

Module : Probabilities-Statistics

Worksheet n°1

(Basic definitions - One-variable statistical series)

Exercise n°1 : (The different types of statistical variables)

- 1) Classify the variables below according to their type :
 - a) Number of people per household b) Height (cm) c) Sex.
 - d) marital status e) Weight (kg) f) Level of education
 - g) Mode of transport
- 2) Specify the modalities or values they can take.

Exercise n°2 : (qualitative variable) The following table shows the blood groups of 2nd year ST students

B	AB	A	A	O	A	A	B	AB	O	A	B	O	AB	B	O	O	A
A	O	B	O	A	A	O	O	O	O	O	A	A	AB	B	A	A	A
A	O	O	A	AB	B	B	A	A	B	AB	AB	B	A	A	AB	A	O

- 1) Indicate the population studied, the statistical unit, the character studied and its nature .
- 2) Draw up the table of the absolute frequencies n_i representing this statistical serie, calculating the relative frequencies f_i and percentages p_i .
- 3) Give two graphical representations suitable for this type of characteristic.
- 4) What is the mode of the series ?

Exercise n°3 : Show that :

- i) The sum of the deviations from the mean is 0 :

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

- ii) Linearity property : If

$$\forall i \in \{1, 2, \dots, k\}, a, b \in \mathbb{R} \text{ we have : } y_i = ax_i + b, \text{ then : } \bar{y} = a\bar{x} + b.$$

- iii) * Property of partial means : Consider two statistical series with respective means $\bar{x}_1$ and $\bar{x}_2$ and respective numbers N_1 and N_2 .

The mean of the two series is

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N_1 + N_2}$$

Exercise n°4 :

The variance is the quantity :

$$V(x) = \sum_{i=1}^k \frac{n_i(x_i - \bar{x})^2}{N} = \sum_{i=1}^k f_i(x_i - \bar{x})^2 \quad (1)$$

Show that :

$$1) V(x) = \sum_{i=1}^k \frac{n_i x_i^2}{N} - \bar{x}^2 = \sum_{i=1}^k f_i x_i^2 - \bar{x}^2$$

2) If $\forall i \in \{1, 2, \dots, k\}, a, b \in \mathbb{R}$ we have : $y_i = ax_i + b, a, b \in \mathbb{R}$, then :

$$V(y) = a^2 V(x),$$

$$\sigma_y = |a| \sigma_x.$$

3) $V(x) \geq 0$.

Exercise n°5 : We are interested in the age of each of the 50 employees in a company. We have the following raw data : 36; 30; 30; 56; 58; 47; 30; 45; 47; 18; 47; 33; 26; 51; 41; 33; 45; 39; 36; 41; 51; 21; 33; 30; 18; 56; 24; 26; 41; 26; 37; 26; 33; 39; 51; 56; 33; 24; 51; 37; 24; 37; 41; 41; 45; 33; 45; 33; 30; 37.

1. Indicate the population studied, the statistical unit, the character studied and its nature;

" 2. Represent the data in a statistical table, calculating the frequencies, percentages and ascending cumulative frequencies;

3. Determine the arithmetic mean of the age, the mode, the median;

4. Group these results into classes of the same widths, then draw up a table of the frequencies, specifying the central of the classes, the percentages and the ascending cumulative frequencies;

5. Construct a histogram for frequency distribution;

6. Determine the mode (by calculation and graphically), median and mean;

7. Compare the two means and comment on the results.

Exercise n°6 : A taxi company is interested in the mileage performed by its vehicles. To this end, it recorded the mileage of 50 of its taxis for a morning's work.

Classes	[10; 20[	[20; 30[	[30; 40[	[40; 60[	[60; 90[	Σ
frequencies	8	12	20	6	4	50

1. Indicate the population, its size, the variable and its type.

2. Draw the histogram of this distribution.

3. Give the modal class, median, mean and standard deviation of the distribution.

***Exercise n°7 :** We consider two groups of students. We record their exam marks in the following two tables :

Group A marks	8	9	10	11	Group B marks	6	8	9	13	14
frequencies	6	6	3	3	frequencies	6	6	3	3	3

1) Calculate the mean and standard deviation for each group.

2) Compare the two groups.

***Exercise n°8 :** In a city, 45 families were surveyed for the number of cell phones they used. Prepare a discrete frequency array based on their replies as recorded below. 1; 3; 2; 2; 2; 2; 1; 2; 1; 2; 2; 3; 3; 3; 3; 3; 3; 2; 3; 2; 2; 6; 1; 6; 2; 1; 5; 1; 5; 3; 2; 4; 2; 7; 4; 2; 4; 3; 4; 2; 0; 3; 1; 4; 3.

1. Indicate the population studied, the statistical unit, the character studied and its nature;

2. Represent the data in a frequency table, calculating the frequencies, percentages and ascending cumulative frequencies;

3. Determine the mean of the number of cell phones, the mode, the median and standard

deviation ;

4. Give the appropriate graphical representation for this type of characteristic.