Chapter III

Measures of Location

<u>D. 3. 1.</u> (*Arithmetic Average or Mean*)
Let $x_i, i = 1, 2, ..., n$, be the observed values of a variate X having the attributes $a_j, j = 1, 2, ..., k$.
The *arithmetic average (mean)* is defined by :

$$\mu := \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \text{(Simple arithmetic average for population data)}$$

$$= \frac{1}{N}\sum_{j=1}^{k} a_{ij} \cdot F\left(a_j\right) \qquad \text{(Weighted arithmetic average for population data)}$$

N : population size.

$$\bar{x} := \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{(Simple arithmetic average for sample data)}$$

$$= \frac{1}{n}\sum_{j=1}^{k} a_{ij} \cdot F\left(a_j\right) \qquad \text{(Weighted arithmetic average for sample data)}$$

n : sample size.

<u>Ex. 3. 1.</u> (See example Ex. 2. 2.)
Let

X : wages per hour in € of 30 workers in a firm:

17.05,	17.80,	17.80,	14.70,	15.15,	18.30,
16.20,	16.20,	16.55,	17.05,	15.15,	15.60,
15.60,	15.60,	16.20,	14.00,	15.00,	15.60,
16.55,	18.00,	17.50,	17.05,	16.55,	15.60,
15.60,	15.00,	16.20,	18.30,	17.50,	15.60.

$n = 30$, $\quad k = 12$.

1.

$$\bar{x} = \frac{1}{30} \cdot \sum_{i=1}^{30} x_i = \frac{489.00}{30} = 16.30 \text{ €}$$

2.

a_j	$F(a_j)$	$a_j.F(a_j)$
14,00	1	14.00
14,70	1	14.70
15,00	2	30.00
15,15	2	30.30
15,60	7	109.20
16,20	4	64.80
16,55	3	49.65
17,05	3	51.15
17,50	2	35.00
17,80	2	35.60
18,00	1	18.00
18,30	2	36.60
Total	30	489.00

$$\bar{x} = \frac{1}{30} \cdot \sum_{i=1}^{12} a_j \cdot F(a_j) = \frac{489.00}{30} = 16.30 \text{ €}$$

<u>D. 3. 2.</u> (***Arithmetic Average of a Grouped Data Set***)
The arithmetic average of a data set grouped in p classes is defined by:

$$\mu :\approx \frac{1}{N} \cdot \sum_{i=1}^{p} \bar{x}_i \cdot F_i \qquad (\textit{Population mean})$$

$$\bar{x} :\approx \frac{1}{n} \cdot \sum_{i=1}^{p} \bar{x}_i \cdot F_i \qquad (\textit{Sample mean})$$

<u>R. 3. 1.</u>

In case the informations $\bar{x}_i$ are not available, they will be replaced by the interval *midpoints*:

$$m_i := \frac{b_i + B_i}{2}$$

b_i, $i = 1, 2, ..., p$: lower bound of the class C_i

B_i, $i = 1, 2, ..., p$: upper bound of the class C_i

<u>R. 3. 2.</u>

$$\frac{1}{N} \cdot \sum_{i=1}^{p} b_i \cdot F_i \le \mu \le \frac{1}{N} \cdot \sum_{i=1}^{p} B_i \cdot F_i$$

$$\frac{1}{n}\cdot\sum_{i=1}^{p} b_i \cdot F_i \le \bar{x} \le \frac{1}{n}\cdot\sum_{i=1}^{p} B_i \cdot F_i$$

<u>Ex. 3. 1.</u> (*cont.*)

i	C_i	F_i	m_i	$m_i \cdot F_i$	$b_i \cdot F_i$	$B_i \cdot F_i$
1	[14.00, 15.00[	2	14.50	29.00	28.00	30.00
2	[15.00, 16.00[	11	15.50	170.50	165.00	176.00
3	[16.00 17.00[	7	16.50	115.50	112.00	119.00
4	[17.00 18.40[	10	17.70	177.00	170.00	184.00
Total		30		492.00	475.00	509.00

$$\bar{x} \approx \frac{492}{30} = 16.40 \text{ €}$$

$$\frac{475}{30} \approx 15.83 \le \bar{x} \le 16.97 \approx \frac{509}{30}$$

<u>R. 3. 3.</u> (***Some Properties of the Arithmetic Average***)

1.

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

2.

$$\sum_{i=}^{n}\left(x_i - \bar{x}\right)^2 \le \sum_{i=1}^{n}\left(x_i - a\right)^2, \quad a \in R^1$$

3.

If $\bar{x}_1$ and $\bar{x}_2$ are the arithmetic averages of two samples of sizes n_1 and n_2 respectively, then the arithmetic average $\bar{x}$ of the distribution combining the two can be calculated as

$$\bar{x} = \frac{n_1 \cdot \bar{x}_1 + n_2 \cdot \bar{x}_2}{n_1 + n_2}$$

(This formula can be extended to still more groups.)

4.

Given following linear transformation

$$\bar{x}_i = \alpha + \beta \cdot x_i, \qquad i = 1, 2, \ldots, n, \qquad \alpha, \beta \in R^1,$$

we obtain

$$\bar{x}^* = \alpha + \beta \cdot \bar{x}$$

(*Hint:* The above properties were formulated for the sample mean. Similar relations hold also for the population mean.)

<u>Ex. 3. 2</u>.
A car hire company charges a fixed sum of 40,00 € per day for its cars. The customer should pay 0.40 € for each kilometre. The average distance driven by the customers each day is about 250 km.
Find the average daily revenue of the company.

Solution:
We have:

$$\alpha = 40.00 \text{ €}, \qquad \beta = 0.40 \text{ €} \qquad \bar{x} = 250 \text{ km}$$

$$\bar{x}^* = 40 + 0.40 \cdot 250 = 140 \text{ €}$$

<u>D. 3. 3.</u> ***(Median of an Ungrouped Data Set)***
The m*edian* of an (ungrouped) data set is defined by:

$$Me := \begin{cases} x_{\left\lceil \frac{n+1}{2} \right\rceil} & \text{if } n \text{ is odd} \\ \frac{1}{2}\left(x_{\left\lceil \frac{n}{2} \right\rceil} + x_{\left\lceil \frac{n}{2}+1 \right\rceil} \right) & \text{if } n \text{ is even} \end{cases}$$

Here gives $\lceil\ \rceil$ the position of the observation in the *arranged* (ascending or descending) data set.

<u>Ex. 3. 3</u>.
Find the average of the following marks:

5, 2, 2, 1, 4, 5, 3, 2, 1, 3, 4.

The marks will be arranges in descending (worsening) order:

Position	1	2	3	4	5	6	7	8	9	10	11
Mark	1	1	2	2	2	3	3	4	4	5	5

$$Me = x_{\left\lceil \frac{11+1}{2} \right\rceil} = x_{[6]} = 3$$

Ex. 3. 4.
The following figures represent the monthly income of 11 people. Find their average monthly income (in €):

2600, 2500, 2650, 2720, 5240, 2450, 2700, 2750, 2500, 2750, 2600

1. As arithmetic average:

$$\bar{x} = \frac{31460}{11} = 2860 \text{ €}$$

The result is obviously unrealistic, since it is determined by an extreme value, namely 5240. All other incomes lie below this average.

2. As median

$$Me = x_{\left[\frac{11+1}{2}\right]} = x_{[6]} = 2650 \text{ €}$$

D. 3. 4. (***Median of a Grouped Data Set***)
Let $C_1, C_2, ..., C_p$ be given classes with the relative frequencies $f_1, f_2, ..., f_p$.
The *Median* of lies in the class C_i if

$$\sum_{j=1}^{i-1} f_j < 0.5 \quad \text{and} \quad \sum_{j=1}^{i} f_j \geq 0.5$$

and will be defined by

$$Me := b_i + \frac{0.5 - \sum_{j=1}^{i-1} f_j}{f_i} \cdot w_i$$

R. 3. 4.
The median is that value of the variate which divides the data set into two equal parts, one part comprising all values greater and the other all values lesser than the median. Geometrically, it divides the area under the histogram into two equal parts.

Ex. 3. 1. (continued)

$C_1 = [14.00, \quad 15.00[\,,$ $\qquad f_1 = \frac{2}{30}$

$C_2 = [15.00, \quad 16.00[$ $\qquad f_2 = \frac{11}{30}$

$C_2 = [16.00, \quad 17.00[$ $\qquad f_3 = \frac{7}{30}$

$C_4 = [17.00, \quad 18.40[$ $\qquad f_4 = \frac{10}{30}$

$$f_1 = \frac{2}{30} < 0.5$$
$$f_1 + f_2 = \frac{2}{30} + \frac{11}{30} = \frac{13}{30} < 0.5$$
$$f_1 + f_2 + f_3 = \frac{2}{30} + \frac{11}{30} + \frac{7}{30} = \frac{20}{30} \geq 0.5 .$$

Therefore,

$$Me \in C_3 ,$$

$$Me \approx 16 + \frac{0.5 - \frac{13}{30}}{\frac{7}{30}} \cdot 1 \approx 16.29 \text{ €}$$

<u>D. 3. 5.</u> (α-*Quantile of an Ungrouped Data Set*)
Given a *ranked* data set, the, α-Quantile ($0 < \alpha < 1$) is defined by :

$$\tilde{x}_\alpha := \begin{cases} x_{[k]} & \text{if } n \cdot \alpha \text{ is not integer} \\ & (k \text{ is then the following integer}) \\ \frac{1}{2}\left(x_{[k]} + x_{[k+1]}\right) & \text{if } n \cdot \alpha \text{ is integer} \\ & (k = n \cdot \alpha) \end{cases}$$

<u>Ex. 3. 4</u>. (continued)
For $\alpha = 0.25$, we have

$$n \cdot \alpha = 11 \cdot 0.25 = 2.75 , \qquad k = 3 , \qquad \tilde{x}_{0.25} = x_{[3]} = 2$$

For $\alpha = 0.75$, we have

$$n \cdot \alpha = 11 \cdot 0.75 = 8.25 , \qquad k = 9 , \qquad \tilde{x}_{0.75} = x_{[9]} = 4$$

<u>D. 3. 6.</u> (α-*Quantile of a Grouped Data Set*)
Let $C_1, C_2, ..., C_p$ be given classes with the relative frequencies $f_1, f_2, ..., f_p$.
The α-*Quantile* ($0 < \alpha < 1$) of the data set lies in the class C_i if

$$\sum_{j=1}^{i-1} f_j < \alpha \qquad \text{and} \qquad \sum_{j=1}^{i} f_j \geq \alpha$$

and will be defined by

$$\bar{x}_\alpha := b_i + \frac{\alpha - \sum_{j=1}^{i-1} f_j}{f_i} \cdot w_i$$

<u>Ex. 3. 1</u>. (continued)
For $\alpha = 0.25$, we have

$$f_1 = \frac{2}{30} < 0.25$$
$$f_1 + f_2 = \frac{2}{30} + \frac{11}{30} = \frac{13}{30} \geq 0.25$$

Therefore,

$$\tilde{x}_{0.25} \in C_2,$$

$$\tilde{x}_{0.25} \approx 15 + \frac{0.25 - \frac{2}{30}}{\frac{11}{30}} \cdot 1 = 15.50 \text{ €}$$

For $\alpha = 0.75$, we have

$$f_1 = \frac{2}{30} < 0.75$$
$$f_1 + f_2 = \frac{2}{30} + \frac{11}{30} = \frac{13}{30} < 0.75$$
$$f_1 + f_2 + f_3 = \frac{2}{30} + \frac{11}{30} + \frac{7}{30} = \frac{20}{30} < 0.75$$
$$f_1 + f_2 + f_3 + f_4 = \frac{2}{30} + \frac{11}{30} + \frac{7}{30} + \frac{10}{30} = 1 \geq 0.75$$

Therefore,

$$\tilde{x}_{0.75} \in C_4,$$

$$\tilde{x}_{0.75} \approx 17 + \frac{0.75 - \frac{20}{30}}{\frac{10}{30}} \cdot 1 \cdot 40 = 17.35 \text{ €}$$

<u>R. 3. 5.</u>

$$Me = \tilde{x}_{0.5}$$

D. 3. 7. (***Percentile***)
A quantile given as percent is called *percentile*.

D. 3. 8. (***Lower and Upper Quartile***)
The 0.25-quantile and the 0.75-Quantilee are called *lower (first) quartile* and *upper (third) quartile* respectively.

D. 3. 9. (***Mode of an Ungrouped Data Set***)
The value of the item which occurs most frequently is called the *mode* if it is unique.

It will be denoted by Mo.

D. 3. 10. (***Mode of a Grouped Data Set***)
The *mode* of a grouped data set lies in the class with the highest frequency. It will be defined by

$$Mo :\approx b_i + \frac{f_i - f_{i-1}}{2f_i - f_{i-1} - f_{i+1}} \cdot w_i$$

Ex. 3. 1. (*continued*)
1. Ungrouped

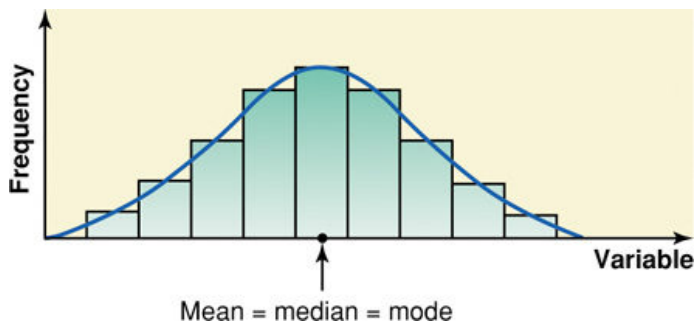
$Mo = 15.60$ € (This value occurs 7 times.)

2. Grouped

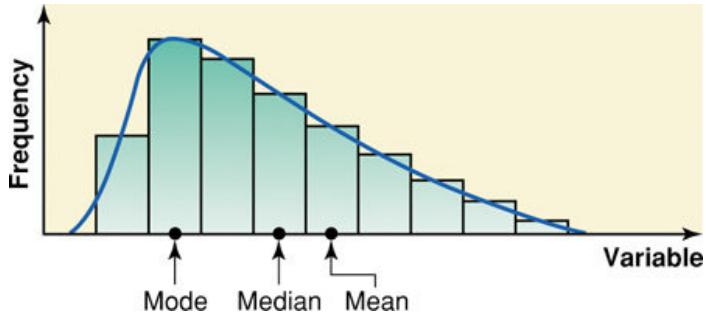
$Mo \in C_2$ (with the highest frequency 11)

$$Mo \approx 15 + \frac{\frac{11}{30} - \frac{2}{30}}{2 \cdot \frac{11}{30} - \frac{2}{30} - \frac{7}{30}} \cdot 1 = 15.69 \text{ €}$$

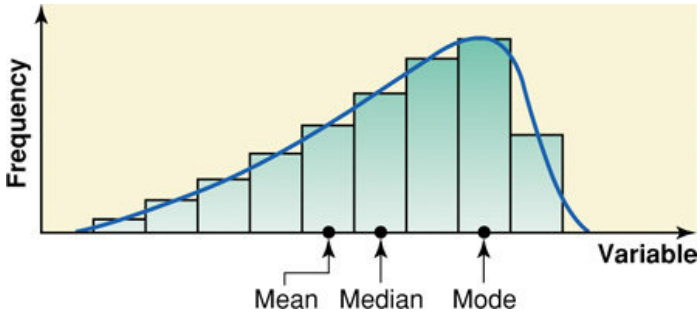
R. 3. 6. (***Relationship among the Mean, Median, and Mode***)
1. For a symmetric histogram and frequency distribution curve with one peak, the values of the mean, median, and mode are identical, and they lie in the centre of the distribution:



2. For a histogram and distribution curve skewed to the right, the value of the mean is the largest, that of the mode is the smallest, and the value of the median lies between the two:



3. If a histogram and a frequency distribution curve are skewed to the left, the value of the mean is the smallest and that of the mode is the largest, with the value of the median lying between these two:



<u>Ex. 3. 5</u>.
Someone has bought a share at the value of 1000 € . In the first year the share's value did not change at all, in the second year it increased by 10 per cent, and in the third year by 20 per cent.
Calculate the average rate of increase of the value of the share.

Solution:
Denoting by

n : year
C_n : value of the share at the end of the year n ,

let us have a look at the actual development of the share's value:

n	C_n
1	$C_1 = 1000$ €
2	$C_2 = 1000 + 1000 \cdot 0.10 = 1100$ €
3	$C_3 = 1100 + 1100 \cdot 0.20 = 1320.00$ €

Using the arithmetic average, we get:

$$\bar{x} = \frac{0.00 + 0.10 + 0.20}{3} = 0.10$$

Assuming now an average annual rate of growth of 10 per cent, we would expect the following development:

n	C_n
1	$C_1 = 1000 + 1000 \cdot 0.10 = 1100$ €
2	$C_2 = 1100 + 1100 \cdot 0.10 = 1210$ €
3	$C_3 = 1210 + 1200 \cdot 0.10 = 1331$ €

Surprisingly, the value of the capital at the end of the third year in the above table is not equal to what it should actually be! The reason is that the arithmetic average is not the appropriate mean to calculate an average rate of growth.

<u>D. 3. 11.</u> (*Geometric Mean*)

$$\bar{x}_g := \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}\,, \qquad x_i > 0, \quad i = 1, 2, \ldots, n$$

<u>Ex. 3. 5</u>. (*continued*)
Using the geometric mean, we get

$$\bar{x}_g := \sqrt[3]{(1 + 0.00) \cdot (1 + 0.10) \cdot (1 + 0.20)} = 1.09696131\,,$$

i. e. the true average annual increase of the value of the share was roughly 9.6 %. This can be verified in the following table:

n	C_n
1	$C_1 = 1000 + 1000 \cdot 0.09696131 = 1096.96$ €
2	$C_2 = 1096.96 + 1096.96 \cdot 0.09696131 = 1203.32$ €
3	$C_3 = 1203.32 + 1203.32 \cdot 0.09696131 = 1320.00$ €

(Last revised: 03.04.09)