

Média

■ Equação:

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

Problema: sensível a *outliers*

Bom indicador apenas se valores são distribuídos simetricamente

Mediana

■ Passos:

- Ordenar os valores de forma crescente
- Calcular a equação:

$$mediana(\mathbf{x}) = \begin{cases} \frac{1}{2} (x_r + x_{r+1}) & \text{se } n \text{ for par } (n = 2r) \\ x_{r+1} & \text{se } n \text{ for ímpar } (n = 2r + 1) \end{cases}$$

Facilita observar se distribuição é assimétrica ou se existem *outliers*

Mediana

■ Exemplos:

■ {17, 4, 8, 21, 4}

■ Ordenando: 4, 4, 8, 17, 21

■ Número ímpar de elementos \Rightarrow mediana = 8

■ Valor do meio na ordenação

■ {17, 4, 8, 21, 4, 15, 13, 9}

■ Ordenando: 4, 4, 8, 9, 13, 15, 17, 21

■ Número par de elementos \Rightarrow mediana = $(9+13)/2 = 11$

■ Média dos dois valores do meio na ordenação

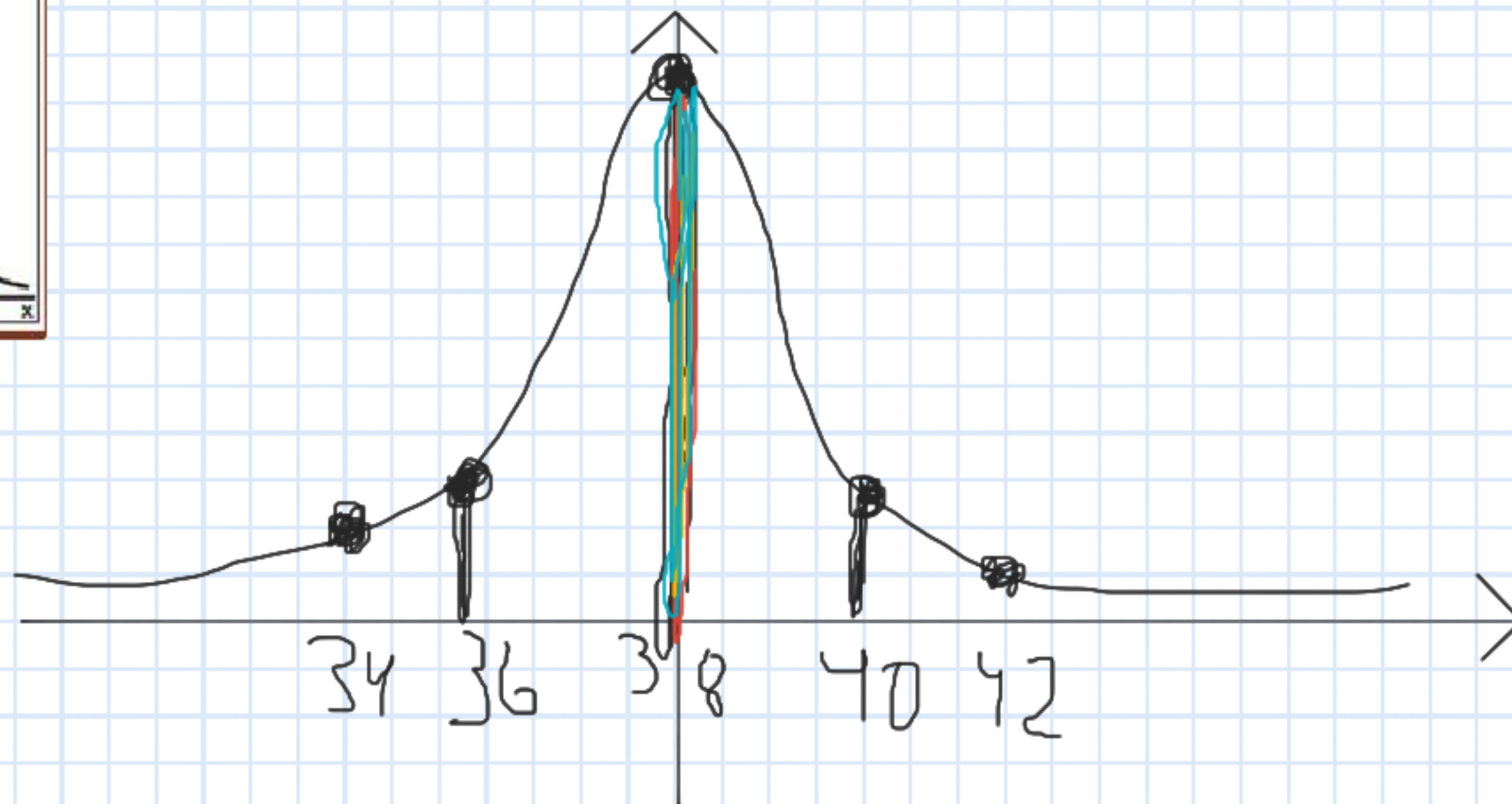
Média e mediana

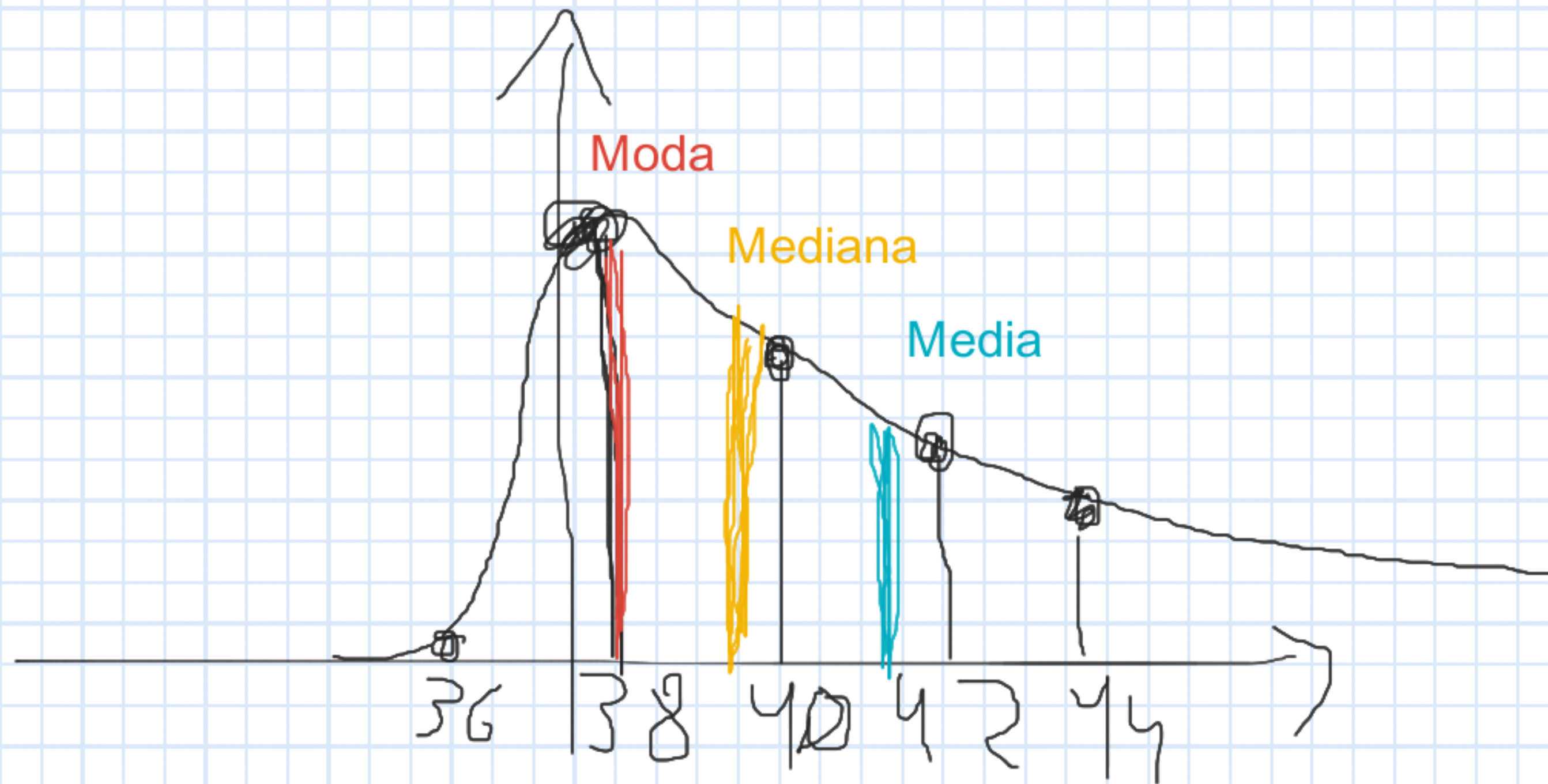
■ Ex. conjunto de dados hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Grandes	38,0	2	SP	Doente
3217	Maria	18	F	67	Pequenas	39,5	4	MG	Doente
4039	Luiz	49	M	92	Grandes	38,0	2	RS	Saudável
1920	José	18	M	43	Grandes	38,5	20	MG	Doente
4340	Cláudia	21	F	52	Médias	37,6	1	PE	Saudável
2301	Ana	22	F	72	Pequenas	38,0	3	RJ	Doente
1322	Marta	19	F	87	Grandes	39,0	6	AM	Doente
3027	Paulo	34	M	67	Médias	38,4	2	GO	Saudável

Média: 26,1

Mediana: 21,5





Média e mediana

■ Ex. conjunto de dados hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Grandes	38,0	2	SP	Doente
3217	Maria	18	F	67	Pequenas	39,5	4	MG	Doente
4039	Luiz	49	M	92	Grandes	38,0	2	RS	Saudável
1920	José	18	M	43	Grandes	38,5	20	MG	Doente
4340	Cláudia	21	F	52	Médias	37,6	1	PE	Saudável
2301	Ana	22	F	72	Pequenas	38,0	3	RJ	Doente
1322	Marta	19	F	87	Grandes	39,0	6	AM	Doente
3027	Paulo	34	M	67	Médias	38,4	2	GO	Saudável

Média: 5
Mediana: 2,5

Média truncada

- Descarta elementos extremos da sequência ordenada de valores

- Minimizar problemas da média
- Necessário definir porcentagem



- Passos:

- Definir porcentagem p
- Ordenar valores
- Descartar $(p/2)\%$ de valores de cada extremo
- Calcular a média dos exemplos restantes

Média truncada

■ Ex. conjunto de dados hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Grandes	38,0	2	SP	Doente
3217	Maria	18	F	67	Pequenas	39,5	4	MG	Doente
4039	Luiz	49	M	92	Grandes	38,0	2	RS	Saudável
1920	José	18	M	43	Grandes	38,5	20	MG	Doente
4340	Cláudia	21	F	52	Médias	37,6	1	PE	Saudável
2301	Ana	22	F	72	Pequenas	38,0	3	RJ	Doente
1322	Marta	19	F	87	Grandes	39,0	6	AM	Doente
3027	Paulo	34	M	67	Médias	38,4	2	GO	Saudável

Média: 26,1

Mediana: 21,5

Média truncada ($p = 25\%$): 23,7

Média truncada

■ Ex. conjunto de dados hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Grandes	38,0	2	SP	Doente
3217	Maria	18	F	67	Pequenas	39,5	4	MG	Doente
4039	Luiz	49	M	92	Grandes	38,0	2	RS	Saudável
1920	José	18	M	43	Grandes	38,5	20	MG	Doente
4340	Cláudia	21	F	52	Médias	37,6	1	PE	Saudável
2301	Ana	22	F	72	Pequenas	38,0	3	RJ	Doente
1322	Marta	19	F	87	Grandes	39,0	6	AM	Doente
3027	Paulo	34	M	67	Médias	38,4	2	GO	Saudável

Média: 5

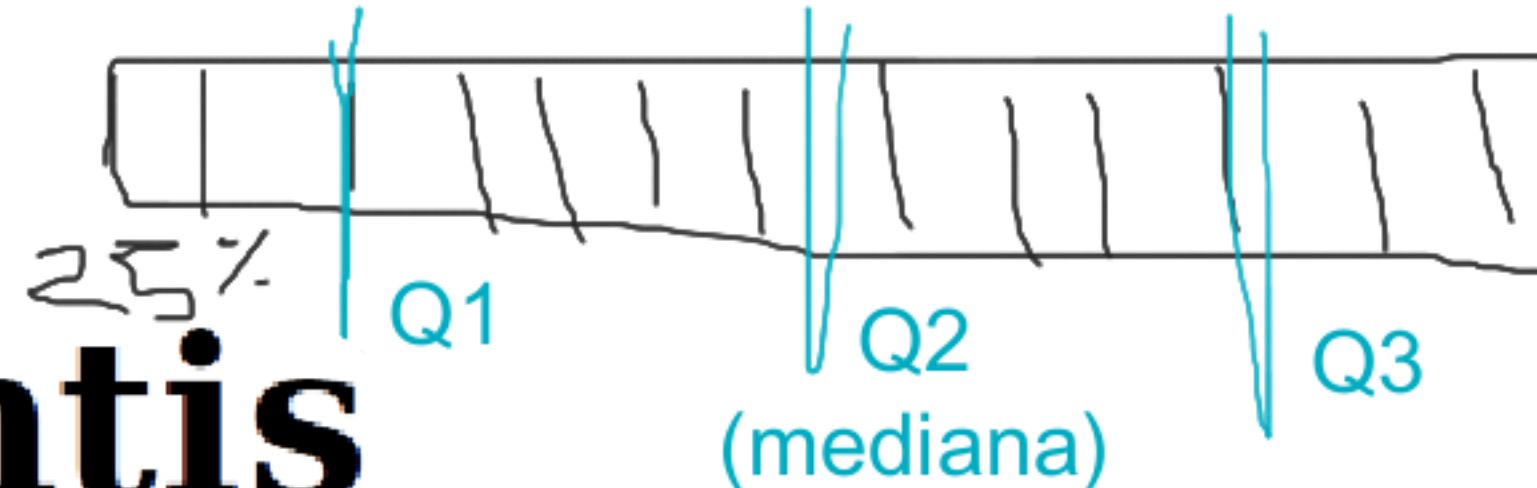
Mediana: 2,5

Média truncada ($p = 25\%$): 3,2

Exercícios

- Dado o conjunto de dados {1, 2, 3, 4, 5, 80}, calcular:
 - Média
 - Mediana
 - Média truncada com $p = 33\%$

Quartis e percentis



■ Mediana divide dados ordenados ao meio

■ Quartis e percentis usam pontos de divisão diferentes

Quartis

- Divide em quartos
- 1º quartil ($Q1$) \Rightarrow valor que tem 25% dos demais valores abaixo dele
- 2º quartil = mediana

Percentil

- Para p entre 0 e 100
- p° percentil = $Pp \Rightarrow x_i$ tal que $p\%$ dos valores observados são menores do que x_i
- $P25 = Q1$
- $P50 = Q2 = \text{mediana}$

Exercício:

**Escreva
funções em
python para
computar:**

*** Média**

*** Média
Truncada**

*** Mediana**

**Bônus:
Percentil**

Quartil e percentil

■ Ex. conjunto de dados hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Grandes	38,0	2	SP	Doente
3217	Maria	18	F	67	Pequenas	39,5	4	MG	Doente
4039	Luiz	49	M	92	Grandes	38,0	2	RS	Saudável
1920	José	18	M	43	Grandes	38,5	20	MG	Doente
4340	Cláudia	21	F	52	Médias	37,6	1	PE	Saudável
2301	Ana	22	F	72	Pequenas	38,0	3	RJ	Doente
1322	Marta	19	F	87	Grandes	39,0	6	AM	Doente
3027	Paulo	34	M	67	Médias	38,4	2	GO	Saudável

(18,18,19,21,22,28,34,49)

Média: 26,1

Mediana: 21,5

Média truncada (p= 25%): 23,7

Q1: 18,5; Q2: 21,5; Q3: 31

P40: 21

Quartil e percentil

■ Ex. conjunto de dados hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Grandes	38,0	2	SP	Doente
3217	Maria	18	F	67	Pequenas	39,5	4	MG	Doente
4039	Luiz	49	M	92	Grandes	38,0	2	RS	Saudável
1920	José	18	M	43	Grandes	38,5	20	MG	Doente
4340	Cláudia	21	F	52	Médias	37,6	1	PE	Saudável
2301	Ana	22	F	72	Pequenas	38,0	3	RJ	Doente
1322	Marta	19	F	87	Grandes	39,0	6	AM	Doente
3027	Paulo	34	M	67	Médias	38,4	2	GO	Saudável

Média: 5

Mediana: 2,5

Média truncada (p= 25%): 3,2

Q1: 2; Q2: 2,5; Q3: 5

P40: 2

Quartil e percentil

■ Exercício conjunto de dados hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.
4201	João	28	M	79	Grandes	38,0
3217	Maria	18	F	67	Pequenas	39,5
4039	Luiz	49	M	92	Grandes	38,0
1920	José	18	M	43	Grandes	38,5
4340	Cláudia	21	F	52	Médias	37,6
2301	Ana	22	F	72	Pequenas	38,0
1322	Marta	19	F	87	Grandes	39,0
3027	Paulo	34	M	67	Médias	38,4

PESO

Média:

Mediana:

Média truncada (p= 25%):

Q1: Q2: Q3:

P30:

Quartil e percentil

■ Exercício conjunto de dados hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.
4201	João	28	M	79	Grandes	38,0
3217	Maria	18	F	67	Pequenas	39,5
4039	Luiz	49	M	92	Grandes	38,0
1920	José	18	M	43	Grandes	38,5
4340	Cláudia	21	F	52	Médias	37,6
2301	Ana	22	F	72	Pequenas	38,0
1322	Marta	19	F	87	Grandes	39,0
3027	Paulo	34	M	67	Médias	38,4

PESO

Média:

Mediana:

Média truncada (p= 25%):

Q1: Q2: Q3:

P30: