3.2-Medidas de localização e de dispersão.

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

Medidas de localização.

Média.

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

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

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 μ.

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

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

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

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

Recorde que $fr_1 = f_1/n$.

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.

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.

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** .

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

Moda

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

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

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

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

Se os dados estiverem agrupados em classes indicaremos a **classe modal.**
Classe modal é a classe à qual corresponde a 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 valor central.

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

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 $\boldsymbol{k = \frac{n+1}{2}}$

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

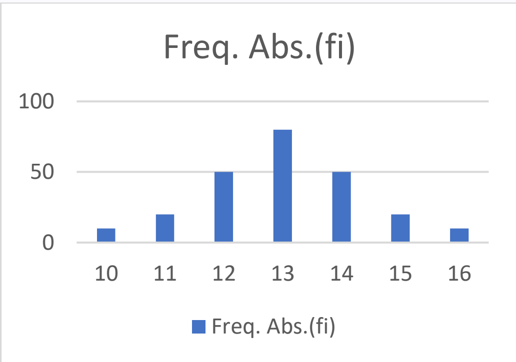
<u>Nota:</u> Se os **dados** estiverem **agrupados** em classes na forma de **intervalos**, determinaremos a classe mediana.

Classe mediana é a classe à qual pertence a mediana.

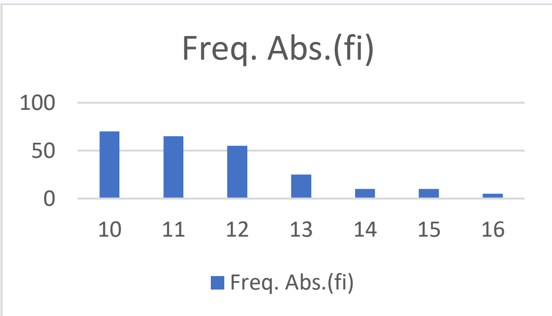
Vantagens, desvantagens e limitações das medidas de tendência central

A média é muito **sensível** a valores discrepantes. A **mediana** é uma medida **resistente**.

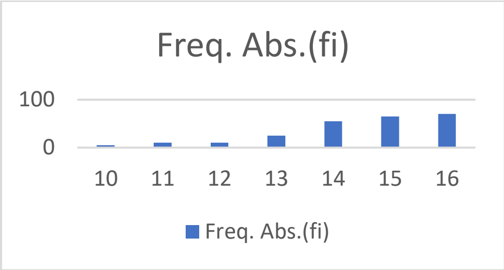
Simetria da distribuição



A média igual à mediana
Neste caso, dizemos que a distribuição é **simétrica.**



A média é maior do que a mediana.
Neste caso, dizemos que a distribuição é **assimétrica positiva, ou enviesada** para a direita.



A média é menor que a mediana.

Neste caso, dizemos que a distribuição é **assimétrica negativa, ou enviesada** para a esquerda.

Outras medidas de localização (quartis e percentis)

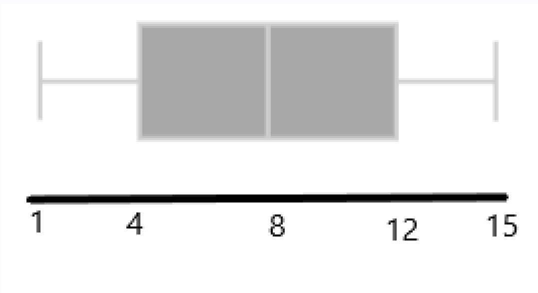
Quartis.

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.

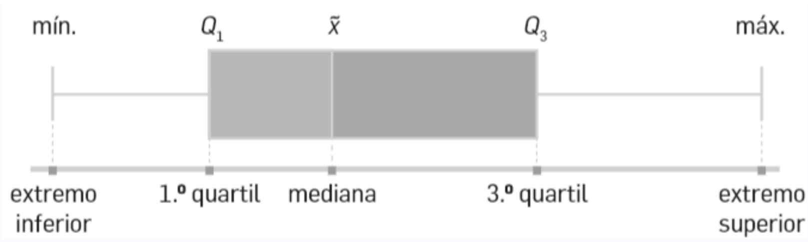
3º quartil (**Q3**) é o valor que divide a amostra (ordenada por ordem crescente) em duas partes, de tal modo que, no mínimo, 75% das observações sejam inferiores ou iguais a esse valor e, no máximo, 25% das observações sejam superiores a esse valor.

Diagrama de extremos e quartis

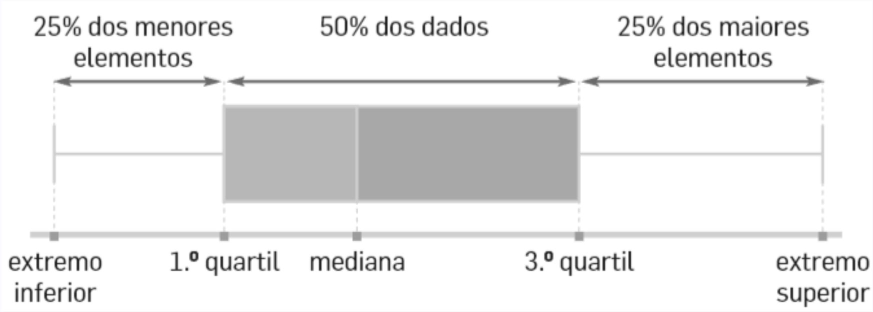


Este chama-se **diagrama de extremos e quartis** e é construído da seguinte forma:

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:



Simetria/ enviesamento.

Dados simétricos: Os dados estão distribuídos de forma simétrica.

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

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

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.

Medidas de dispersão.

Amplitude.

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

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

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.

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:

$$A_q = Q_3 - Q_1$$

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.**

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

Variância para dados agrupados:

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

Propriedades do desvio padrão:

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

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