STATISTICS

Module 2 *of*

MATHEMATICS AND STATISTICS

Marina Santacroce

Department of Decision Sciences, Bocconi

L2: Descriptive statistics: central tendency and variability measures

Outline

Introduction

Frequency tables and graphs give a first summary of the data set.

In this lecture we present some *summary statistics*, i.e. numerical quantities whose value is determined by the data, that are

- measures of *tendency* (or *location)*
- measures of *variability* (or *spread*)

Among the first we will study the mean, median and mode, which are *representative* values describing the "*center*" of the data set.

Mean

Let $x_1, \ldots, x_n$ be the n observations and $v_1, \ldots, v_k$ and $k \leq n$, the distinct values observed in this data set.

The mean is the arithmetic *mean* of the values

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + \ldots + x_n).$$

Denoting the *frequency distribution* by

$$X = \begin{Bmatrix} v_1 & \ldots & v_k \\ f_1 & \ldots & f_k \end{Bmatrix}$$

then, the mean is given by

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{k} v_i f_i \qquad \text{(absolute frequencies)}$$

or equivalently

$$\bar{x} = \sum_{i=1}^{k} v_i \frac{f_i}{n} \qquad \text{(relative frequencies)}$$

in these two cases the mean is a *weighted* average: any distinct value v_i is weighted with the corresponding frequency f_i (relative frequency f_i/n).

Example

$X = \{$ "maturità"[1] grade$\}$ 10 students

$$\{x_i\}_{i=1,\dots,10} = \{90, 100, 88, 90, 60, 60, 60, 100, 60, 90\}$$

$$\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = \frac{1}{10}(90 + 100 + 88 + 90 + 60 + 60 + 60 + 100 + 60 + 90) = 79.8$$

or, alternatively,

$$X = \begin{Bmatrix} 60 & 88 & 90 & 100 \\ 4 & 1 & 3 & 2 \end{Bmatrix}$$

and we compute the mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} v_i f_i = \frac{1}{10}(60 \cdot 4 + 88 \cdot 1 + 90 \cdot 3 + 100 \cdot 2) = 79.8$$

Yet another way is

$$X = \begin{Bmatrix} 60 & 88 & 90 & 100 \\ 0.4 & 0.1 & 0.3 & 0.2 \end{Bmatrix}$$

$$\bar{x} = \sum_{i=1}^{k} v_i \frac{f_i}{n} = 60 \cdot 0.4 + 88 \cdot 0.1 + 90 \cdot 0.3 + 100 \cdot 0.2 = 79.8$$

[1] Italian Baccalaureate

Mean for grouped data

If we have data grouped in classes, we can only find an approximate value of the mean.

Since, for example, $X = \begin{cases} (0,5] & (5,10] & \dots \\ 2 & 3 & \dots \end{cases}$

2 data values belong to $(0,5]$, but the exact value is unknown!

Thus given the class frequency distribution

$$X = \begin{cases} (x_0, x_1] & (x_1, x_2] & \dots & (x_{k-1}, x_k] \\ f_1 & f_2 & \dots & f_k \end{cases}$$

- the central value of each class is taken as representative $\bar{x}_i = \frac{x_{i-1}+x_i}{2}, \qquad i = 1, \dots, k$
- the approximate mean is computed as

$$\bar{x} \cong \frac{1}{n} \sum_{i=1}^{k} \bar{x}_i f_i = \sum_{i=1}^{k} \bar{x}_i \frac{f_i}{n}$$

- the approximation depends on the choice of the classes and on the representativeness of the central values.

Example

Let us consider the data

$$X = \begin{cases} 18 & 19 & 20 & 21 & 22 & 23 \\ 15 & 10 & 8 & 8 & 5 & 4 \end{cases} \quad \Rightarrow \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} v_i f_i = 19.8$$

and now *grouped in 2 classes*

$$X = \begin{cases} [18, 20] & (20, 23] \\ 33 & 17 \end{cases}$$

- we compute the central values

$$\bar{x}_1 = \frac{18+20}{2} = 19 \qquad \text{and} \qquad \bar{x}_2 = \frac{20+23}{2} = 21.5$$

- and the mean with the central values

$\Rightarrow \bar{x} \cong \frac{1}{n}\sum_{i=1}^{k} \bar{x}_i f_i = \frac{1}{50}(19 \cdot 33 + 21.5 \cdot 17) = 19.85$

The mean is an approximation of the true value because of the loss of information due to the data grouping.

Linear transformation

The computation of the mean can often be simplified by considering a *linear transformation of* X, that is a $Y = a + bX$, for two suitable constants $a, b \in \mathbb{R}$.

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n}(a + bx_i) = \frac{1}{n}\sum_{i=1}^{n} a + b\frac{1}{n}\sum_{i=1}^{n} x_i = a + b\bar{x}.$$

Or, analogously, by considering the frequency distribution of X

$$X = \begin{Bmatrix} v_1 & \dots & v_k \\ f_1 & \dots & f_k \end{Bmatrix}$$

and the related distribution of Y

$$Y = \begin{Bmatrix} a + bv_1 & \dots & a + bv_k \\ f_1 & \dots & f_k \end{Bmatrix} = \begin{Bmatrix} u_1 & \dots & u_k \\ f_1 & \dots & f_k \end{Bmatrix}$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{k} u_i f_i = \frac{1}{n}\sum_{i=1}^{k}(a + bv_i)f_i = \frac{1}{n}\sum_{i=1}^{k} af_i + b\frac{1}{n}\sum_{i=1}^{k} v_i f_i = a + b\bar{x}.$$

Example

Example: Winning scores in the US Masters golf tournament from 2004 to 2013 are

$$280, 278, 272, 276, 281, 279, 276, 281, 289, 280$$

$\Rightarrow$ Find the mean.

Solution:

We can consider the transformation $y = x - 280$ and thus the data set of y_i

$$0, -2, -8, -4, 1, -1, -4, 1, 9, 0$$

and compute $\bar{y} = -8/10 \qquad \Rightarrow \bar{x} = -0.8 + 280 = 279.2.$

Or, alternatively, use the frequency distribution,

$$X = \begin{cases} 272 & 276 & 278 & 279 & 280 & 281 & 289 \\ 1 & 2 & 1 & 1 & 2 & 2 & 1 \end{cases}$$

- $$Y = \begin{cases} -8 & -4 & -2 & -1 & 0 & 1 & 9 \\ 1 & 2 & 1 & 1 & 2 & 2 & 1 \end{cases}$$

$\Rightarrow \qquad \bar{y} = -0.8 \qquad$ and $\qquad \bar{x} = -0.8 + 280$

Other properties

- $\sum_{i=1}^{n} \bar{x} = \sum_{i=1}^{n} x_i$.
- The sum of the deviations of the data from the mean is zero, that is

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

where *deviation* means the difference between the observed value from the mean.

- The mean minimizes the sum of the *squared deviations* of the data from an arbitrary constant, i.e.

$$\min_{a \in \mathbb{R}} \sum_{i=1}^{n} (x_i - a)^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

In this respect we can say that the mean is a good representative of the data set since it is close to the observations. More precisely, if we look for the value which minimizes its distance from the observed data and the distance is chosen to be

$$d(x_i, a) = (x_i - a)^2,$$

the mean $\bar{x}$ is the value for which the minimum is attained.

Median

The *median* is the *center* of the *frequency distribution*, i.e it is the *value which splits the ordered values into two groups of the same size.*

How to compute it?

If the data set is $x_1, \ldots, x_n$

1) order the values of the data set from smallest to largest

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$$

2) compute the median:

- if n is odd, $\frac{n+1}{2}$ is an integer and

$$\text{med} = x_{(\frac{n+1}{2})}$$

- if n is even, the median is the average of the two central values

$$\text{med} = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}.$$

Examples:

Ex.1:

- $\{18, 19, 21, 24, 29\} \quad n = 5$
- $\frac{n+1}{2} = \frac{5+1}{2} = 3$
- $\text{med} = x_{(\frac{n+1}{2})} = x_{(3)} = 21$

Ex.2:

- $\{18, 19, 21, 24, 29, 30\} \quad n = 6$
- $\text{med} = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} = \frac{x_{(3)} + x_{(4)}}{2} = \frac{21 + 24}{2} = 22.5$

Median

If the data set is represented by a relative frequency distribution

1) we compute the *cumulated frequencies* $F_i = f_1/n + \ldots + f_i/n$

v_i	f_i/n	F_i
v_1	f_1/n	$F_1 = f_1/n$
v_2	f_2/n	$F_2 = f_1/n + f_2/n$
$\vdots$	$\vdots$	$\vdots$
v_k	f_k/n	$F_k = f_1/n + \ldots f_k/n = 1$

2) as before we have 2 possible cases:

   i) if there is not a v_i such that $F_i = 0.5$ in this case the median is

$$\text{if } F_i < 0.5 \text{ and } F_{i+1} > 0.5 \quad \Rightarrow \quad \text{med} = v_{i+1}$$

   ii) there exists v_i such that $F_i = 0.5$ in this case the median is the average between this value and the following one, i.e.

$$\text{if } F_i = 0.5 \quad \Rightarrow \quad \text{med} = \frac{v_i + v_{i+1}}{2}$$

Examples

Ex.1:

v_i	f_i/n	F_i
0	0.35	0.35
1	0.25	0.6
2	0.2	0.8
3	0.15	0.95
4	0.03	0.98
5	0.02	1

$F_1 < 0.5$ and $F_2 > 0.5$, therefore

$$\text{med} = v_2 = 1.$$

Ex.2:

v_i	f_i/n	F_i
0	0.35	0.35
1	0.15	0.5
2	0.3	0.8
3	0.15	0.95
4	0.03	0.98
5	0.02	1

v_2 is such that $F_2 = 0.5$, therefore

$$\text{med} = \frac{v_2 + v_3}{2} = \frac{1+2}{2} = 1.5$$

Properties of the median

- If Y is the linear transformation $Y = a + bX$, then

$$\text{med}(Y) = a + b\,\text{med}(X).$$

- The median is the value which minimizes the sum of absolute values of the deviations of the data from a constant, that is

$$\min_{a \in \mathbb{R}} \sum_{i=1}^{n} |x_i - a| = \sum_{i=1}^{n} |x_i - \text{med}|$$

Note that the mean makes use of all the data values, thus it is affected by extreme values.
In contrast, the median is *robust* with respect to *outliers*.

We have to be careful when the values of the mean and median are quite different!

Mode: definition and properties

The *mode* of a data set

- is the *value* that occurs with *highest frequency* or
- the *class* with the *highest density*, if the data are split in classes.

Example:

$$X = \left\{ \begin{array}{ccccc} 1 & 2 & 3 & \textcircled{4} & 5 \\ 0.2 & 0.1 & 0.1 & 0.4 & 0.2 \end{array} \right.$$

The mode is 4.

NOTE:

- the *mode* is a value and not a frequency!
- it could be *not unique*.

- If Y is the linear transformation $Y = a + bX$, then

$$\text{mode}(Y) = a + b\,\text{mode}(X).$$

Examples

Ex. 1:

$$X = \begin{cases} \textcircled{1} & 2 & \textcircled{3} & 4 & 5 \\ 0.4 & 0.1 & 0.4 & 0.05 & 0.05 \end{cases}$$

The mode is not unique $\Rightarrow$ 1 and 3.

Ex.2:

	$\frac{f_i}{n}$	a_i	h_i
(0,10]	0.35	10	0.035
(10,30]	0.6	20	0.03
(30,35]	0.05	5	0.01

Recall $a_i = x_i - x_{i-1}$ is the length of the class i and $h_i = \frac{f_i/n}{a_i}$ is its density.

The mode here is the class $(0, 10]$.

Introduction

Mean, median and mode are values which describe the central tendency of the data set, but they do not give information about the *variability* (*dispersion* or *spread*) of the data.

Es.:

- $X = \{$ sq.m. flats from group 1$\}$, $n = 50$
- $Y = \{$sq.m. flats from group 2$\}$, $n = 50$

v_i	$\frac{f_i}{n}$	F_i
110	0.1	0.1
130	0.2	0.3
150	0.4	0.7
170	0.2	0.9
190	0.1	1

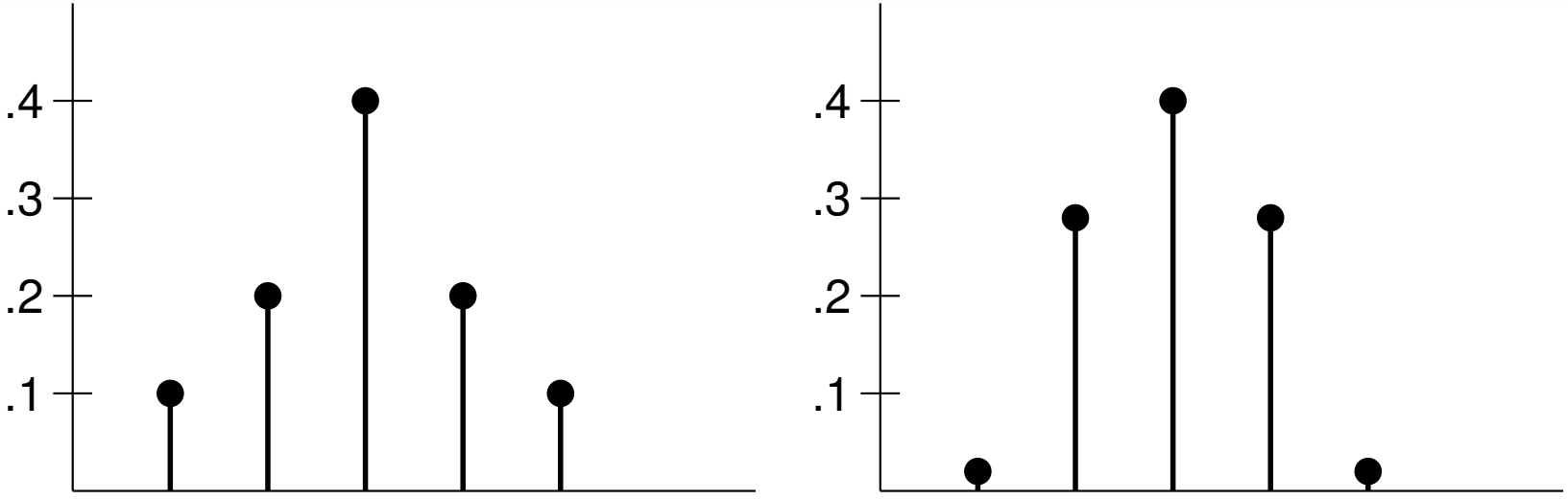
u_i	$\frac{f_i}{n}$	F_i
110	0.02	0.02
130	0.28	0.3
150	0.4	0.7
170	0.28	0.98
190	0.02	1

$$\bar{x} = \sum_{i=1}^{5} v_i \frac{f_i}{n} = 150 \qquad \bar{y} = \sum_{i=1}^{5} u_i \frac{f_i}{n} = 150$$

$$\text{mode} = 150 \qquad \text{mode} = 150$$

$$\text{med} = 150 \qquad \text{med} = 150$$

Mean, median and mode are the same in the two data sets.



- Both frequency distributions are symmetric with respect to 150, but the one of the second group is more concentrated around it.

We want to introduce a statistics which describes the dispersion.

The less is the dispersion of the data set, the more is the central value representative of the data.
The extreme case will be when all observations coincide and

$$\text{mean} = \text{mode} = \text{median} = \text{observation}$$

Variance and standard deviation

The most important measure of the variability is the variance which measures the *dispersion around the mean*.

Given $\{x_1, \ldots, x_n\}$ or $X = \begin{cases} v_1 & \ldots & v_k \\ f_1 & \ldots & f_k \end{cases}$

the *variance*[2] is the *mean of the squared deviations of the data from the mean*

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad \text{or, equivalently} \quad s^2 = \sum_{i=1}^{k}(v_i - \bar{x})^2 \frac{f_i}{n}.$$

- The square root of the variance called *standard deviation* is also used.

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\sum_{i=1}^{k}(v_i - \bar{x})^2 \frac{f_i}{n}}.$$

- the *standard deviation* is measured in the same units as the data.

[2] the (sample) variance is defined by $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Properties of the variance

- $s^2 \geq 0$
- $s^2 = 0$ if and only if all data values coincide.
  $s^2 = 0$ (or $s = 0$) $\Leftrightarrow$ *no variability in the data set*.
- Given a data set with mean $\bar{x}$ and variance s_X^2 and $Y = a + bX$, the variance of Y is
$$s_Y^2 = b^2 s_X^2$$

The following formula is very useful from a computational point of view

$$s^2 = \bar{x}_{(2)} - \bar{x}^2$$

where

$$\bar{x}_{(2)} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{k} v_i^2 \frac{f_i}{n}.$$

Coefficient of variation

To compare the variability of two very different data sets it is often useful to resort to the coefficient of variation or relative standard deviation. If, for example

- the values of the data are measured in different units or
- they are measured in the same units but they are of different size orders.

The *coefficient of variation* (CV) is

$$\text{CV} = \frac{s}{|\bar{x}|}, \quad \text{if } \bar{x} \neq 0$$

Variance for data grouped in classes

Given the class frequency distribution

$$X = \begin{cases} (x_0, x_1] & (x_1, x_2] & \dots & (x_{k-1}, x_k] \\ f_1 & f_2 & \dots & f_k \end{cases}$$

we compute the class central values $\bar{x}_i = \frac{x_{i-1}+x_i}{2}, \qquad i = 1, \dots, k$.
Then, as we did for the mean, we calculate the variance using the central values $\bar{x}_1, \dots, \bar{x}_k$ and frequencies $f_1, \dots, f_k$ and we find an approximation of the variance given by

$$s^2 \cong \frac{1}{n}\sum_{i=1}^{k}(\bar{x}_i - \widetilde{\bar{x}})^2 f_i = \left(\frac{1}{n}\sum_{i=1}^{k}(\bar{x}_i)^2 f_i\right) - \widetilde{\bar{x}}^2$$

where we denoted by $\widetilde{\bar{x}}$ the approximation $\frac{1}{n}\sum_{i=1}^{k}\bar{x}_i f_i$.

Quantile

The quantiles extend the concept of median.
The *quantile* of order $\alpha \in (0,1)$ denoted by q_α is the value such that the proportion of data smaller or equal to it is at least α and the proportion of data greater or equal to it is at least $1-\alpha$.

Particular cases:

- quantile of order $\alpha = 0.25$, $\alpha = 0.5$ and $\alpha = 0.75$ are called respectively 1°, 2° and 3° QUARTILE:
  - $q_{0.25}$, denoted by Q_1, has (roughly) 25% of data on its left
  - $q_{0.5}$, denoted by Q_2, is the median $Q_2 =$med
  - $q_{0.75}$, denoted by Q_3, has (roughly) 75% of data on its left

  the quartiles split the distribution into 4 parts with the same proportion of data.

The *interquartile range* (I.R.) is defined by the quantity

$$Q_3 - Q_1$$

It increases with the variability of the data set in the central part of the distribution. Note that between Q_1 and Q_3 there is the central 50% of data.

The *interquartile range* compared to the variance does not depend on all data and is robust with respect to outliers or extreme values.

Computing quantiles

As for computing the median, we have two cases:

- there exists i such that $F_i < \alpha$ and $F_{i+1} > \alpha \Rightarrow\ q_\alpha = v_{i+1}$
- otherwise, there exists $F_i = \alpha \Rightarrow q_\alpha = \dfrac{v_i + v_{i+1}}{2}$

Ex.: Find the three quartiles and $q_{0.43}$.

v_i	$\frac{f_i}{n}$	F_i
15	0.4	0.4
16	0.1	0.5
17	0.25	0.75
29	0.05	0.8
30	0.2	1

- Q_1: $\nexists\, v_i$ such that $F_i = 0.25$, and $F_1 > 0.25 \quad \Rightarrow \quad Q_1 = v_1 = 15$
- Q_2 is the median, therefore $Q_2 = \dfrac{16 + 17}{2} = 16.5$
- Q_3: v_3 such that $F_3 = 0.75 \quad \Rightarrow \quad Q_3 = \dfrac{v_3 + v_4}{2} = \dfrac{17 + 29}{2} = 23$
- as for the 43° percentile: $F_1 < 0.43$ and $F_2 > 0.43 \quad \Rightarrow \quad q_{0.43} = v_2 = 16$

Box-plot

The box-plot is an important type of graph which illustrates the main features of the data set. In particular, it gives information:

- center
- variability
- shape (symmetry)
- existence of outliers or extreme values

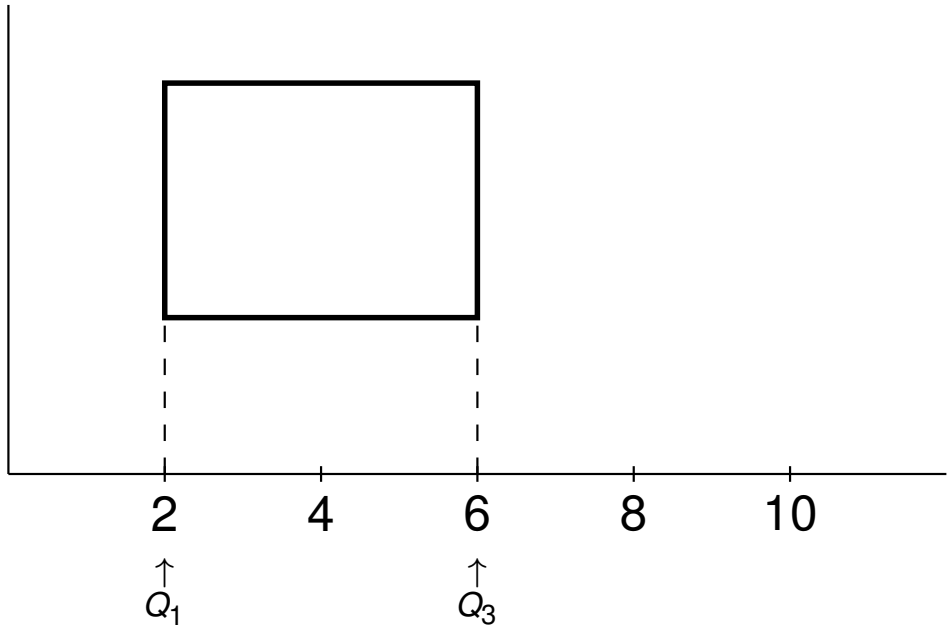
The ingredients to build a box-plot are:

- the three quartiles: Q_1, med and Q_3
- minimum and maximum values

The box-plot can be plotted in horizontal or in vertical with the same interpretation. We see it in horizontal.

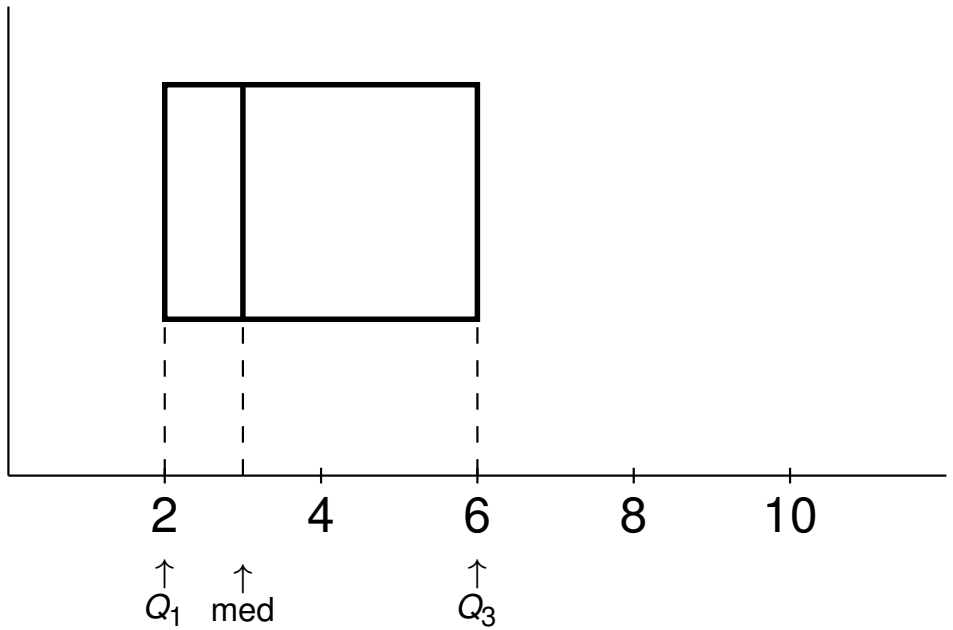
Box-plot: first and second steps

1) on the horizontal axis we put the values X
2) the first and third quartiles determine the box which contains the 50% central observations. For ex. if $Q_1 = 2$ and $Q_3 = 6$



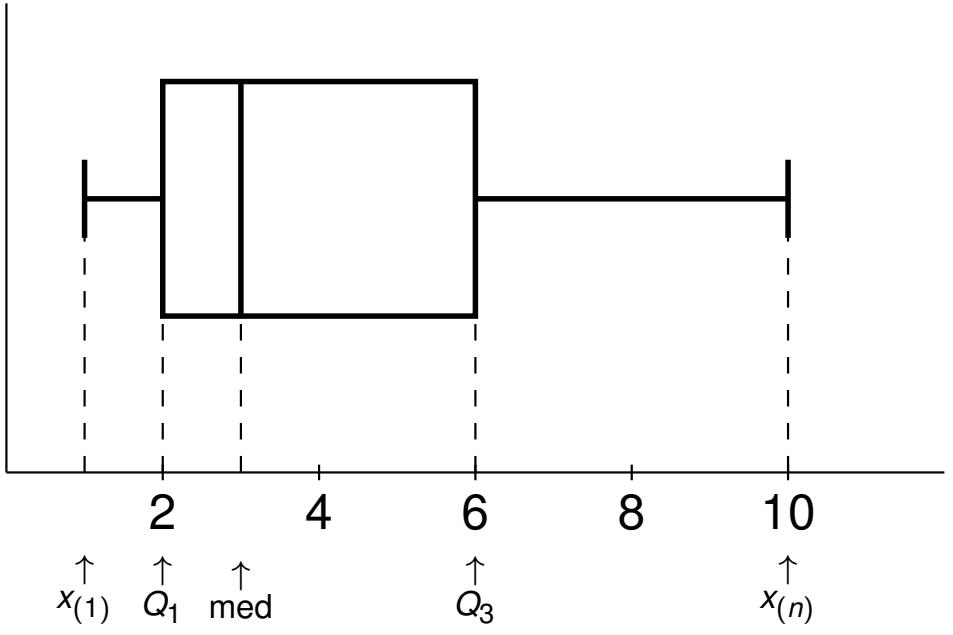
Box-plot: third step

1) on the horizontal axis we put the values X
2) the first and third quartiles determine the box which contains the 50% central observations. For ex. if $Q_1 = 2$ and $Q_3 = 6$
3) the median splits the box in 2. For ex. if med $= 3$



box-plot: fourth step

1) on the horizontal axis we put the values X
2) the first and third quartiles determine the box which contains the 50% central observations. For ex. if $Q_1 = 2$ and $Q_3 = 6$
3) the median splits the box in 2. For ex. if med $= 3$
4) lastly the lines connect the box to the minimum value and maximum value, respectively $x_{(1)}$ and $x_{(n)}$. For ex. if $x_{(1)} = 1$ and $x_{(n)} = 10$

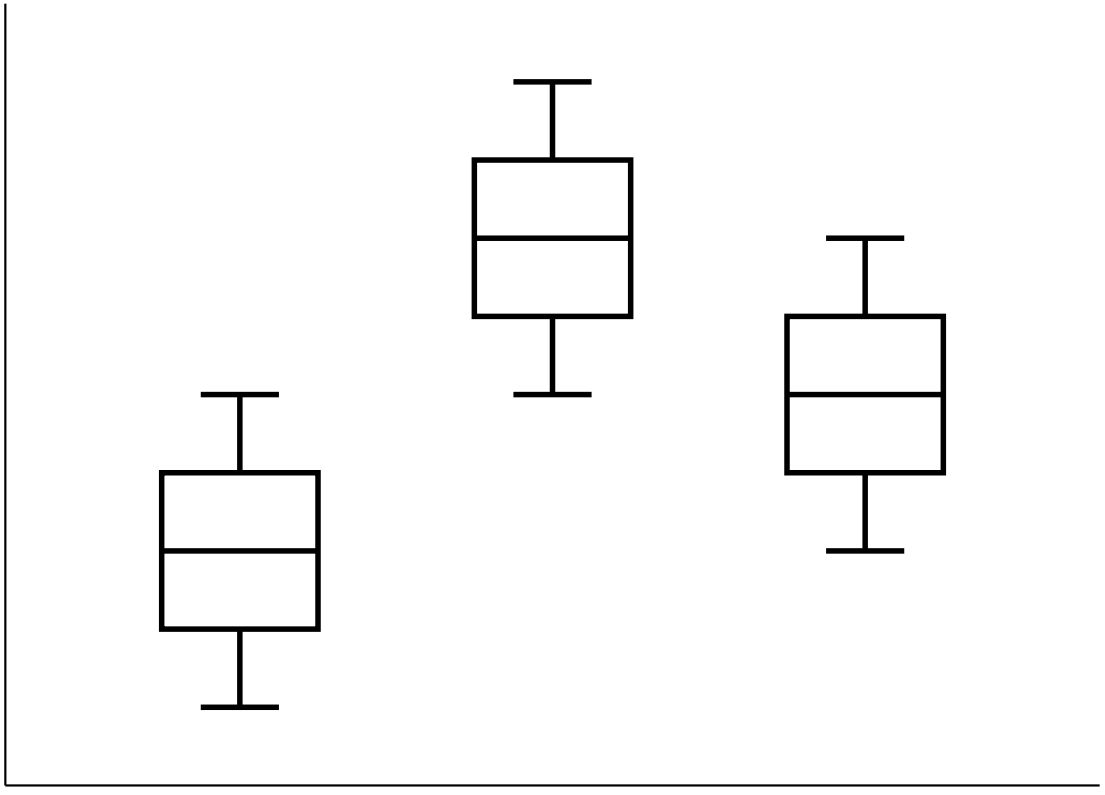


Features:

- tendency: median will represent the center
- variability: length of the box and of the lines
- shape: if the median is approximately at the center of the box and the lines are equally long, then we can say that the distribution is approximately symmetric
- in our example it was positively asymmetric.

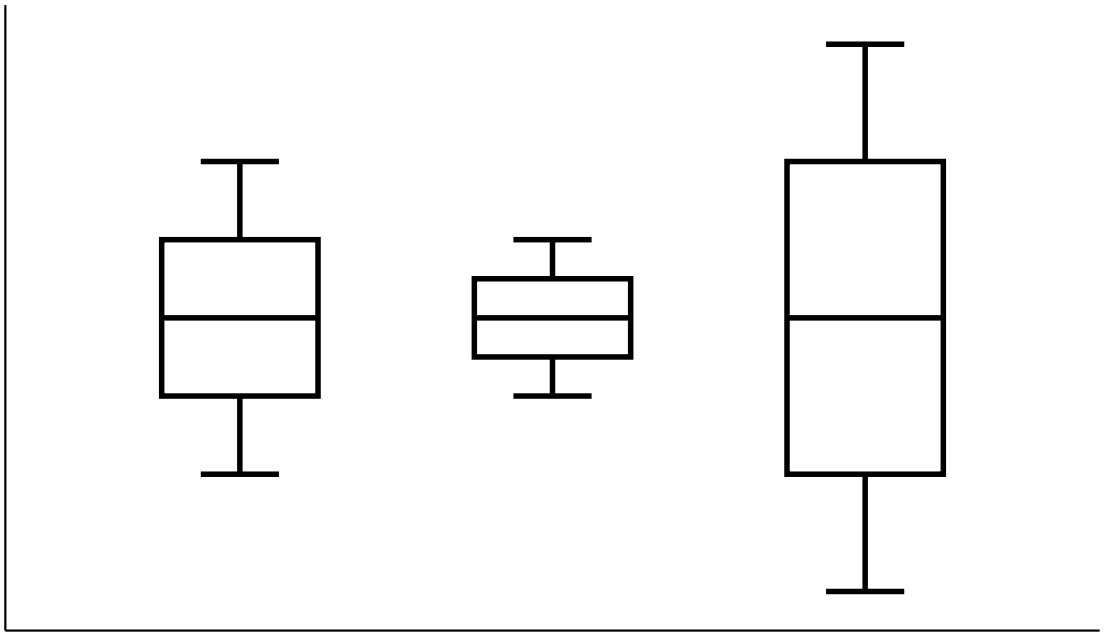
Comparing box-plots

Lining up box-plots related to different data sets one can better understand the different features of the data sets.



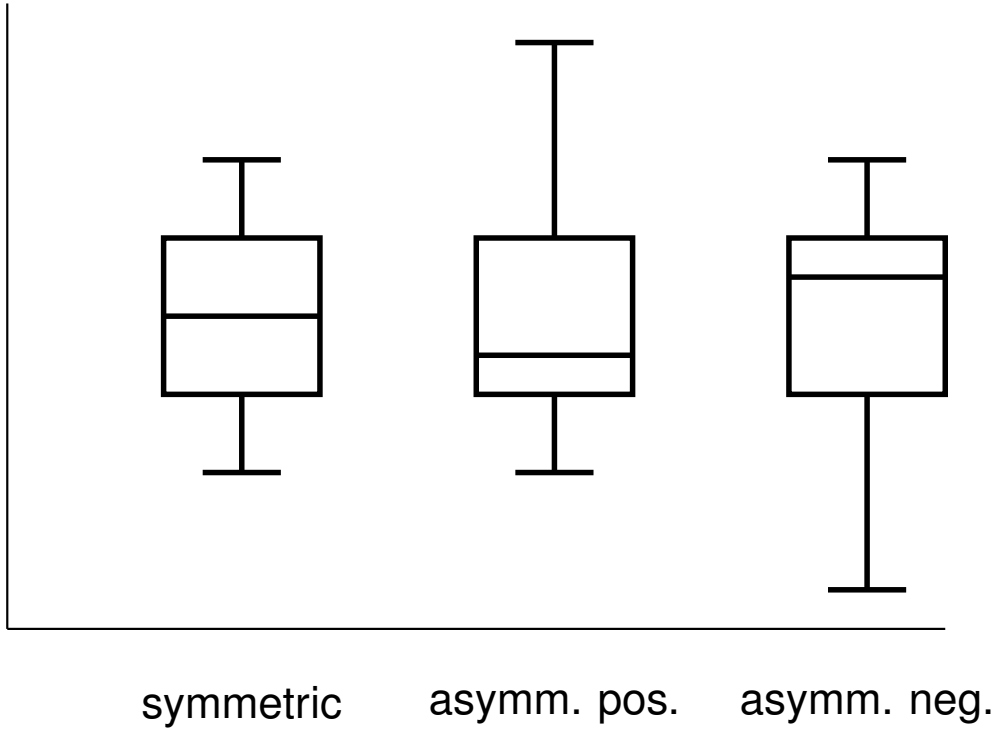
different centers, same shape and variability

Comparing box-plots



different variability

Comparison



different shape

Existence and representation of outliers

These are some conventional criteria to define outliers:

- a value is *adjacent*, if it belongs to the interval

$$[r_1, r_2] = [Q_1 - 1.5\text{I.R.},\ Q_3 + 1.5\text{I.R.}]$$

  where I.R.$= Q_3 - Q_1$ is the interquartile range
- a value is considered *far*, if it is outside $[r_1, r_2]$
- a value is considered *very far*, if it is outside the interval

$$[R_1, R_2] = [Q_1 - 3\text{I.R.},\ Q_3 + 3\text{I.R.}]$$

Example

$$x_{(1)} = 1,\ Q_1 = 2,\ \text{med} = 5,\ Q_3 = 9,\ x_{(n)} = 11$$

- I.R. $= Q_3 - Q_1 = 9 - 2 = 7$
- r_1, r_2

$$\begin{aligned} r_1 &= Q_1 - 1.5\text{I.R.} = 2 - 1.5(7) = 2 - 10.5 = -8.5 \\ r_2 &= Q_3 + 1.5\text{I.R.} = 9 + 1.5(7) = 9 + 10.5 = 19.5 \end{aligned} \quad \Rightarrow\ [r_1, r_2] = [-8.5, 19.5]$$

therefore $x_{(1)} = 1 \in [r_1, r_2]$ and $x_{(n)} = 11 \in [r_1, r_2]$; there are not values outside $[r_1, r_2]$.

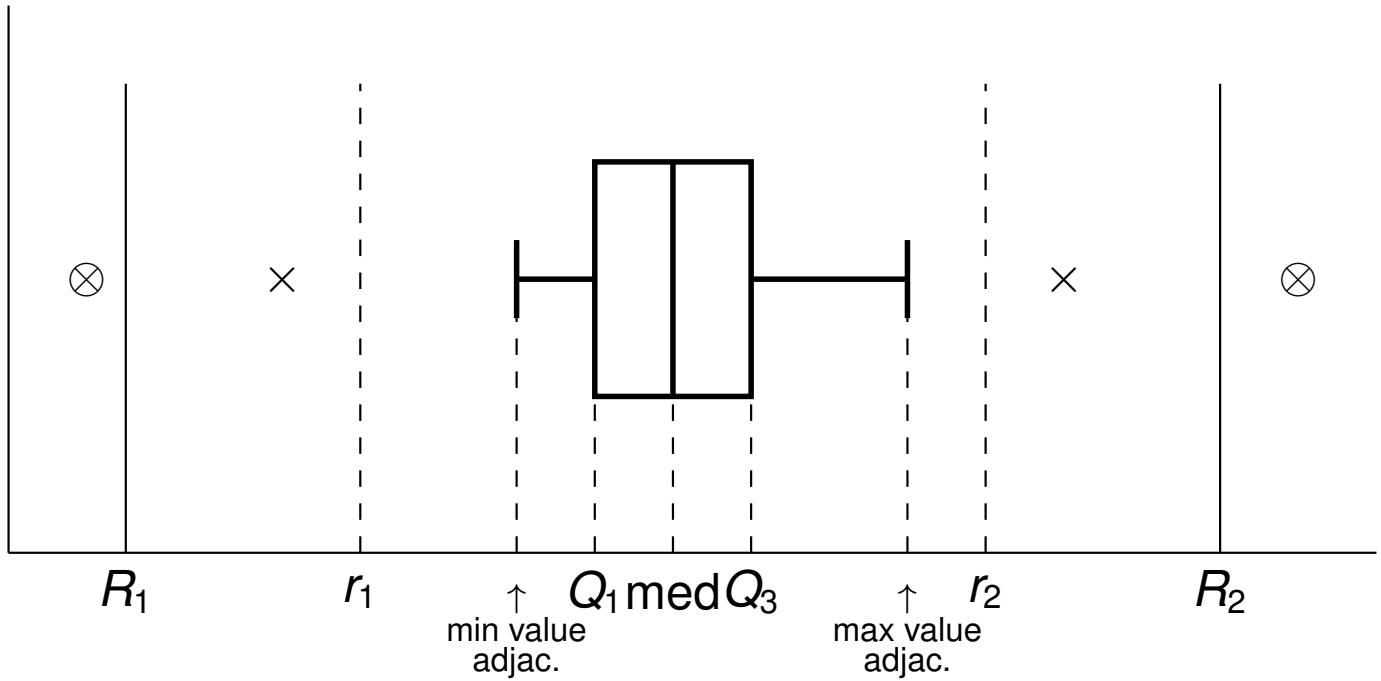
- R_1, R_2

$$\begin{aligned} R_1 &= Q_1 - 3\text{I.R.} = 2 - 3(7) = 2 - 21 = -19 \\ R_2 &= Q_3 + 3\text{I.R.} = 9 + 3(7) = 9 + 21 = 30 \end{aligned} \quad \Rightarrow \quad [R_1, R_2] = [-19, 30]$$

there are not values outside $[R_1, R_2]$

These information can be plotted in the box-plot, in the following way:

- the lines connect,
  - Q_1 to the smallest adjacent value;
  - Q_3 to the largest adjacent value.
- if there are not external values, they will arrive up to the minimum and maximum
- otherwise the external values are suitably highlighted



Last Example

$\{x_i\}_{i=1,\ldots,n} = \{-3, 10, 11, 13, 15, 17, 18, 19, 25, 48\}$

v_i	$\frac{f_i}{n}$	F_i
-3	0.1	0.1
10	0.1	0.2
11	0.1	0.3
13	0.1	0.4
15	0.1	0.5
17	0.1	0.6
18	0.1	0.7
19	0.1	0.8
25	0.1	0.9
48	0.1	1

- med: $\exists\, v_i$ t.c. $F_i = 0.5 \Rightarrow$ $\text{med} = \frac{v_5+v_6}{2} = \frac{15+17}{2} = 16$
- Q_1: $F_2 < 0.25$ e $F_3 > 0.25 \Rightarrow$ $Q_1 = v_3 = 11$
- Q_3: $F_7 < 0.75$ e $F_8 > 0.75 \Rightarrow$ $Q_3 = v_8 = 19$
- I.R.$= Q_3 - Q_1 = 19 - 11 = 8$

- $[r_1, r_2] = [Q_1 - 1.5\text{D.I.},\ Q_3 + 1.5\text{D.I.}] = [11 - 1.5(8),\ 19 + 1.5(8)] = [-1, 31]$
- $[R_1, R_2] = [Q_1 - 3\text{D.I.},\ Q_3 + 3\text{D.I.}] = [11 - 3(8),\ 19 + 3(8)] = [-13, 43]$

Outliers:

- $x_1 = -3 \quad \Rightarrow \quad x_1 \notin [r_1, r_2],\ x_1 \in [R_1, R_2]$
- $x_{10} = 48 \quad \Rightarrow \quad x_{10} \notin [R_1, R_2]$

Lines (max and min are adjacent values):

- $x_2 = 10$
- $x_9 = 25$