

Statistical Measures

Different measures can be used :

① Measures of central tendency

مقاييس
التركيبة المركزية

② Measures of variation

→ We use measures like mean, median, or mode to represent the center of the data with a single number.

→ The variation can also be expressed with a single number, most simply by finding the range, or difference between the highest and lowest values. It can also be expressed using the standard deviation or variance.

مقاييس التباين
والانحراف

Measures of central tendency (Measures of location)

Raw data

(A) Mean

Mean is the average of the values.

Let x_1, x_2, \dots, x_n be a sample of size n , then the mean (\bar{x}) is given by

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

Ex. let us have the data set

$$\begin{array}{cccccc} 3, & 2, & 7, & 6, & 4 & (n=5) \\ x_1 & x_2 & x_3 & x_4 & x_5 \end{array}$$

$$\text{Then } \bar{x} = \frac{1}{5} (3+2+7+6+4) = \frac{22}{5} = 4.4$$

(5) ~ (البيانات) ~ (توسط) ~ (معدل) ~ (معدل)

(B) Median defined,

Median is the value separating the higher half from the lower half of a data sample.

Let x_1, x_2, \dots, x_n be a sample of size n , and $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ be the sorted sample.

Then the median M is defined as

$$M = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & , \text{ if } n \text{ is odd.} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} & , \text{ if } n \text{ is even.} \end{cases}$$

Ex. For the set of the data

3, 2, 7, 6, 4.

Find the median.

Sol. Order the data to be

2, 3, 4, 6, 7

$x_{(1)}$ $x_{(2)}$ $x_{(3)}$ $x_{(4)}$ $x_{(5)}$

$(n=5)$
↓

odd

$$\text{Median } M = x_{\left(\frac{n+1}{2}\right)} = x_{(3)} = 4.$$

Ex. For the set of data

2, 3, 4, 6, 7, 120
 $X_{(1)}$ $X_{(2)}$ $X_{(3)}$ $X_{(4)}$ $X_{(5)}$ $X_{(6)}$ ($n=6$
even)

$$M = \frac{X_{(3)} + X_{(4)}}{2} = \frac{4 + 6}{2} = 5.$$

Note: Same result can be obtained
from the data 3, 2, 4, 4, 7, 120
 $n=6$ \rightarrow outlier.

$$M = \frac{X_{(3)} + X_{(4)}}{2} = \frac{4 + 4}{2} = 4.$$

الوسيط يرتكز على البيانات وليس على القيم المتطرفة لا تؤثر على الوسط الحسابي
أي أن الوسط لا يتأثر بالقيم المتطرفة بل يوجه للبيانات

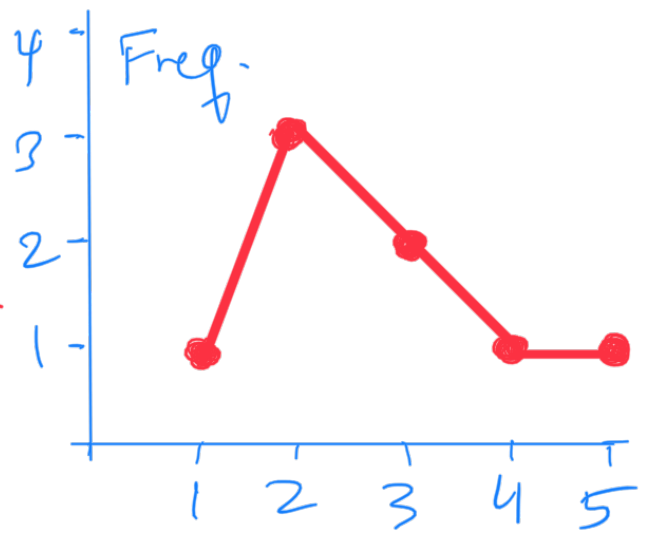
(C) Mode الكثرة

Mode of a set of data is the value that appears most often.

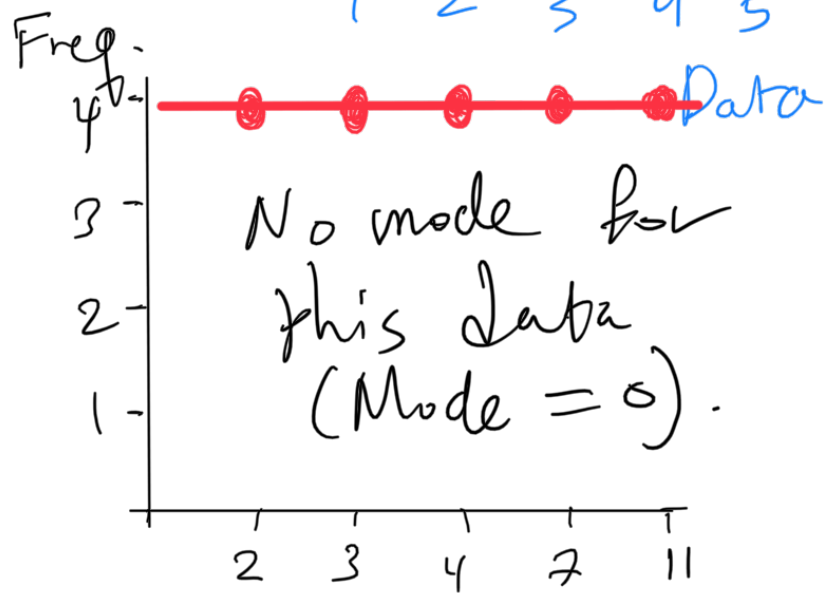
Ex. The data set 2, 3, 2, 4, 1, 2, 7, 3.

Value	Frequency
1	1
2	3
3	2
4	1
5	1

Mode = 2.



Ex.	x	f
1	2	4
	3	4
	4	4
	7	4
	11	4



No mode for this data (Mode = 0).

Ex.	x	f
1	2	2
	3	7
	4	3
	7	7
	11	1

Mode = 3 and 7.

Ex. Given that for the sample of 50 measurements, the mean is equal to 6. If you know that the first and the second observations are 4 and 6, find the sample mean for their other 48 observations.

Soln. $\bar{X}_{50} = \text{mean} = \frac{1}{n} \sum_{i=1}^n X_i$

So $\sum_{i=1}^{50} X_i = (50)(6) = 300.$

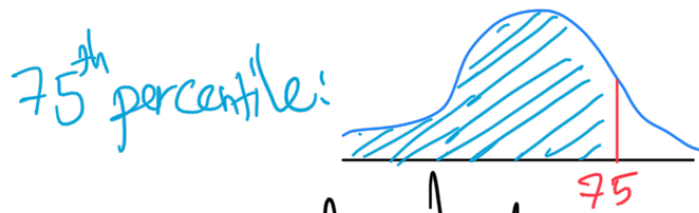
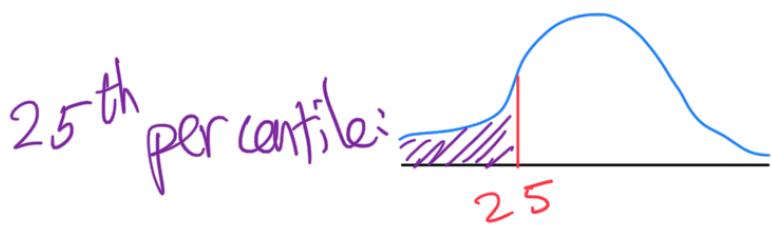
Hence $\sum_{i=1}^{48} X_i = \sum_{i=1}^{50} X_i - (X_1 + X_2)$

$= 300 - (6 + 4) = 290.$

Thus, $\bar{X}_{48} = \frac{\sum_{i=1}^{48} X_i}{48} = \frac{290}{48} \approx 6.042$

(D) Percentiles \rightarrow limited

Percentile is a measure indicating the value below which a given percentage of observations in a group of observations falls.



Let X_1, X_2, \dots, X_n be our sorted data,
then the 100th percentile is given by

$$P_{100p} = \begin{cases} \frac{X_{np} + X_{np+1}}{2}, & \text{if } np \text{ is an integer,} \\ X_{[np]}, & \text{if } np \text{ is not an integer.} \end{cases}$$

The following percentiles are special cases:

Quartiles (الرباعيات)



Q_1 : The first Q (الرباع الأولى) = P_{25}

The value for which %25 of observations are less than Q_1 .

Q_2 : The second Q (الثاني) = P

Median \rightarrow The value for which %50 of observations are less than Q.

Q_3 : The third Q (الثالث) = P

The value for which %75 of observations are less than Q.

Quartile range = $Q_3 - Q_1$

Ex. Consider the following data

x	f
3	4
4	7
5	6
6	3
7	4
8	6
Sum	30

Find (1) Q_1
(2) P_{60}

Soln. $n = 30$.

(1) $Q_1 = P_{25}$

So $P = \frac{25}{100}$

$$P_{100} = P_{25}. \text{ Now } np = \binom{30}{1} \left(\frac{25}{100} \right) = 75$$

$$\text{Then } Q = X_{\lceil 7.5 \rceil} = X_8 = 4.$$

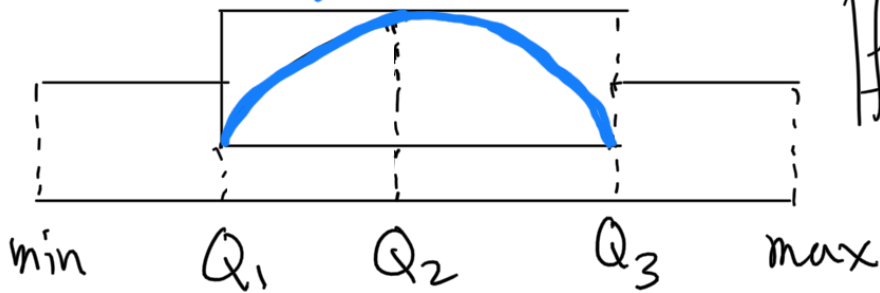
$$\rightarrow X_i = \begin{cases} 3, & 1 \leq i \leq 4, \\ 4, & 5 \leq i \leq 11, \\ 5, & 12 \leq i \leq 17, \\ 6, & 18 \leq i \leq 20, \\ \vdots & \end{cases}$$

$$(2) P_{60}$$

$$np = \binom{30}{1} \left(\frac{60}{100} \right) = 18 \text{ (integer)}$$

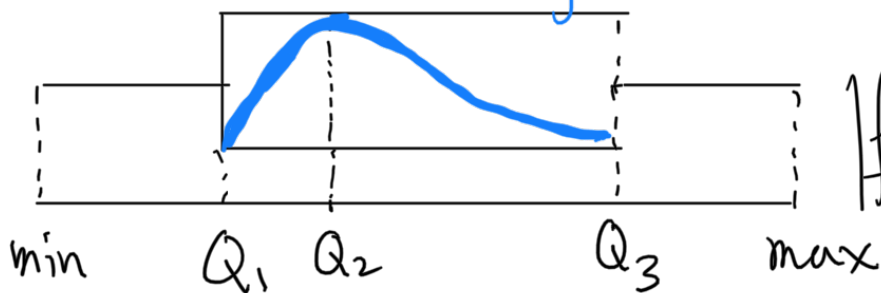
$$\text{So, } P_{60} = \frac{X_{18} + X_{19}}{2} = \frac{6 + 6}{2} = 6$$

skewness Symmetric

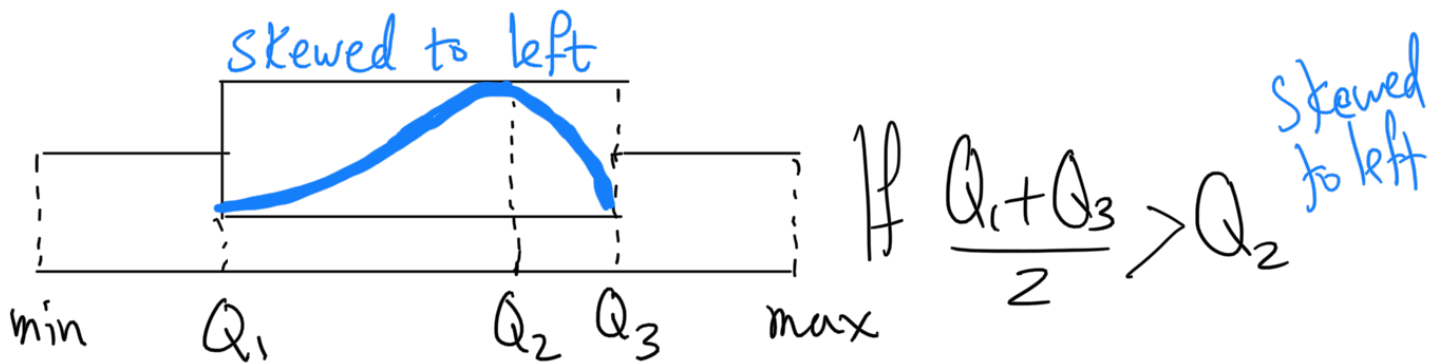


$$\text{If } \frac{Q_1 + Q_3}{2} = Q_2 \text{ Symmetric}$$

Skewed to right



$$\text{If } \frac{Q_1 + Q_3}{2} < Q_2 \text{ Skewed to right}$$



Ex.

(1) $Q_1 = 12, Q_2 = 16, Q_3 = 20 \Rightarrow$ Symmetric.

(2) $Q_1 = 12, Q_2 = 14, Q_3 = 20 \Rightarrow$ Skewed to right

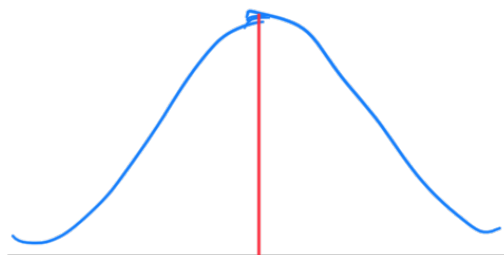
(3) $Q_1 = 12, Q_2 = 18, Q_3 = 20 \Rightarrow$ Skewed to left.

Facts:

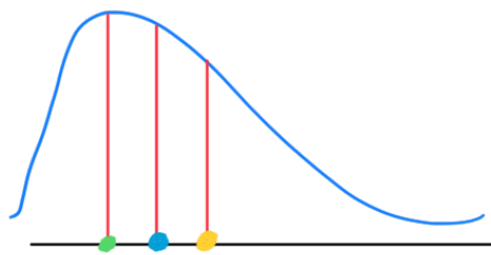
(1) $\text{Mode} < \text{Median} < \text{Mean} \Rightarrow$ Skewed to the right.

(1) $\text{Mode} > \text{Median} > \text{Mean} \Rightarrow$ Skewed to the left.

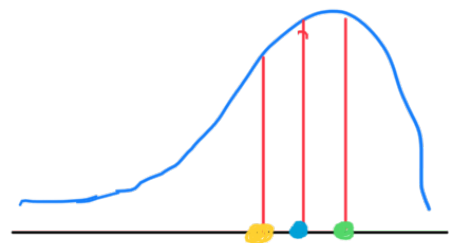
(1) $\text{Mode} = \text{Median} = \text{Mean} \Rightarrow$ Symmetric.



$\text{Mode} = \text{Median} = \text{Mean}$



$\text{Mode} < \text{Median} < \text{Mean}$



$\text{Mode} > \text{Median} > \text{Mean}$

Skewness describes the amount of asymmetry of the distribution of data about the mean, and is given by

$$\beta = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{s^3}$$

If $\beta = 0$, then the distribution is symmetric.

If $\beta < 0$, then the distribution is skewed to the left (negatively skewed).

If $\beta > 0$, then the distribution is skewed to the right (positively skewed).

Searching keywords:

- Measures of central tendency.
- Mean, median, mode.
- Percentiles, quartiles.
- Skewness.
- The University of Jordan الجامعة الأردنية
- Principles of Statistics مبادئ الإحصاء
- Baha Alzalg بهاء الزالق

References: See the course website

<http://sites.ju.edu.jo/sites/Alzalg/Pages/131.aspx>

For any comments or concerns, please use my email to contact me.



د. بهاء محمود الزالق
The University of Jordan
Dr. Baha Alzalg
baha2math@gmail.com

This material might not be used for commercial purposes.

Copyright: All Rights Reserved. B. Alzalg, 2020, Amman, Jordan.