3.2-Medidas de localização e de dispersão.

(Média, moda, mediana, quartis, percentis, amplitude, variância, desvio padrão, extremos e quartis, Python na estatística.)

Cálculo de estatísticas.

Medidas de localização/Medidas de dispersão.

Nota: As medidas de localização indicam onde estão localizados ou centrados os dados.
As medidas de dispersão informam-nos se os dados estão próximos uns dos outros, ou se estão afastados, ou dispersos entre si.

Medidas de localização.

Medidas de tendência central: (média, moda e mediana).

Média.

Média ($\bar{x}$) é o quociente da soma de todos os dados (valores da variável) pelo número de dados, ou seja:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Exemplo

Consideremos os dados de uma amostra: 4; 5; 6.

Calculemos a média amostral.

$$\bar{x} = \frac{4+5+6}{3} = \frac{15}{3} = 5$$

Nota: para a média de uma **amostra** usamos o símbolo $\bar{x}$. Quando de trata de uma **população**, o valor médio populacional é representado por $\boldsymbol{\mu}$.

Nota: quando os dados estão agrupados, a média pode ser dada por:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_M f_M}{n}$$

O que seria equivalente a multiplicar cada número pela respetiva frequência relativa:

$$\bar{x} = x_1 fr_1 + x_2 fr_2 + \ldots + x_M fr_M.$$

Recorde que $fr_1 = f_1/n$.

Exemplo Dados agrupados.
Consideremos os valores de xi e as respetivas frequências absolutas:

x_i	fi
4	10
5	15
6	5
Total	**30**

A média será dada por :

$$\bar{x} = \frac{10 \times 4 + 15 \times 5 + 5 \times 6}{30} \approx 4.83$$

Nota: no caso de termos dados **agrupados em classes** na forma de **intervalos**, apenas podemos calcular uma aproximação do valor da média. Para tal, usamos a marca da classe e a frequência respetiva.

A **marca da classe** (m_i) obtém-se fazendo a média dos valores extremos do intervalo considerado.

Exemplo: a marca da classe [10, 20[é (10+20) / 2 =15.
A marca da classe [20, 30[é (20+30) / 2 =25

Média:
(Dados agrupados em classes na forma de intervalos)

$$\bar{x} \approx \frac{m_1 f_1 + m_2 f_2 + \ldots + m_M f_M}{n}$$

Onde m_1, m_2, …m_M são as marcas das várias classes, f_1, f_2, .. f_n são as frequências absolutas e n é o número total de elementos.

Exemplo

Numa empresa fez-se um estudo sobre o tempo que os seus empregados demoravam no percurso de casa para o emprego e obtiveram-se os resultados que constam na tabela seguinte:
Calculemos o valor aproximado da média.

Tempo (Minutos)	Número de empregados
[10, 20[	12
[20, 30[	6
[30, 40[	6
[40, 50[	7
[50, 60[	6
[60, 70[	3

Temos de recorrer às marcas de cada uma das classes.
Estas são respetivamente 15; 25; 35; 45; 55; 65

Média: $\bar{x} = \frac{12\times15+6\times25+6\times35+7\times45+6\times55+3\times65}{40} \approx 34.5$

O valor aproximado da média é 34.5 minutos.

Nota: observemos duas propriedades da média a partir do exemplo seguinte:

Exemplo consideremos os números 1; 2; 3.
A sua média é (1+2+3)/3 = 6/3 = 2.

Se, a cada um dos números ***somarmos*** cinco unidades:

1+**5**=6; 2+**5**=7; 3+**5**=8 ,

a nova média será (6+7+8)/3 = 21/3 = 7.
A nova média também virá somada de 5 unidades, pois 2+5=7

Do mesmo modo, se ***multiplicarmos*** todos os números por 4, teremos:

1×**4**= 4; 2×**4**= 8; 3×**4**= 12.

A nova média será (4+8+12)/3 = 24/3 = 8.
A nova média também virá multiplicada por 4 unidades, pois 2×4=8.

Propriedades da média.

1) Se a cada valor da variável x adicionarmos uma constante c≠ 0, obtemos uma nova variável, y , cuja média é $\bar{y} = \bar{x}$ **+ c** .

Exemplo
A média de uma amostra é 5. Se somarmos 4 unidades a todos os elementos dessa amostra, a média passará a ser 5+4=9

2) Se multiplicarmos cada valor da variável x por uma constante c≠ 0, obtemos uma nova variável, y , cuja média é $\bar{y} = \bar{x}$ × **c** .

Exemplo
A média de uma amostra é 5. Se multiplicarmos por 4 todos os elementos dessa amostra, a média passará a ser 5×4=20

Moda

Moda (M_o) é o valor da variável ao qual corresponde uma maior frequência (absoluta ou relativa).

Exemplo
Consideremos os números: 1; 1; 2; 2; 2; 2; 3; 4; 5; 6.
A moda é 2.porque é o número que tem a maior frequência absoluta.

Se todos os valores da variável têm a mesma frequência, diz-se **amodal**.

Exemplo

1; 1; 2; 2; 3; 3; 4; 4; 5; 5; 6; 6. Esta amostra é amodal, pois todos os elementos têm frequência 2. Nenhum é mais frequente que os outros.

No caso de existirem vários valores com a frequência mais alta, diz-se que a amostra é **plurimodal**.

Exemplo

1; 1; 1; 1; 2; 2; 3; 3;3; 3; 4; 4; 4; 4; 5; 5; 6; 6.

Esta amostra é plurimodal pois o 1, o 3 e o 4 têm a maior frequência.

Se existirem exatamente dois valores com a maior frequência — diz-se **bimodal**.

Exemplo

1; 1; 1; 1; 2; 2; 3; 3;3; 3; 4; 4; 5; 5; 6; 6.

Esta amostra é bimodal pois o 1, o 3 têm a maior frequência.

Se os dados estiverem agrupados em classes indicaremos a **classe modal.**

Classe modal é a classe à qual corresponde a maior frequência.

Exemplo: nos dados da tabela seguinte,

Tempo (Minutos)	Número de empregados
[10, 20[	12
[20, 30[	6
[30, 40[	6
[40, 50[	7
[50, 60[	6
[60, 70[	3

A classe modal é [10, 20[pois é a classe com maior frequência.

Mediana.

A **mediana** ($\tilde{x}$ ou med. ou M_e) é o valor que divide o conjunto de dados (depois de ordenados) em duas partes com o mesmo número de observações.

-Se o número de dados é **ímpar**, a mediana é o <u>valor central</u>.

-Se o número de dados é **par**, a mediana é a <u>média aritmética dos valores centrais</u>.

Exemplo 1

10; 13; **15**; 18; 19; $\tilde{x} = 15$

Exemplo 2

10; 13; **<u>15; 18</u>**; 19; 22 $\tilde{x} = \frac{15+18}{2} = 16.5$

Nota: se o número de dados for muito grande, pode não ser prático apresentar todos. Será mais adequado procurar a ordem K da mediana.

Se o número de dados, *n,* é **ímpar**, a ordem k da mediana é dada por $\boldsymbol{k} = \frac{\boldsymbol{n+1}}{\boldsymbol{2}}$

Se o número de dados, *n,* é **par**, a mediana corresponde à média dos valores de ordens $\frac{\boldsymbol{n}}{\boldsymbol{2}}$ e $\frac{\boldsymbol{n}}{\boldsymbol{2}} + \boldsymbol{1}$.

Exemplo

Consideremos os pesos de 34 pessoas, cujos valores já estão ordenados:

35 39 41 43 49 51 51 51 55 58 60 60 60 61 65 68 69 74 75 76 76 77 78 79 79 86 87 88 88 89 90 90 95 95

Como são 34, que é um número par, a mediana será a média dos valores de ordens $\frac{\boldsymbol{34}}{\boldsymbol{2}}$ e $\frac{\boldsymbol{34}}{\boldsymbol{2}}$+1 **isto é, a média entre o 17º e 18º elementos ordenados.**

35 39 41 43 49 51 51 51 55 58 60 60 60 61 65 68 **<u>69 74</u>** 75 76 76 77 78 79 79 86 87 88 88 89 90 90 95 95

Neste caso, corresponde a $\frac{\boldsymbol{69+74}}{\boldsymbol{2}} = \boldsymbol{71.5}$

Exemplo

Consideremos os pesos de 35 pessoas, cujos valores já estão ordenados:

35 39 41 43 49 51 51 51 55 58 60 60 60 61 65 68 69 74 75 76 76 77 78 79 79 86 87 88 88 89 90 90 95 95 97

Como são 35, que é um número ímpar, a mediana será o valor ordenado de ordem $\frac{\boldsymbol{35+1}}{\boldsymbol{2}} = 18$ 18º elemento ordenado.

35 39 41 43 49 51 51 51 55 58 60 60 60 61 65 68 69 **<u>74</u>** 75 76 76 77 78 79 79 86 87 88 88 89 90 90 95 95 97

Neste caso, corresponde a 74.

Nota: Se os **dados** estiverem **agrupados** em classes na forma de **intervalos**, determinaremos a classe mediana.

Classe mediana é a classe à qual pertence a mediana.

Exemplo
Consideremos os dados agrupados da tabela seguinte:

Tempo (Minutos)	Número de Empregados (f_i)
[10, 20[	12
[20, 30[	6
[30, 40[	6
[40, 50[	7
[50, 60[	6
[60, 70[	3
Total:	40

Pode ser útil acrescentar uma coluna com as frequências acumuladas (Fi) para facilitar a identificação da classe mediana.

Tempo (Minutos)	Número de Empregados (f_i)	Fi (freq. Acumulada)
[10, 20[	12	12
[20, 30[	6	18
[30, 40[	6	24
[40, 50[	7	31
[50, 60[	6	37
[60, 70[	3	40
Total:	40	

Como o número total de elementos é 40, a mediana fica situada entre os 20º e 21º elementos ordenados.

Repare que a classe [20, 30[apenas inclui até ao 18º elemento ordenado. A classe [30, 40[inclui desde o 19º elemento ordenado até ao 24º.

Assim, os 20º e 21º elementos ordenados pertencem à classe [30, 40[.

Conclusão [30, 40[é a classe mediana.

Nota: Para identificarmos a **classe mediana**, determinamos as frequências acumuladas e procuramos a primeira classe que tem frequência acumulada maior ou igual a **n/2**(ou 50%).

Exemplo

Vamos agora calcular um valor aproximado para a mediana, quando os dados estão agrupados em intervalos de classe.

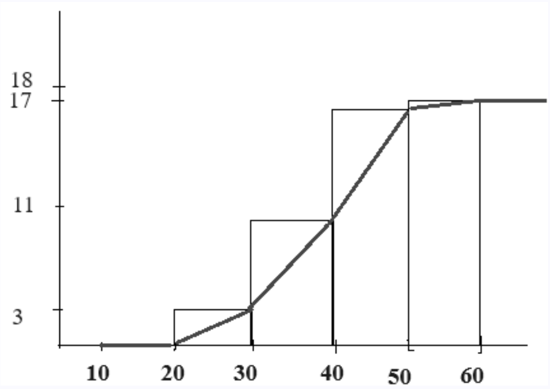
Num grupo de 18 alunos foi analisado o tempo que demoraram a realizar uma tarefa na disciplina de Matemática. Os valores estão na tabela seguinte:

Tempo (em minutos)	[20, 30[	[30, 40[	[40, 50[	[50, 60[
Número de alunos (fi)	3	8	6	1

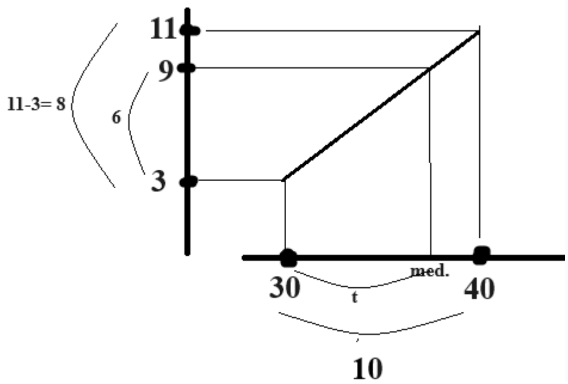
Neste caso, como temos os valores em intervalos de classe, e não temos os valores originais, vamos basear-nos nas frequências acumuladas (última linha):

Tempo (em minutos)	[20, 30[	[30, 40[	[40, 50[	[50, 60[
Número de alunos (fi)	3	8	6	1
Freq. Abs. Acumulada (Fi)	3	11	17	18

e procurar a classe que inclui o n/2, isto é , o elemento de ordem 9.
Corresponde à classe [30, 40[.
A partir daqui, construímos o histograma das frequências absolutas acumuladas.



Lembremos que:
o total de elementos é 18,
18/2=9, e que
a mediana está na classe [30, 40[.
A classe anterior acumula 3 elementos. A classe [30, 40[acumula 3+8=11 elementos.
Vamos determinar geometricamente uma aproximação do 9º elemento.



Usamos a regra dos 3 simples para calcular o ***t***.

10→ 8

t→ 6 t=(10×69/8 ⇔ t= 7.5 med.≈ 30+7.5 ≈37.5

Logo, o valor aproximado da mediana é 37.5

Vantagens, desvantagens e limitações das medidas de tendência central

Exemplo: Consideremos os números: 1, 1, 3, 4, 61

Média: 14

Mediana: 3

Se substituirmos o 61 por 6:

1, 1, 3, 4, 6

Obtemos:

Média: 3

Mediana: 3

Nota: basta mudar o último elemento para a média mudar significativamente o seu valor. Curiosamente a mediana ficou inalterada apesar da alteração de 61 para 6.

Dizemos que:

A média é muito **sensível** a valores discrepantes.

A **mediana** é uma medida **resistente**.

Exemplo

Numa empresa com 9 funcionários, incluindo o gestor, os ordenados são (em euros):

900; 900; 950; 950; 960; 990; 1100; 1150; 9920

A média é 1980 €

e a mediana é 960 €.

Olhando para os dados, verificamos facilmente que o gestor tem um vencimento muito acima dos restantes funcionários e por essa razão a média é 1980.

A mediana, 960, representa melhor a realidade destes vencimentos para a maioria dos funcionários.

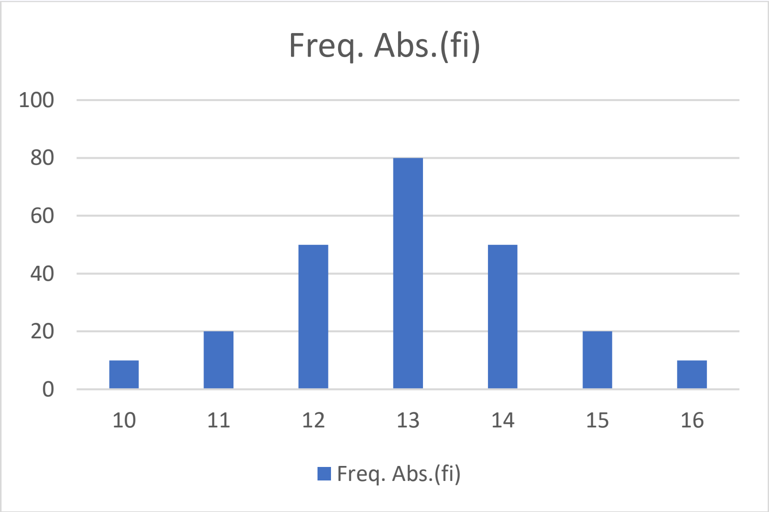
Simetria da distribuição

Exemplo

Considere o as idades de um grupo alunos que participaram numa atividade desportiva:

Idade(anos)	Freq. Abs.(fi)
10	10
11	20
12	50
13	80
14	50
15	20
16	10

Graficamente:



Média: 13   mediana 13

A média igual à mediana

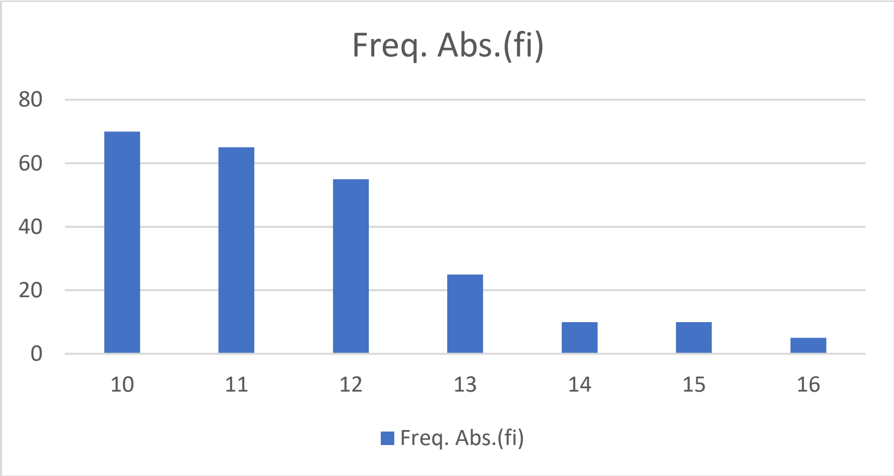
Nota: neste caso, dizemos que a distribuição é **simétrica.**

Exemplo

Considere o as idades de um outro grupo alunos que participaram numa atividade desportiva:

Idade(anos)	Freq. Abs.(fi)
10	70
11	65
12	55
13	25
14	10
15	10
16	5

Graficamente:



Média 11.54   mediana: 11

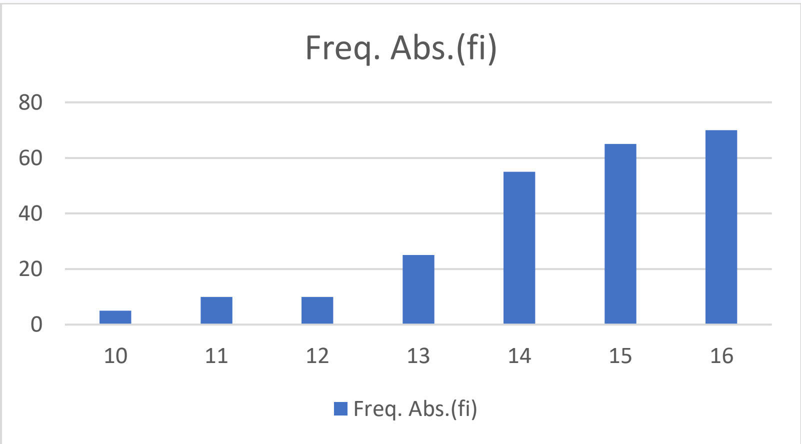
A média é maior do que a mediana.

Nota: neste caso, dizemos que a distribuição é **assimétrica positiva, ou enviesada** para a direita.

Exemplo

Considere ainda as idades de um outro grupo alunos que participaram numa atividade desportiva:

Idade(anos)	Freq. Abs.(fi)
10	5
11	10
12	10
13	25
14	55
15	65
16	70



Média 14.45   mediana: 15

A média é menor que a mediana.

Nota: neste caso, dizemos que a distribuição é **assimétrica negativa, ou enviesada** para a esquerda.

Outras medidas de localização (quartis e percentis)

Quartis.

De modo semelhante à mediana, que divide os dados ordenados em duas partes com o mesmo número de elementos, os quartis dividem a distribuição dos dados ordenados em quatro partes, de modo que cada uma das partes contenha o mesmo número de observações.

Exemplo

Se tivermos os números de 1 a 8,

"1 2 3 4 5 6 7 8" a mediana estará entre 4 e 5.

O primeiro quartil fará o papel de mediana da primeira parte:

1 2 3 4, isto é, estará entre 2 e 3.

Do mesmo modo, o terceiro quartil funcionará como mediana da segunda parte:

5 6 7 8, isto é, estará entre 6 e 7.

De outro modo:

1 2 3 4 5 6 7 8 (n=8)

1 2 | 3 4 | 5 6 | 7 8

(1º quartil) Q_1=2.5 Mediana: 4.5 (3º quartil) Q_3=6.5

1º quartil (Q_1) é um valor que divide a amostra (depois de organizados os dados por ordem crescente) em duas partes, de tal modo que, no mínimo, 25% das observações sejam inferiores ou iguais a esse valor e, no máximo, 75% das observações sejam superiores a esse valor.

2º quartil (**Q_2** ou **med.**) corresponde à mediana.

3º quartil (**Q3**) é o valor que divide a amostra (ordenada por ordem crescente) em duas partes, de tal modo que, no mínimo, 75% das observações sejam inferiores ou iguais a esse valor e, no máximo, 25% das observações sejam superiores a esse valor.

Exemplo

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 (n=15)

1 2 3 **4** 5 6 7 **8** 9 10 11 **12** 13 14 15

(1º quartil) Q_1=4 Mediana: 8 (3º quartil) Q_3=12

Nota: Por vezes, as definições de mediana e de quartis podem variar ligeiramente de livro para livro. Aqui o importante é que, estamos usando um procedimento semelhante ao que é feito nas calculadoras gráficas, para evitar incoerências, uma vez que podemos obter estes valores através da tecnologia permitida e disponível.

Exemplo

Numa empresa fez-se um estudo sobre o tempo que os seus empregados demoravam no percurso de casa para o emprego e obtiveram-se os resultados que constam na tabela seguinte:

Identifique a classe que contém o primeiro quartil e a classe que contém a mediana.

Tempo (Minutos)	Número de empregados
[10, 20[	12
[20, 30[	6
[30, 40[	6
[40, 50[	7
[50, 60[	6
[60, 70[	3

Resolução:

Se construirmos a tabela de frequências acumuladas:

Tempo (Minutos)	Número de empregados	Freq. Abs. Acumul.(F_i)	Freq. Rel. Acumul.Fr_i (%)
[10, 20[	12	12	30
[20, 30[	6	18	45
[30, 40[	6	24	60
[40, 50[	7	31	77.5
[50, 60[	6	37	92.5
[60, 70[	3	40	100

Podemos constatar que

A classe que contém o primeiro quartil é [10, 20[, pois a sua frequência acumulada é 30%, que ultrapassa os 25% associados ao primeiro quartil.

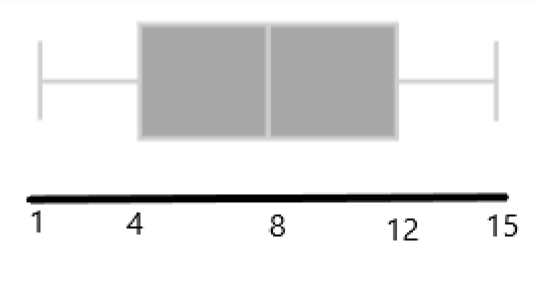
A classe que contém a mediana é [30, 40[, cuja frequência acumulada é 60%, e é a primeira classe que ultrapassa os 50%.

Diagrama de extremos e quartis

Retomando um dos exemplos anteriores:

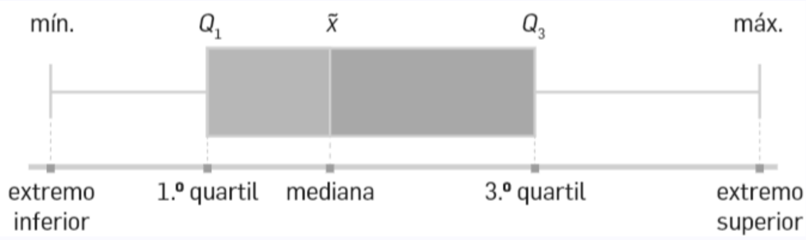
1 2 3 **4** 5 6 7 **8** 9 10 11 **12** 13 14 15

existe uma forma de representar os dados com os extremos, a mediana e os quartis, colocando uma "caixa" na zona:

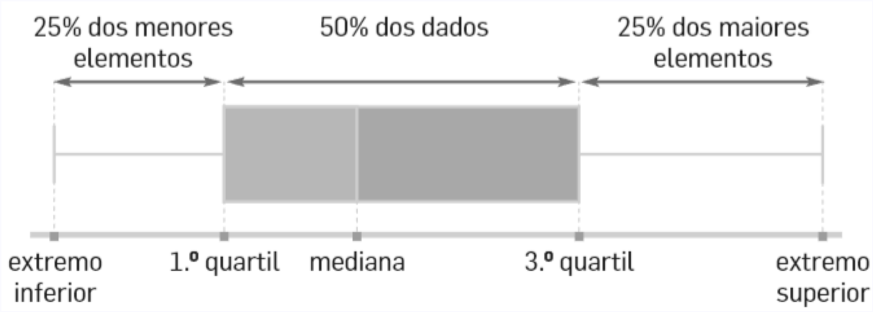


Este chama-se **diagrama de extremos e quartis** e é construído da seguinte forma:

1º) Desenhamos uma linha onde se marcam os extremos e os quartis.
2º) Construímos dois retângulos contíguos — o primeiro entre Q_1 e a mediana e o segundo entre esta e Q_3.
3º) Construímos um segmento de reta entre o mínimo e Q_1 e outro entre Q_3 o máximo.



Percentagens:



Nota: podemos obter o **diagrama de extremos e quartis** usando a **calculadora gráfica**.

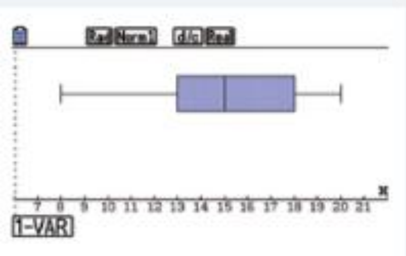
Sugestão: Introduza os números numa lista, na secção "estatística" e peça o gráfico, escolhendo este tipo de gráfico.

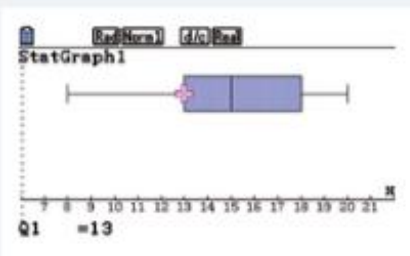
Calculadora gráfica- Diagrama de Extremos e quartis.

Exemplo:
Coloque os valores 8; 12; 14; 15; 15; 17; 17; 19; 20 na "lista 1" da calculadora gráfica. (Stat).

Casio:

Depois de lançar os dados, GRAPH(F1);
SET(F6) Graph Type: **MedBox**
Xlist: **list 1**
Frequency: **1**
EXE
GRAPH 1





Texas

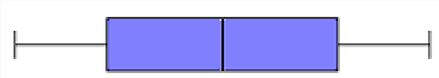
Depois de lançar os dados em L1, STAT PLOT (2nd Y=)
On   e escolha em Type o diagrama de extremos e quartis.
Xlist: L1
Freq: 1
GRAPH

Simetria/ enviesamento.

Dados simétricos: Os dados estão distribuídos de forma simétrica.

Exemplo:
Consideremos os dados 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

Min=1   Q1=3   med=5.5   Q3=8   Máx=10



Simétrico.

Enviesamento para a direita: Os dados estão mais dispersos à direita de Q2 e mais concentrados à esquerda de Q2 .

Exemplo:
Consideremos os dados 1, 2, 3, 4, 5, 6, 10, 14, 18, 22.

Min=1   Q1=3   med=5.5   Q3=14   Máx=22

Enviesado para a direita.

Enviesamento para a esquerda: Os dados estão mais dispersos à esquerda de Q2 e mais concentrados à direita de Q2 .

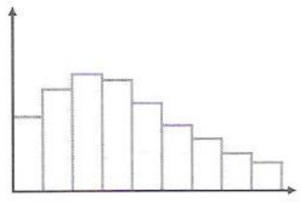
Exemplo:
Consideremos os dados 1, 5, 9, 13, 17, 18, 19, 20, 21, 22.

Min=1   Q1=9   med=17.5   Q3=20   Máx=22

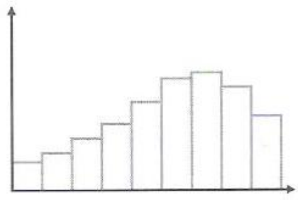
Enviesado para a esquerda.

Nota: Por estranho que possa parecer, em alguns livros o enviesamento é definido ao contrário do que aqui está apresentado. Nesta definição, consideramos “enviesado” o lado mais “comprido”, isto é, onde os dados estão mais dispersos.

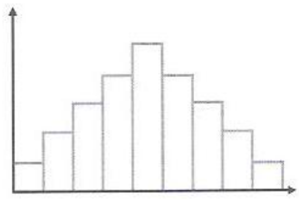
Exemplo:
Numa empresa com três filiais, em três cidades distintas, registaram-se os tempos que os trabalhadores gastam de casa ao local de trabalho e obtiveram-se os seguintes histogramas:



Cidade A



Cidade B



Cidade C

Sabendo que os diagramas de extremos e quartis seguintes também representam as mesmas distribuições, faça corresponder a cada uma deles uma cidade. Justifique, convenientemente, a sua escolha.

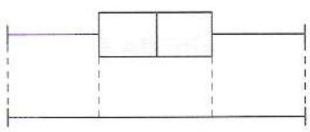


Diagrama 1

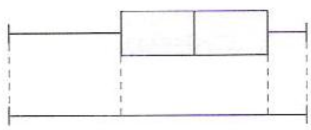


Diagrama 2

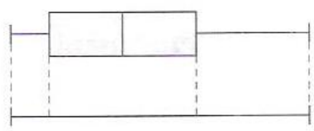


Diagrama 3

Resposta:

Na cidade A, os dados estão mais concentrados mais à esquerda, logo corresponde ao diagrama 3.
Na cidade B, os dados estão mais concentrados à direita, logo corresponde ao diagrama 2.
Na cidade C, os dados estão distribuídos de forma simétrica em relação ao centro, pelo que corresponderá ao diagrama 1.

Percentis.

Os **percentis** dividem um conjunto de dados ordenados em cem partes iguais.

O percentil P_k de <u>ordem k</u> tem, no mínimo, k% dos dados da amostra inferiores ou iguais a ele e, no máximo, (100 –k)% dos dados da amostra superiores a ele.

Por ***exemplo***, 20% dos elementos da amostra são inferiores ou iguais ao percentil 20 (**P_{20}**).

Exemplo:

Se considerarmos os números inteiros de 1 a 20:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,12, 13, 14, 15, 16, 17, 18, 19, 20.

Como 1/20=0.05 ou 5%, podemos dizer que 1 é inferior ou igual a 5% dos dados.
Assim. P_5=1.
Do mesmo modo, podemos escrever:
P_{10}=2; P_{15}=3; P_{20}=4 P_{25}=5 (apesar de Q_1 ser 5.5).

P_{30}=6; P_{35}=7; P_{40}=8; P_{45}=9; P_{50}=10 (apesar de Q_2, ou seja, a mediana ser 10.5).
P_{100}=20 (por ser o máximo).
P_{95}=19; P_{90}=18; P_{85}=17;
P_{80}=16; P_{75}=15; (apesar de Q_3 ser 15.5).

Exemplo

Numa empresa fez-se um estudo sobre o tempo que os seus empregados demoravam no percurso de casa para o emprego e obtiveram-se os resultados que constam na tabela seguinte:

Identifique as classes que contêm o 46º percentil e o 88º percentil.

Tempo (Minutos)	Número de empregados
[10, 20[	12
[20, 30[	6
[30, 40[	6
[40, 50[	7
[50, 60[	6
[60, 70[	3

Resolução:

Se construirmos a tabela de frequências acumuladas:

Tempo (Minutos)	Número de empregados	Freq. Abs. Acumul.(F_i)	Freq. Rel. Acumul.Fr_i (%)
[10, 20[	12	12	30
[20, 30[	6	18	45
[30, 40[	6	24	60
[40, 50[	7	31	77.5
[50, 60[	6	37	92.5
[60, 70[	3	40	100

Podemos constatar que:
A primeira classe cuja frequência relativa acumulada ultrapassa 46% é a terceira, logo

P_{46} está na classe [30, 40[

Do mesmo modo, P_{88} está na classe [50, 60[.

Medidas de dispersão.

Nota: As medidas que se seguem servem para analisar se os números estão muito próximo (concentrados) entre si, tais como: {10; 10; 10; 11; 11; 11; 12; 12} ou se estão muito afastados (ou dispersos), tais como {1; 5; 16; 22; 34; 55; 61} .

Amplitude.

Amplitude (h) é a diferença entre o valor máximo e o valor mínimo da, ou seja:

$$\mathbf{h = x_{max} - x_{min}}$$

Exemplo

Considere os dados: 1; 5; 16; 22; 34; 50; 51

A amplitude é h=51-1=50

Nota: Quando os dados estão agrupados em classes, a amplitude é a diferença entre o limite superior da última classe e o limite inferior da primeira classe.

Exemplo

No conjunto de dados agrupados em intervalos de classe:

Tempo (Minutos)
[**10**, 20[
[20, 30[
[30, 40[
[40, 50[
[50, 60[
[60, **70**[

A amplitude é 70-10=60

Nota: Como a amplitude é muito sensível a valores discrepantes, isto é, valores demasiado altos ou demasiado baixos em relação aos restantes elementos da amostra, por vezes usamos a amplitude interquartil.

Amplitude interquartil.

Amplitude interquartil (A_q) é a diferença entre o 3.º quartil e o 1.º quartil:

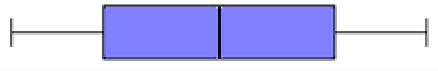
$$\mathbf{A_q = Q_3 - Q_1}$$

Exemplo:

Considere os dados

1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

Min=1 **Q1=3** med=5.5 **Q3=8** Máx=10



A amplitude interquartil é dada por:

$$\mathbf{A_q = Q_3 - Q_1 = 8-3=5}$$

Desvio médio. *(é pouco utilizado!...)*

Desvio médio (d_m) é a média dos desvios entre as observações e a média, ou seja:

$$d_m = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{n}$$

em que x_1, x_2, ...x_n são as observações, $\bar{x}$ é a média e n é o número de observações.

Exemplo:
Consideremos os dados de uma amostra: 6; 7; 8.

1º calculemos a média amostral.

$$\bar{x} = (6+7+8)/3 = 21/3 = 7$$

Desvio médio

$$d_m = \frac{|6 - \mathbf{7}| + |7 - \mathbf{7}| + |8 - \mathbf{7}|}{3} = \frac{1 + 0 + 1}{3} = \frac{2}{3}$$

Variância (amostral).

Variância (amostral) (s^2) é o quociente entre a soma dos quadrados dos desvios em relação à média e o número de observações menos 1, ou seja:

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}$$

Desvio padrão (amostral) (s) é a raiz quadrada da **Variância.**

Exemplo 1: Variância amostral(**S^2**) e desvio padrão amostral**(S)**.

Consideremos os dados de uma amostra: 4; 5; 6.

1º calculemos a média amostral.

$$\bar{x} = (4+5+6)/3 = 5$$

Variância amostral:

$$S^2 = \frac{(4 - \mathbf{5})^2 + (5 - \mathbf{5})^2 + (6 - \mathbf{5})^2}{3 - 1} = \frac{1 + 0 + 1}{2} = \frac{2}{2} = 1$$

O desvio padrão é a raiz quadrada da variância:

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

Neste caso, $S = \sqrt{1} = 1$

Exemplo Desvio padrão amostral.
Consideremos os dados da <u>amostra</u>: 2; 3; 7.
1º calculemos a média da amostra:

$$\bar{x}=(2+3+7)/3 = 4$$

Variância:

$$S^2 = \frac{(2-\mathbf{4})^2+(3-\mathbf{4})^2+(7-\mathbf{4})^2}{3-1} = \frac{4+1+9}{2} = \frac{14}{2} = 7$$

O desvio padrão é a raiz quadrada da variância:

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

Neste caso, $S = \sqrt{7} \approx 2.646$

Nota: se os **dados** forem discretos e estiverem **agrupados**, usamos o mesmo raciocínio, agrupando os respetivos valores:

Variância para dados agrupados:

$$s^2 \approx \frac{f_1\,(m_1-\bar{x})^2 + f_2\,(m_2-\bar{x})^2 + \ldots + f_M\,(m_M-\bar{x})^2}{n-1}$$

Exemplo Desvio padrão amostral.
Considere os dados agrupados:

xi	*fi*
2	*15*
3	*10*
6	*5*
Total:	***30***

1º calculemos a média da amostra:

$$\bar{x} = \frac{15\times 2 + 10\times 3 + 5\times 6}{30} = 3$$

Variância:

$$S^2 = \frac{15\times(2-\mathbf{3})^2 + 10\times(3-\mathbf{3})^2 + 5\times(6-\mathbf{3})^2 +}{30-1} = \frac{15+0+45}{29} = \frac{60}{29} \approx 2.069$$

O desvio padrão é a raiz quadrada da variância:

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

Neste caso, $S = \sqrt{2.069} \approx 1.438$

Sugestões para a calculadora gráfica:

Casio:

(CG-Casio)Exemplo 1: dados simples 14; 17; 15; 17; 15; 15; 18:
Colocamos os dados numa lista da secção estatística(Lista 1).

Menu/Estatística.
Introduza os valores na lista 1.
Calc(F2) SET(F6)/ 1Var Xlist: **List 1** 1Var **Freq: 1**. (EXE). 1-VAR(F1)
Aparece a média $\overline{x}$ (aprox. 15.86)
e também outras grandezas. (use a seta para descer.)

(CG-Casio)Exemplo 2: dados Agrupados.
"Lista1": 0; 1; 2; 3; 4. "Lista 2": 3; 5; 7; 10; 5.
Colocamos os dados nas listas da secção estatística(Lista 1, Lista 2).

Casio: Menu/Estatística.
Introduzir os valores na listas 1 e Lista2.
Calc(F2) SET(F6)/ 1Var Xlist: **List 1** 1Var **Freq: List 2**. (EXE). 1-VAR(F1)
Aparece a média $\overline{x}$ (2.3)
e também outras grandezas (use a seta para descer.)
A mediana é representada por "med".

Texas

(CG-Texas)Exemplo 1: dados simples 14; 17; 15; 17; 15; 15; 18:
Colocamos os dados numa lista da secção estatística(Lista 1).

Tecla STAT / edit
Introduza os valores na lista "L1".
STAT/ CÁlCULO(ou CALC)/ 1-Var Stats/
1-Var Stats L1 (ENTER)
Aparece a média $\overline{x}$ (aprox. 15.86)
e também outras grandezas (use a seta para descer.)

(CG-Texas)Exemplo 2: dados Agrupados.
"Lista1": 0; 1; 2; 3; 4. "Lista 2": 3; 5; 7; 10; 5.
Colocamos os dados nas listas da secção estatística(STAT) (Edit) (Lista 1, Lista 2).

Introduza os valores nas listas "L1" e "L2".

STAT/ CÁlCULO(ou CALC)/ 1-Var Stats/
1-Var Stats L1, L2 (ENTER) ou preencha os espaços(…)
Aparece a média $\overline{x}$ (2.3)
e também outras grandezas (use a seta para descer.)
A mediana é representada por "med".

TI-Nspire… procure em:
Primeira página de:
https://pedronoia.net/nspire.pdf

NumWork

Procure em:
https://www.numworks.com/pt/professores/tutoriais/estatistica/

Notas importantes:

Nota 1) Tal como a média, o desvio padrão também é uma medida **pouco resistente** pois é influenciado por valores discrepantes. No que diz respeito a **valores discrepantes** e, comparando as três medidas de dispersão estudadas (amplitude, amplitude interquartil e desvio padrão), podemos afirmar que a amplitude interquartil é a mais robusta(ou mais resistente) das três.

Nota 2) O desvio padrão toma sempre valores maiores ou iguais a zero. Se todos os dados forem iguais, o desvio padrão é zero.

Nota 3) Para amostras com a mesma média, quanto **mais dispersos** forem os dados, **maio**r é o valor do **desvio padrão**. ***Sugestão:*** experimente com a CG para as populações: 9; 10; 11 e 5; 10; 15. Obtenha a média e o desvio padrão para cada uma.

Propriedades do desvio padrão:

Propriedade1: Se adicionarmos a cada um dos valores uma constante k, o novo desvio padrão não se altera.

Exemplo Desvio padrão amostral.
Consideremos os dados da amostra: 2; 3; 7.
O desvio padrão é $S \approx 2.646$ (pode verificar com a c.g)

Se somar 4 unidades a cada um dos dados, 2+4=6; 3+4=7; 7+4=11, obtemos 6; 7; 11.

Calculando o seu desvio padrão, obtemos $S \approx 2.646$, como seria de esperar.

Nota: esta propriedade é intuitiva, pois, ao adicionarmos o mesmo número a todos os valores, as distâncias entre os números mantêm-se.

Propriedade2: Se multiplicarmos cada um dos valores por uma constante k, o novo desvio padrão será igual ao desvio padrão original, multiplicado por k.

Exemplo Desvio padrão amostral.
Consideremos os dados da amostra: 2; 3; 7.
O desvio padrão é $S \approx 2.646$ (pode verificar com a c.g)

Se multiplicar todos os dados por 5, obteremos 2×5=10; 3×5=15; 7×5=35, obtemos

10; 15; 35.

Calculando o seu desvio padrão, obtemos $S \approx 13.23 = 5 \times 2.646$, como seria de esperar, o desvio padrão fica 5 vezes maior.

Nota: Esta propriedade também é intuitiva, pois, ao multiplicarmos o mesmo número a todos os valores, as distâncias entre os números aumentam nessa mesma proporção.

Intervalo:

$$]\bar{x} - s,\ \bar{x} + s[$$

Nota: Poderá ser importante saber a elementos de uma amostra, que pertence ao intervalo

$$]\bar{x} - s,\ \bar{x} + s[$$

Exemplo:

Consideremos os dados de uma amostra:

22 23 24 24 25 25 25 26 26 27 27 27 28 29 31 31 31 32 33 34

Se colocarmos da calculadora gráfica, podemos constatar que

$$\bar{x} = 27.5 \text{ e } s \approx 3.49$$

Assim,

$$]\ \bar{x} - s; \bar{x} + s[\ =]27.5 - 3.49; 27.5 + 3.49[\ =]24.01; 30.99[.$$

Procuremos os elementos que pertencem a este intervalo:

22 23 **24 24 25 25 25 26 26 27 27 27 28 29** 31 31 31 32 33 34

São 12 elementos num total de 20. Como 12/20 = 0.6, isto corresponde à percentagem de 60%.

Conclusão: 60% dos elementos pertencem ao intervalo $]\ \bar{x} - s; \bar{x} + s[$.

Pensamento computacional- Python.

Nota: dentro do Python, existe um módulo que permite trabalhar com termos técnicos da Estatística. Para trabalhar com esse módulo é preciso importar “statistics”, isto é,

```
import statistics
```

para poder utilizar os termos técnicos da estatística tais como a média:

```
statistics.mean
```

Ou desvio padrão:

```
statistics.stdev
```

Sugestão: consulte vídeos tutoriais sobre python:
https://pedronoia.pt/python/10m26a.htm
ou
https://pedronoia.pt/python/10m26b.htm
Ou
https://pedronoia.pt/python/pyl6.htm