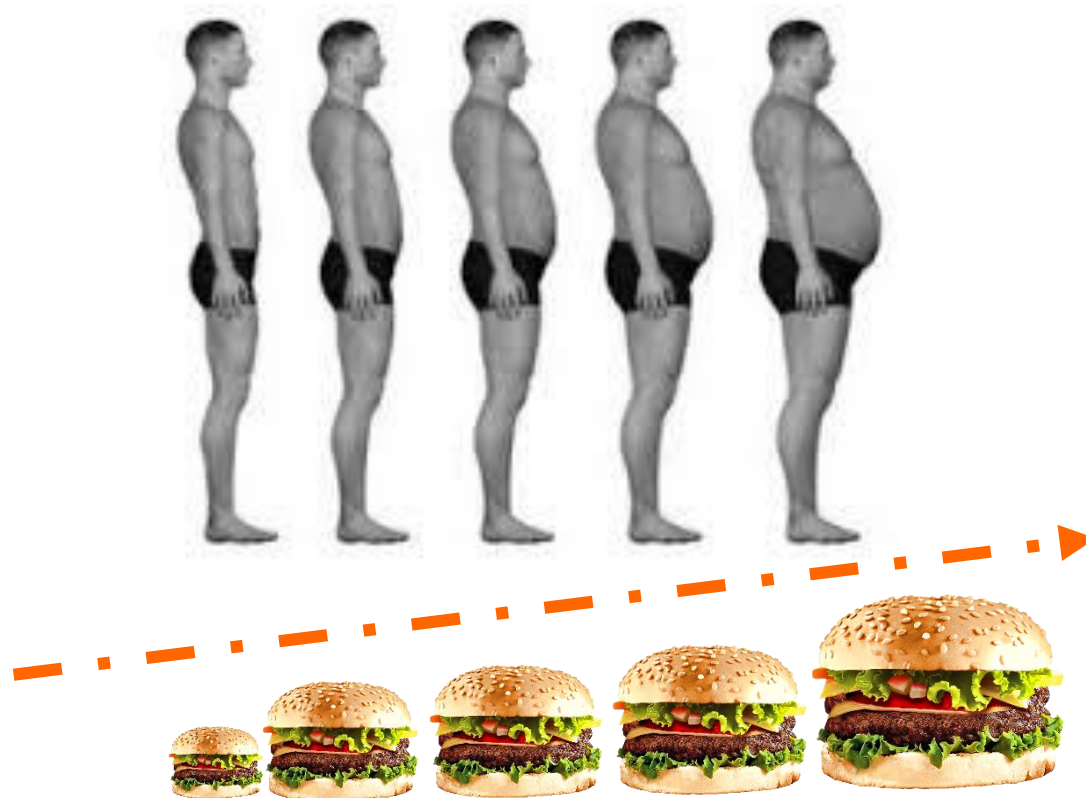
Correlação

...O que você entende
por **Correlação** ?



Correlação

...olhe pra essa imagem e pense **um pouco mais!**



Correlação

“É uma técnica estatística capaz de avaliar a existência de **relação entre duas variáveis...**
Essa relação pode ser expressa através de uma **força e direção**”



Correlação

- Mede o grau da correlação (positiva ou negativa) entre duas variáveis de escala métrica;

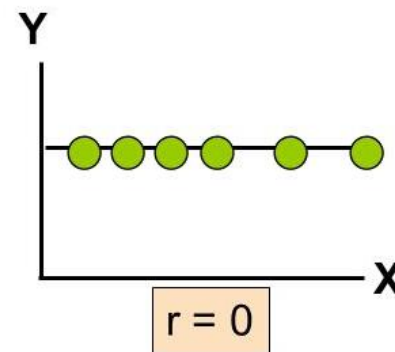
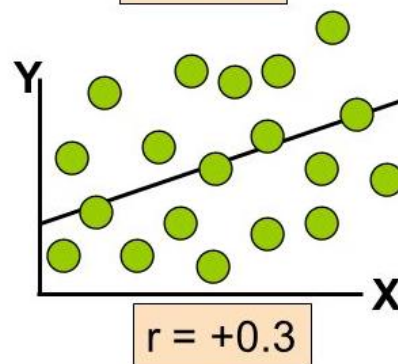
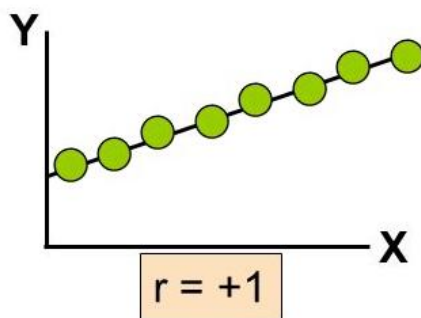
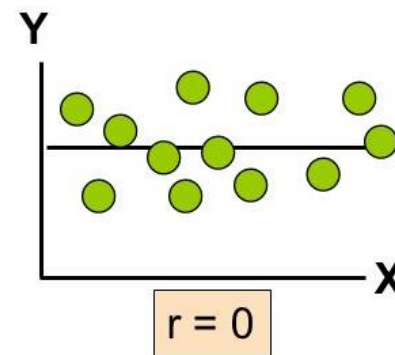
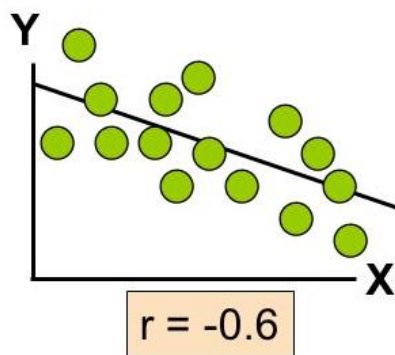
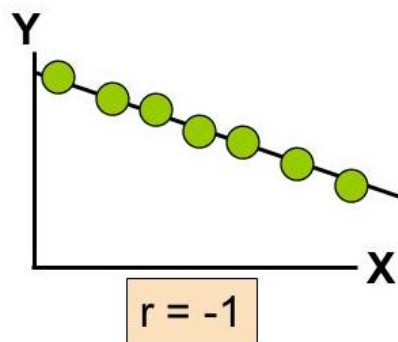
- ✓ $P > 0,4$ Significa uma correlação positiva entre as duas variáveis.
- ✓ $-0,4 < 0 < 0,4$: Significa que as duas variáveis não dependem linearmente uma da outra.
- ✓ $P < -0,4$ Significa uma correlação negativa entre as duas variáveis - Isto é, se uma aumenta, a outra sempre diminui.

Coeficiente de correlação de Pearson

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \times \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Correlação

- Representação gráfica para a correlação;



Correlação

- **Exemplo:**

- ✓ A tabela abaixo mostra duas variáveis relacionadas a profissionais de Data Science:

Profissional	Tempo de Experiência (anos)	Salário
1	0,1	R\$ 1.290,00
2	0,8	R\$ 2.330,00
3	0,6	R\$ 1.800,00
4	3	R\$ 3.452,00
5	5,1	R\$ 5.890,00
6	7,4	R\$ 6.730,00
7	8,6	R\$ 8.600,00

Correlação

- **Exemplo:**

- Inicialmente é necessário calcular a média de cada uma das variáveis:

- ✓ Média do Tempo de Experiência (anos) = 3,66;
- ✓ Média do salário = 4298,86.

- Em seguida aplica os somatórios:

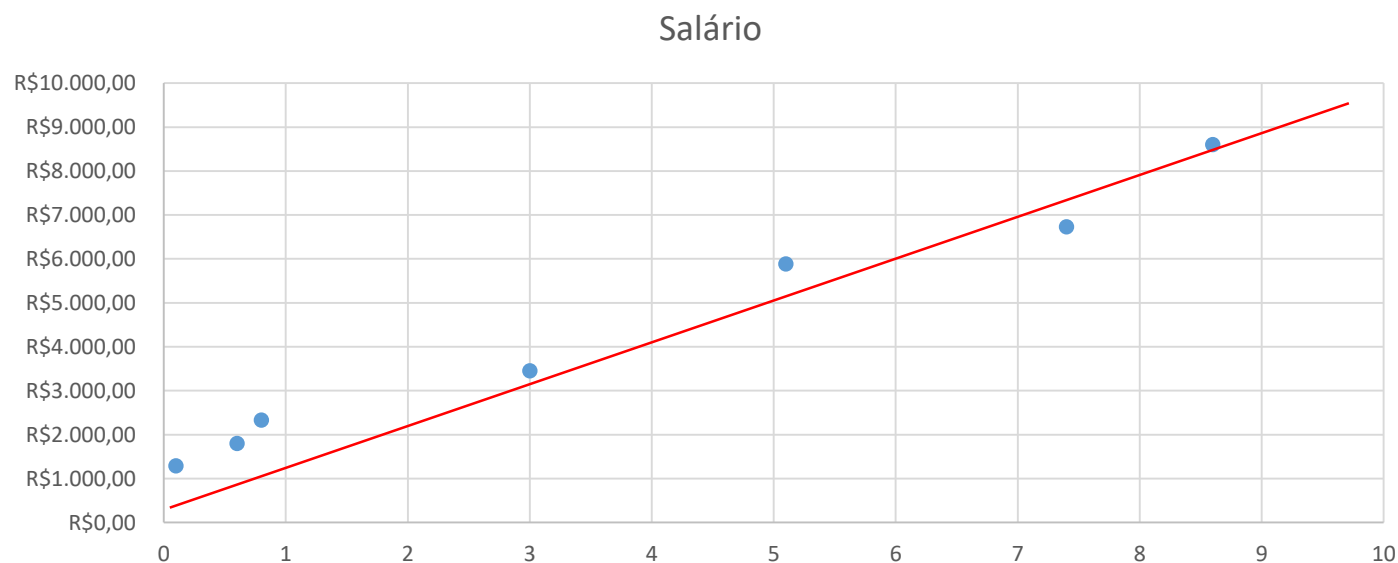
$$r = \frac{((0,1 - 3,66) * (1290 - 4298,86)) + ... + ((8,6 - 3,66) * (8600 - 4298,86))}{((0,1 - 3,66)^2 * (1290 - 4298,86)^2) + ... + ((8,6 - 3,66)^2 * (8600 - 4298,86)^2)}$$

$$r = 0,990774749$$

Correlação

- **Exemplo:**

- ✓ Este valor indica uma correlação forte e positiva entre as duas variáveis;
- ✓ Podemos ver a relação visualmente através do gráfico de dispersão:



$$r = 0,990774749$$

Agora é com vocês!

- Quais são os valores do intervalo do coeficiente de correlação de Pearson ?
- Qual o melhor gráfico pra representar a força e o sentido da correlação ?
- Qual a natureza das variáveis que são utilizadas para calcular a correlação ?



Exercício



Exercício

- Será entregue uma folha contendo dez questões abertas (Obs: fazer em grupo);
- Utilize ao máximo sua criatividade e seu poder de argumentação;



Dúvidas



- **Contatos:**

- ✓ Email: rodrigo.linsrodrigues@ufrpe.br

- ✓ Facebook: [/rodrigomuribec](https://www.facebook.com/rodrigomuribec)