

Noções de Estatística

Medidas de posição e dispersão

Objetivos

Apresentar as principais medidas de posição central: moda, média e mediana

Apresentar as principais medidas de dispersão: amplitude, variância, assimetria e curtose

The diagram shows a data table with the following structure and annotations:

- Column names:** Name, Team, Number, Position, Age, Height, Weight, College, Salary.
- Index axis=0:** Points to the row index (0-6).
- Columns axis=1:** Points to the column headers.
- Index label:** Points to the row index.
- Missing value:** Points to the 'NaN' value in the 'Number' column for row 3.
- Data:** Points to the numerical values in the 'Age', 'Weight', and 'Salary' columns for row 5.

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0.0	PG	25.0	6-2	180.0	Texas	7730337.0
1	John Holland	Boston Celtics	30.0	SG	27.0	6-5	205.0	Boston University	NaN
2	Jonas Jerebko	Boston Celtics	8.0	PF	29.0	6-10	231.0	NaN	5000000.0
3	Jordan Mickey	Boston Celtics	NaN	PF	21.0	6-8	235.0	LSU	1170960.0
4	Terry Rozier	Boston Celtics	12.0	PG	22.0	6-2	190.0	Louisville	1824360.0
5	Jared Sullinger	Boston Celtics	7.0	C	NaN	6-9	260.0	Ohio State	2569260.0
6	Evan Turner	Boston Celtics	11.0	SG	27.0	6-7	220.0	Ohio State	3425510.0

Annotations in the diagram:

- Column names:** Points to the header row.
- Columns axis=1:** Points to the column headers.
- Index label:** Points to the row index.
- Index axis=0:** Points to the row index.
- Missing value:** Points to the 'NaN' value in the 'Number' column for row 3.
- Data:** Points to the numerical values in the 'Age', 'Weight', and 'Salary' columns for row 5.

Logo: OG

Objetivos

Apresentar as principais medidas de posição central: moda, média e mediana

Apresentar as principais medidas de dispersão: amplitude, variância, assimetria e curtose

	nome	slug	apelido	atleta_id	rodada_id	clube_id	posicao_id	status_id	pontos_num	preco_num	variacao_num	media_num	jogos_num	DD	GS	PI	SG	CA	FC	FD	FS	G	A	DS	FF	I	CV	FT	GC	PP	DP	
0	Aderbar Melo dos Santos Neto	santos	Santos	69012	11	293	1	7	0.00	6.80	0.00	1.30	8	7.0	9.0	46.0	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Jean Paulo Fernandes Filho	jean	Jean	90933	11	373	1	7	-1.20	7.46	-2.24	3.58	10	14.0	15.0	73.0	2.0	1.0	1.0	1.0	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	Vagner Carmo Mancini	vagner-mancini	Vagner Mancini	39850	11	373	6	7	4.43	6.76	0.07	3.55	10	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
3	Gustavo Henrique Ferrareis	gustavo-ferrareis	Gustavo Ferrareis	89815	11	373	4	7	9.30	10.08	1.40	4.00	9	NaN	NaN	50.0	NaN	1.0	21.0	3.0	19.0	2.0	1.0	14.0	8.0	2.0	NaN	NaN	NaN	NaN	NaN	
4	Éder Ferreira Graminho	eder	Éder	91203	11	373	3	6	-1.60	3.21	-1.16	0.90	10	NaN	NaN	46.0	2.0	1.0	9.0	1.0	8.0	NaN	NaN	5.0	NaN							

Média aritmética simples

Apresentar as principais medidas de posição central: **média**, moda e mediana

Apresentar as principais medidas de dispersão: amplitude, variância, assimetria e curtose

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

Média ponderada

Apresentar as principais medidas de posição central: **média**, moda e mediana

Apresentar as principais medidas de dispersão: amplitude, variância, assimetria e curtose

$$M_p = \frac{x_1p_1 + x_2p_2 + \dots + x_np_n}{p_1 + p_2 + \dots + p_n}$$

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

p = Ponderação

Média aritmética com dados agrupados - sem intervalo de classe

Apresentar as principais medidas de posição central: **média**, moda e mediana

Apresentar as principais medidas de dispersão: amplitude, variância, assimetria e curtose

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

Quando quero sem intervalo de classe

Média aritmética com dados agrupados - sem intervalo de classe

Apresentar as principais medidas de posição central: **média**, moda e mediana

Apresentar as principais medidas de dispersão: amplitude, variância, assimetria e curtose

$$\bar{X} = \frac{\sum_{i=1}^n \bar{x}_i n_i}{n}$$

Quando quero com intervalo de classe

Moda

Apresentar as principais medidas de posição central: média, **moda** e mediana

Apresentar as principais medidas de dispersão: amplitude, variância, assimetria e curtose

Multimodal: mais de uma moda

Amodal: não existe valor predominante

Moda

Apresentar as principais medidas de posição central: média, **moda** e mediana

Apresentar as principais medidas de dispersão: amplitude, variância, assimetria e curtose

	nome	slug	apelido	atleta_id	rodada_id	clube_id	posicao_id	status_id	pontos_num	preco_num	variacao_num	media_num	jogos_num	DD	GS	PI	SG	CA	FC	FD	FS	G	A	DS	FF	I	CV	FT	GC	PP	DP	
0	Aderbar Melo dos Santos Neto	santos	Santos	69012	11	293	1	7	0.00	6.80	0.00	1.30	8	7.0	9.0	46.0	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Jean Paulo Fernandes Filho	jean	Jean	90933	11	373	1	7	-1.20	7.46	-2.24	3.58	10	14.0	15.0	73.0	2.0	1.0	1.0	1.0	1.0	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	Vagner Carmo Mancini	vagner-mancini	Vagner Mancini	39850	11	373	6	7	4.43	6.76	0.07	3.55	10	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
3	Gustavo Henrique Ferrareis	gustavo-ferrareis	Gustavo Ferrareis	89815	11	373	4	7	9.30	10.08	1.40	4.00	9	NaN	NaN	50.0	NaN	1.0	21.0	3.0	19.0	2.0	1.0	14.0	8.0	2.0	NaN	NaN	NaN	NaN	NaN	
4	Éder Ferreira Graminho	eder	Éder	91203	11	373	3	6	-1.60	3.21	-1.16	0.90	10	NaN	NaN	46.0	2.0	1.0	9.0	1.0	8.0	NaN	NaN	5.0	NaN							

Fazer no dataframe

Mediana

Apresentar as principais medidas de posição central: média, moda e **mediana**

Apresentar as principais medidas de dispersão: amplitude, variância, assimetria e curtose

	nome	slug	apelido	atleta_id	rodada_id	clube_id	posicao_id	status_id	pontos_num	preco_num	variacao_num	media_num	jogos_num	DD	GS	PI	SG	CA	FC	FD	FS	G	A	DS	FF	I	CV	FT	GC	PP	DP	
0	Aderbar Melo dos Santos Neto	santos	Santos	69012	11	293	1	7	0.00	6.80	0.00	1.30	8	7.0	9.0	46.0	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Jean Paulo Fernandes Filho	jean	Jean	90933	11	373	1	7	-1.20	7.46	-2.24	3.58	10	14.0	15.0	73.0	2.0	1.0	1.0	1.0	1.0	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	Vagner Carmo Mancini	vagner-mancini	Vagner Mancini	39850	11	373	6	7	4.43	6.76	0.07	3.55	10	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	Gustavo Henrique Ferrareis	gustavo-ferrareis	Gustavo Ferrareis	89815	11	373	4	7	9.30	10.08	1.40	4.00	9	NaN	NaN	50.0	NaN	1.0	21.0	3.0	19.0	2.0	1.0	14.0	8.0	2.0	NaN	NaN	NaN	NaN	NaN	
4	Éder Ferreira Graminho	eder	Éder	91203	11	373	3	6	-1.60	3.21	-1.16	0.90	10	NaN	NaN	46.0	2.0	1.0	9.0	1.0	8.0	NaN	NaN	5.0	NaN							

Fazer no dataframe

Dispersão

Apresentar as principais medidas de posição central: média, moda e mediana

Apresentar as principais medidas de dispersão: **amplitude**, variância, assimetria e curtose

	nome	slug	apelido	atleta_id	rodada_id	clube_id	posicao_id	status_id	pontos_num	preco_num	variacao_num	media_num	jogos_num	DD	GS	PI	SG	CA	FC	FD	FS	G	A	DS	FF	I	CV	FT	GC	PP	DP	
0	Aderbar Melo dos Santos Neto	santos	Santos	69012	11	293	1	7	0.00	6.80	0.00	1.30	8	7.0	9.0	46.0	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Jean Paulo Fernandes Filho	jean	Jean	90933	11	373	1	7	-1.20	7.46	-2.24	3.58	10	14.0	15.0	73.0	2.0	1.0	1.0	1.0	1.0	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	Vagner Carmo Mancini	vagner-mancini	Vagner Mancini	39850	11	373	6	7	4.43	6.76	0.07	3.55	10	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	Gustavo Henrique Ferrareis	gustavo-ferrareis	Gustavo Ferrareis	89815	11	373	4	7	9.30	10.08	1.40	4.00	9	NaN	NaN	50.0	NaN	1.0	21.0	3.0	19.0	2.0	1.0	14.0	8.0	2.0	NaN	NaN	NaN	NaN	NaN	
4	Éder Ferreira Graminho	eder	Éder	91203	11	373	3	6	-1.60	3.21	-1.16	0.90	10	NaN	NaN	46.0	2.0	1.0	9.0	1.0	8.0	NaN	NaN	5.0	NaN							

Fazer no dataframe

Dispersão

Apresentar as principais medidas de posição central: média, moda e mediana

Apresentar as principais medidas de dispersão: amplitude, **variância**, assimetria e curtose

	nome	slug	apelido	atleta_id	rodada_id	clube_id	posicao_id	status_id	pontos_num	preco_num	variacao_num	media_num	jogos_num	DD	GS	PI	SG	CA	FC	FD	FS	G	A	DS	FF	I	CV	FT	GC	PP	DP	
0	Aderbar Melo dos Santos Neto	santos	Santos	69012	11	293	1	7	0.00	6.80	0.00	1.30	8	7.0	9.0	46.0	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Jean Paulo Fernandes Filho	jean	Jean	90933	11	373	1	7	-1.20	7.46	-2.24	3.58	10	14.0	15.0	73.0	2.0	1.0	1.0	1.0	1.0	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	Vagner Carmo Mancini	vagner-mancini	Vagner Mancini	39850	11	373	6	7	4.43	6.76	0.07	3.55	10	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	Gustavo Henrique Ferrareis	gustavo-ferrareis	Gustavo Ferrareis	89815	11	373	4	7	9.30	10.08	1.40	4.00	9	NaN	NaN	50.0	NaN	1.0	21.0	3.0	19.0	2.0	1.0	14.0	8.0	2.0	NaN	NaN	NaN	NaN	NaN	
4	Éder Ferreira Graminho	eder	Éder	91203	11	373	3	6	-1.60	3.21	-1.16	0.90	10	NaN	NaN	46.0	2.0	1.0	9.0	1.0	8.0	NaN	NaN	5.0	NaN							

Para uma população

Fazer no dataframe

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

Dispersão

Apresentar as principais medidas de posição central: média, moda e mediana

Apresentar as principais medidas de dispersão: amplitude, **variância**, assimetria e curtose

	nome	slug	apelido	atleta_id	rodada_id	clube_id	posicao_id	status_id	pontos_num	preco_num	variacao_num	media_num	jogos_num	DD	GS	PI	SG	CA	FC	FD	FS	G	A	DS	FF	I	CV	FT	GC	PP	DP	
0	Aderbar Melo dos Santos Neto	santos	Santos	69012	11	293	1	7	0.00	6.80	0.00	1.30	8	7.0	9.0	46.0	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Jean Paulo Fernandes Filho	jean	Jean	90933	11	373	1	7	-1.20	7.46	-2.24	3.58	10	14.0	15.0	73.0	2.0	1.0	1.0	1.0	1.0	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	Vagner Carmo Mancini	vagner-mancini	Vagner Mancini	39850	11	373	6	7	4.43	6.76	0.07	3.55	10	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
3	Gustavo Henrique Ferrareis	gustavo-ferrareis	Gustavo Ferrareis	89815	11	373	4	7	9.30	10.08	1.40	4.00	9	NaN	NaN	50.0	NaN	1.0	21.0	3.0	19.0	2.0	1.0	14.0	8.0	2.0	NaN	NaN	NaN	NaN	NaN	
4	Éder Ferreira Graminho	eder	Éder	91203	11	373	3	6	-1.60	3.21	-1.16	0.90	10	NaN	NaN	46.0	2.0	1.0	9.0	1.0	8.0	NaN	NaN	5.0	NaN							

Para uma amostra

Fazer no dataframe

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

Dispersão

Apresentar as principais medidas de posição central: média, moda e mediana

Apresentar as principais medidas de dispersão: amplitude, **variância** (**Desvio Padrão**), assimetria e curtose

	nome	slug	apelido	atleta_id	rodada_id	clube_id	posicao_id	status_id	pontos_num	preco_num	variacao_num	media_num	jogos_num	DD	GS	PI	SG	CA	FC	FD	FS	G	A	DS	FF	I	CV	FT	GC	PP	DP	
0	Aderbar Melo dos Santos Neto	santos	Santos	69012	11	293	1	7	0.00	6.80	0.00	1.30	8	7.0	9.0	46.0	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Jean Paulo Fernandes Filho	jean	Jean	90933	11	373	1	7	-1.20	7.46	-2.24	3.58	10	14.0	15.0	73.0	2.0	1.0	1.0	1.0	1.0	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	Vagner Carmo Mancini	vagner-mancini	Vagner Mancini	39850	11	373	6	7	4.43	6.76	0.07	3.55	10	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
3	Gustavo Henrique Ferrareis	gustavo-ferrareis	Gustavo Ferrareis	89815	11	373	4	7	9.30	10.08	1.40	4.00	9	NaN	NaN	50.0	NaN	1.0	21.0	3.0	19.0	2.0	1.0	14.0	8.0	2.0	NaN	NaN	NaN	NaN	NaN	
4	Éder Ferreira Graminho	eder	Éder	91203	11	373	3	6	-1.60	3.21	-1.16	0.90	10	NaN	NaN	46.0	2.0	1.0	9.0	1.0	8.0	NaN	NaN	5.0	NaN							

Para uma amostra

Fazer no dataframe

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

Dispersão: coeficiente de variação (CV)

$$CV = \frac{S}{\bar{x}}$$

Diz-se que uma distribuição tem:

Baixa dispersão: $CV \leq 15\%$

Média dispersão: $15\% <$

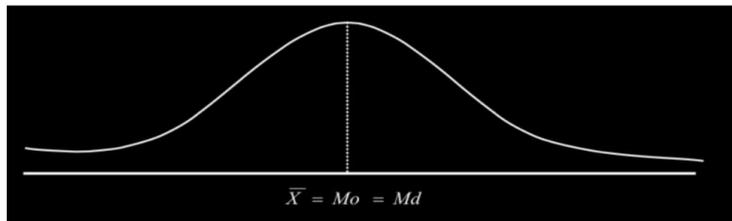
$CV < 30\%$

Alta dispersão: $CV \geq 30\%$

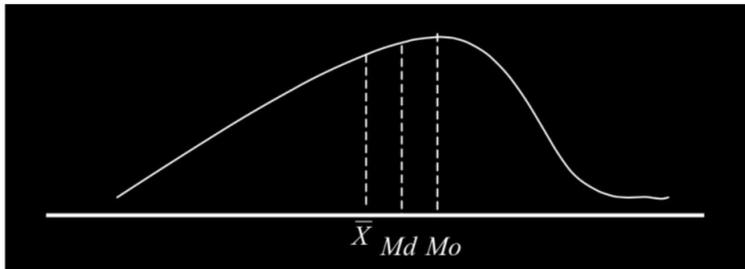
Assimetria:

Assimetria

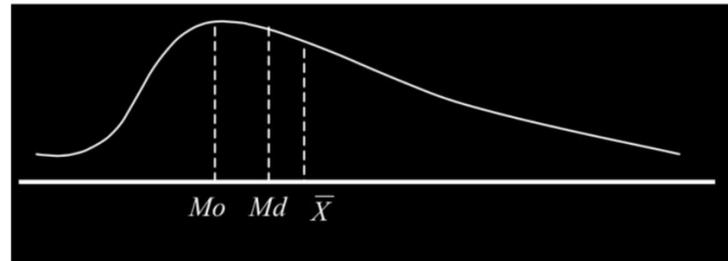
- Curva Simétrica



- Assimetria à Esquerda (Negativa)



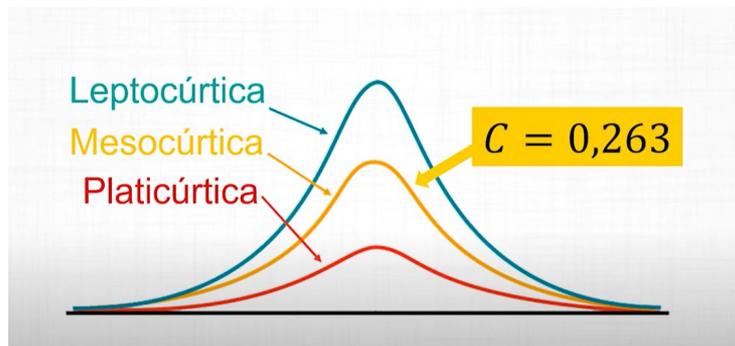
- Assimetria à Direita (Positiva)



Fim:

Espera aí! E o curtose?

Espera aí! E o curtose?



$$K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} \left. \begin{array}{l} K = 0,263 - \text{Distribuição Mesocúrtica} \\ K > 0,263 - \text{Distribuição Platicúrtica} \\ K < 0,263 - \text{Distribuição Leptocúrtica} \end{array} \right\}$$

Onde:

K = Coeficiente de curtose

Q_1 = 1º quartil

Q_3 = 3º quartil

P_{90} = Percentil 90

P_{10} = Percentil 10