Chapter 3: Numerical IndicatorsTables and Graphical Representations

I. Introduction: Numerical indicators (or parameters) are measurable data-driven metrics used to summarise data. They include location indicators and dispersion indicators. Common examples include means, medians, standard deviations.

II. Location Indicators (Central Tendency measures):

II.1. The Mode: (denoted by m_o) the mode is the modality or value which occurs most frequently i.e. the one corresponding to the greatest frequency. Notice that the mode is not necessary unique. We speak about modal class in the case of continuous character.

Example:

- Tthe statistical series 2,2,5,7,9,9,9,10,10,11,12,18 has as mode 9.
- The statistical series 3, 5, 8, 10, 12,15, 16 has no mode.
- The statistical series 2, 3, 4, 4, 4, 5, 5, 7, 7,7, 9 has two modes 4 et 7. We say that this series is bimodal.

*A series owing just mode is called unimodal.

*In the case of a continuous variable, we apply the following formula:

$$m_o = a_{i_0} + l\frac{\nabla_1}{\nabla_1+\nabla_2}$$

a_{i_0}: is the lower limit of the modal class,
l: is the length of the modal class,
∇_1: the difference between the frequency (relative frequency) of the modal class and the revious one.
∇_2: the difference between the frequency (relative frequency) of the modal class and the following one.

Example: Recall the constructed example in the two previous chapters on first year maths student.

- For The color of eyses X: (nominal)

x_i	n_i
Black	5
Blue	5
Green	3
Brown	4

We see that this statistical series is bimodal where both the modalities «Black» and «Blue» are modes.

- For The behaviour towrds morning coffee Y. (ordinal)

y_i	n_i
Never	3
Often	7
Somtimes	5
Always	5

The mode for this statistical series is the modality «Often»

- For the number of sisters and brothers Z. (discrete)

z_i	n_i
0	1
1	1
2	3
3	6
4	4
5	3
6	1
7	1

The mode $m_o = 3$. It corresponds to the geatest counts $n_4 = 6$

- For the heights which is a continuous character H (measured in m):

Classes	n_i
[1.45 , 1.53)	3
[1.53 , 1.61)	4
[1.61 , 1.69)	5
[1.69 , 1.77)	7
[1.77 , 1.85)	1

Here, we have a model class [1.69 , 1.77) corresponding to the heighest frequency n_4 =7 .

Then we calculate the mode as follow:

$$m_o = a_{i_0} + l\frac{\nabla_1}{\nabla_1 + \nabla_2} = 1.96 + 0.08\frac{(7-5)}{(7-5)+(7-1)}$$

II.2. The Median: (denoted by m_e): is the value that divides the ordered statistical series into two equals parts.

Example:

- The statistical series 3, 4, 4, 5, 6, 8, 8, 8, 10 has 6 as median.
- The statistical series 5, 5, 7, 9, 11, 12,15,18 has as median (9+11)/2= 10.

*The median corresponds to the cumulative relative frequency 0,5.

*In the case of a continuous variable, we apply the following formula:

$$m_e = l_{i_e} + (l_{i_e+1} - l_{i_e})\frac{0.5 * n - n_{i_e-1}^{cum}}{n_{i_e}^{cum} - n_{i_e-1}^{cum}}$$

with

l_{i_0}, l_{i_0+1}: are the lower and the upper limit of the median class respectively,

n: is the sample size

$n_{i_e-1}^{cum}$ (respectively $n_{i_e+1}^{cum}$): is the difference between the cumulative frequency of the median class and the cumulative frequency of the previous one (respectively the following one). Notice that this difference is not but the counts of the median class.

Remark:

1) The median is to be calculate for the quantitative characters only.

2) We can replace the absolute frequencies by the relatives ones (just devide the expression m_e of by n)

$$m_e = l_{i_e} + (l_{i_e+1} - l_{i_e})\frac{0.5 - \mathrm{f}_{i_e-1}^{cum}}{f_{i_e}^{cum} - \mathrm{f}_{i_e-1}^{cum}}$$

Example: Recall the constructed example in the two previous chapters on first year maths student.

- For the number of sisters and brothers Z. The median $m_e = 3$. It corresponds to $\frac{\mathrm{n}}{2} = 10$

z_i	n_i	$\mathrm{n}_{\mathrm{i}}^{\mathrm{cum}}$
0	1	1
1	1	2
2	3	5
3	6	$\frac{\mathrm{n}}{2} = 10$ 11
4	4	15
5	3	18
6	1	19
7	1	20
$\sum$	20	

- For the heights which is a continuous character H (measured in m):

z_i	n_i	$\mathrm{n}_{\mathrm{i}}^{\mathrm{cum}}$
[1.45 , 1.53)	3	3
[1.53 , 1.61)	4	7
[1.61 , 1.69)	5	$\frac{\mathrm{n}}{2} = 10$ 12
[1.69 , 1.77)	7	19
[1.77 , 1.85)	1	20
$\sum$	20	

Here, we have a modiane class [1.61 , 1.69) corresponding to the $\frac{\mathrm{n}}{2}$ =7 .

Then we calculate the mediane as follow:

$$m_e = l_{i_e} + (l_{i_e+1} - l_{i_e})\frac{0.5 * n - n_{i_e-1}^{cum}}{\mathrm{n}_{i_e}} = 1.61 + (1.96 - 1.61)\frac{7 - 7}{5}$$

II.2. The Mean: Let X be a discrete character having k values $x_1, \ldots, x_k$. Its mean $\bar{X}$ is given by:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{k} n_i\, x_i$$

Example: Let X be the number of sisters and brothers . (discrete)

x_i	n_i	$n_i * x_i$
0	1	0
1	1	1
2	3	6
3	6	18
4	4	16
5	3	15
6	1	6
7	1	7
$\sum$	20	69

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{k} n_i\, x_i = \frac{69}{20}$$

*If X is a continuous character owing k classes $[x_i, \ldots, x_{i+1}[,\ \ i = \overline{1,k}$. Its mean $\bar{X}$ is given by:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{k} n_i\, c_i$$

where c_i is the midvalue of the class $[x_i, \ldots, x_{i+1}[$,i.e. $c_i = \frac{x_{i+1} - x_i}{2}$.

Example: For the heights which is a continuous character H (measured in m):

Classes	n_i	c_i	$n_i * c_i$
[1.45 , 1.53)	3	1.49	4.47
[1.53 , 1.61)	4	1.57	6.28
[1.61 , 1.69)	5	1.65	8.25
[1.69 , 1.77)	7	1.73	12.11
[1.77 , 1.85)	1	1.79	1.77
$\sum$	20		32.88

The mean in this case is

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{k} n_i\, c_i = \frac{32.88}{20}$$

III. Dispersion Indicators (Spread parameters):

III.1 The Variance: If X is a discrete character having k values $x_1, \dots, x_k$ of mean $\bar{X}$. Its variance is given by:

$$V_X = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \bar{X})^2 \qquad \text{or} \qquad V_X = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \bar{X}^2$$

*If X is a continuous character owing k classes $[x_i, \dots, x_{i+1}[\quad i = \overline{1,k}$ of mean $\bar{X}$, It is given by:

$$V_X = \frac{1}{n}\sum_{i=1}^{k} n_i(c_i - \bar{X})^2 \qquad \text{or} \qquad V_X = \frac{1}{n}\sum_{i=1}^{k} n_i c_i^2 - \bar{X}^2$$

where c_i is the midvalue of the class $[x_i, \dots, x_{i+1}[$, $c_i = \frac{x_{i+1} - x_i}{2}$

III.1 The Standard Deviation: $\sigma_X = \sqrt{V_X}$

Examples:

- Let X be the number of sisters and brothers . (discrete)

x_i	n_i	$n_i * x_i$	$n_i * x_i^2$
0	1	0	0
1	1	1	1
2	3	6	12
3	6	18	54
4	4	16	64
5	3	15	75
6	1	6	36
7	1	7	49
$\sum$	20	69	291

$$V_X = \frac{291}{20} - \left(\frac{69}{20}\right)^2$$

- Le X be the heights measured in m:

Classes	n_i	c_i	$n_i * c_i$	$n_i * c_i^2$
[1.45 , 1.53)	3	1.49	4.47	6.6603
[1.53 , 1.61)	4	1.57	6.28	9.8596
[1.61 , 1.69)	5	1.65	8.25	13.6125
[1.69 , 1.77)	7	1.73	12.11	20.9503
[1.77 , 1.85)	1	1.79	1.77	3.1683
$\sum$	20		32.88	54.251

The variance in this case is

$$V_X = \frac{54.251}{20} - \left(\frac{32.88}{20}\right)^2$$