Mohamed Boudiaf University - M'sila **University Year : 2025/2026**
Faculty of Technology
Department of Electrical Engineering
Module : Probabilities and Statistics

Solutions to the Directed Works Series n°1

***Exercice 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
frequencies	6	6	3	3

Group B marks	6	8	9	13	14
frequencies	6	6	3	3	3

1) Calculate the mean and standard deviation for each group.
2) Compare the two groups.

Solution :

1) Mean and Standard Deviation

Group A : $N_A = 18$

$$\bar{x}_A = \frac{6\times 8 + 6\times 9 + 3\times 10 + 3\times 11}{18} = \frac{165}{18} \approx 9.17$$

$$\sigma_A^2 = \frac{1533}{18} - (9.17)^2 \approx 1.14 \quad \Rightarrow \quad \sigma_A \approx 1.07$$

Group B : $N_B = 21$

$$\bar{x}_B = \frac{6\times 6 + 6\times 8 + 3\times 9 + 3\times 13 + 3\times 14}{21} = \frac{192}{21} \approx 9.14$$

$$\sigma_B^2 = \frac{1938}{21} - (9.14)^2 \approx 8.61 \quad \Rightarrow \quad \sigma_B \approx 2.94$$

Coefficient of Variation

$$CV_A = \frac{\sigma_A}{\bar{x}_A} = \frac{1.07}{9.17} \approx 0.117 \quad (11.7\%)$$

$$CV_B = \frac{\sigma_B}{\bar{x}_B} = \frac{2.94}{9.14} \approx 0.322 \quad (32.2\%)$$

2) Comparison of the Two Groups

- The means are almost identical (≈ 9.1). - Group A has a much lower standard deviation and CV, showing more consistency. - Group B has a higher dispersion (CV 32.2%), indicating greater variability in marks.

Conclusion : Group A is more homogeneous and stable in performance, while Group B is more dispersed.

Exercice 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.

Solution :

Ages of 50 employees
We are interested in the age of each of the 50 employees in a company.

1) Study elements
— Population : 50 employees of the company.
— Statistical unit : One employee.
— Character studied : Age.
— Nature : Quantitative discrete.

2) Statistical table with frequencies

Age x_i	Absolute n_i	Relative $f_i = n_i/50$	Percentage p_i	Cum. N_i
18	2	0.04	4%	2
21	1	0.02	2%	3
24	3	0.06	6%	6
26	4	0.08	8%	10
30	5	0.10	10%	15
33	7	0.14	14%	22
36	2	0.04	4%	24
37	4	0.08	8%	28
39	2	0.04	4%	30
41	5	0.12	12%	35
45	4	0.08	8%	39
47	3	0.06	6%	42
51	4	0.08	8%	46
56	3	0.06	6%	49
58	1	0.02	2%	50
Total	50	1	100%	/

3) Mean, mode, median (raw data)

$$\bar{x} = \frac{\sum n_i \times x_i}{50} = \frac{1868}{50} = 37.36 \text{ years}$$

Mode : $Mo = 33$ (highest frequency $n = 7$).
Median : position 25 and 26 fall in value 37, so :

$$Me = 37 \text{ years.}$$

4) Grouping into equal-width classes Grouping into equal-width classes and frequency table

we can view the frequency table as follow ::

Step1 : Range :$w = x_{max} - x_{min} = 58 - 18 = 40$

Step2 : With Sturges' rule : the number of classes $1 + 3.322Log50 = 6.6$

$$k = \lfloor k \rfloor + 1 = \lfloor 6.6 \rfloor + 1 = 6 + 1 = 7$$

Step3 : the Width of classes $L = \frac{w}{k} = \frac{40}{7} \simeq 5.7 \simeq 6.$

Class	Frequency n_i	Midpoint c_i	Percentage p_i	Cumulative N_i
$[18; 24[$	3	21	6%	3
$[24; 30[$	7	27	14%	10
$[30; 36[$	12	33	24%	22
$[36; 42[$	13	39	28%	35
$[42; 48[$	7	45	14%	42
$[48; 54[$	4	51	8%	46
$[54; 60]$	4	57	6%	50
Total	50	/	100%	/

5) Histogram of grouped frequencies

Histogram for the grouped frequency distribution

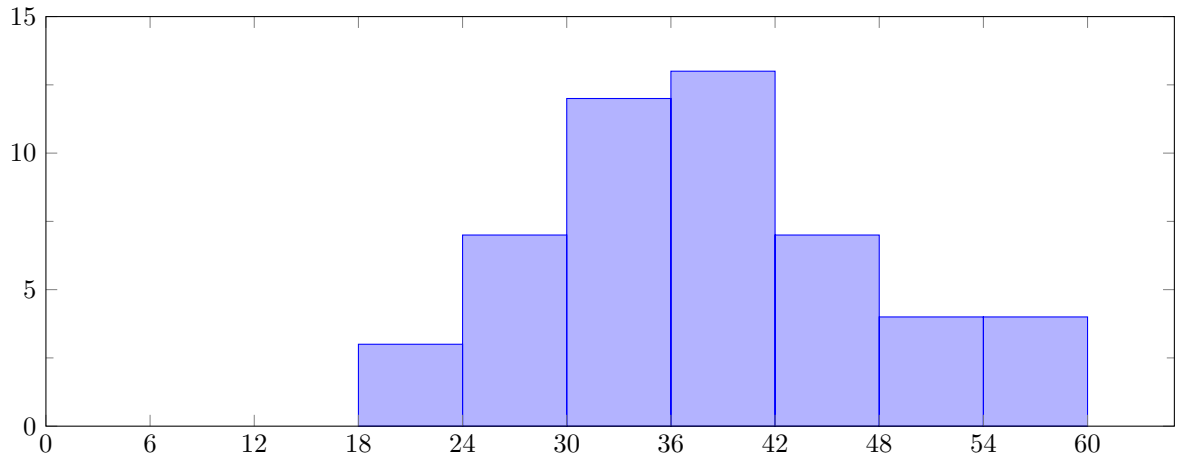

FIGURE 1 – Histogram of the frequency distribution by age class.

6) Central tendency for grouped data

Grouped mean :

$$\bar{x}_g = \frac{\sum n_i \times c_i}{50} = \frac{3\times 21 + 7\times 27 + 12\times 33 + 13\times 39 + 7\times 45 + 4\times 51 + 4\times 57}{50} = \frac{1902}{50} = 38.04.$$

Mode (modal class $[36;42[$, with $n_m = 13$, previous $n_p = 12$, next $n_s = 7$) :

$$\text{Mo} \approx L + \frac{(n_m - n_p)}{(n_m - n_p) + (n_m - n_s)} \times h = 36 + \frac{(13-12)}{(13-12)+(13-7)} \times 6 = 36 + \frac{1}{7} \times 6 \approx 36.86.$$

Median (interpolated in median class $[36;42[$) :

$$Me \approx L + \frac{(N/2 - \text{cum.before})}{n_{\text{class}}} \times h = 36 + \frac{(25-22)}{13} \times 6 \approx 37.38.$$

7) Comparison of means

Raw mean : 37.36 ; Grouped mean : 38.04. They are close, but the grouped mean is slightly higher due to class midpoints approximation.

Exercice 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.

Solution :

1) Study components

— **Population :** 50 taxis of the company.
Size : $N = 50$.

— **Variable :** mileage (km).
Type : Quantitative continuous.

2) Histogram (using adjusted frequencies) Let class widths be $a_i = (10, 10, 10, 20, 30)$. Compute adjusted frequencies

$$h_i = \frac{a}{a_i} n_i, \quad a = \text{GCD}(a_1, \ldots, a_5) = 10.$$

Class centers c_i are the midpoints.

classes	$[10;20[$	$[20;30[$	$[30;40[$	$[40;60[$	$[60;90[$
Frequencies n_i	8	12	20	6	4
Adjusted $h_i = \frac{10}{a_i} n_i$	8	12	20	3	1.333
Cumulative N_i	8	20	40	46	50
Centers c_i	15	25	35	50	75

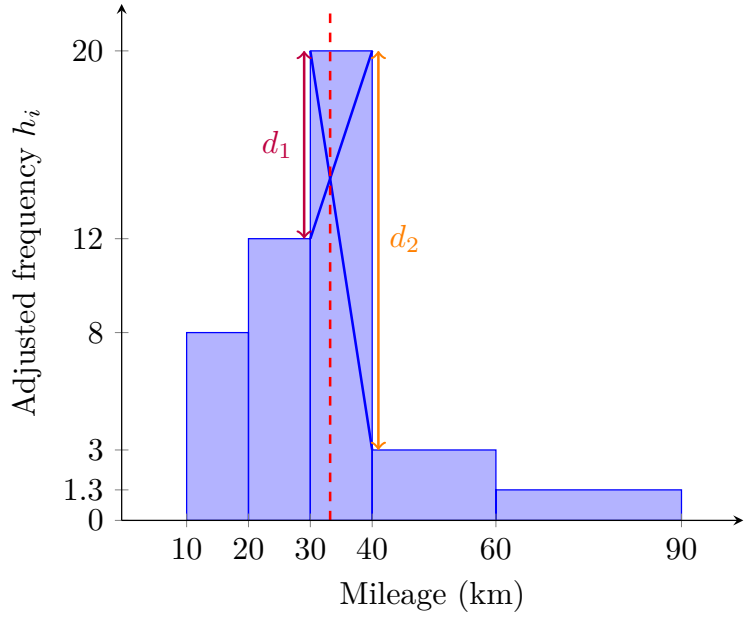

FIGURE 2 – Histogram of the Mileage of Company Taxi.

4) Measures (mode, median, mean, standard deviation)
Modal class is $[30; 40[$ (highest frequency a).

$$\hat{x} = L + \frac{d_1}{d_1 + d_2} \cdot h$$

where :
— L : lower boundary of the modal class,
— h : class width,
— n_m : frequency of the modal class,
— n_{prev}, n_{next} : frequencies of neighboring classes,
— $d_1 = n_m - n_{\text{prev}}$, $d_2 = n_m - n_{\text{next}}$.
Applied here : $L = 30$, $h = 10$, $n_m = 20$, $n_{\text{prev}} = 12$, $n_{\text{next}} = 6$.

$$Mo = 30 + \frac{20 - 12}{(20 - 12) + (20 - 3)} \cdot 10 = 30 + \frac{8}{25} \cdot 10 = 33.2 \text{ km}.$$

Median (grouped)

$N/2 = 25$. $[30; 40[$ is 20, so $Me \in ([30; 40[$. Then

$$Me = L + \left(\frac{\frac{N}{2} - F_{\text{prev}}}{f_{\text{class}}} \right) \cdot h = 30 + \frac{25 - 20}{20} \cdot 10 = 32.5 \text{ km}.$$

Mean (grouped)

Using class centers $c_i = \{15, 25, 35, 50, 75\}$:

$$\bar{x} = \frac{\sum c_i n_i}{N} = \frac{15 \cdot 8 + 25 \cdot 12 + 35 \cdot 20 + 50 \cdot 6 + 75 \cdot 4}{50} = \frac{1720}{50} = 34.4 \text{ km}.$$

Standard deviation (grouped, population)

$$\sigma = \sqrt{\frac{\sum c_i^2 n_i}{N} - \bar{x}^2} = \sqrt{\frac{71300}{50} - 34.4^2} = \sqrt{1426 - 1183.36} \approx \sqrt{242.64} \approx 15.58 \text{ km}.$$

Note : Adjusted frequencies are a convenient scaled form of densities. For drawing, densities $\frac{n_i}{a_i}$ are the canonical heights.

***Exercice 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 ; 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.

Solution :

1) Study components

- **Population studied :** 45 families in the city.
- **Statistical unit :** One family.
- **Characteristic studied :** Number of cell phones per family.
- **Nature of the characteristic :** Quantitative discrete.

2) Frequency table

Cell phones	**Frequency** n_i	**Percentage** p_i	**Cumulative** N_i
0	1	2.22%	1
1	7	15.56%	8
2	15	33.33%	23
3	12	26.67%	35
4	5	11.11%	40
5	2	4.44%	42
6	2	4.44%	44
7	1	2.22%	45
Total	45	100%	

TABLE 1 – Discrete frequency distribution of the number of cell phones per family.

3) Mean, mode, median, and standard deviation

- **Mean :**

$$\bar{x} = \frac{\sum x_i n_i}{N} = \frac{122}{45} \approx 2.71$$

- **Mode :** 2 (highest frequency : 15).
- **Median :** For $N = 45$, the 23rd observation is 2 Median = 2.
- **Standard deviation :**

$$\sigma = \sqrt{\frac{\sum x_i^2 n_i}{N} - \bar{x}^2} = \sqrt{\frac{426}{45} - \left(\frac{122}{45}\right)^2} \approx 1.46$$

4) Graphical representation For discrete quantitative data, the appropriate representation is a **bar chart** :

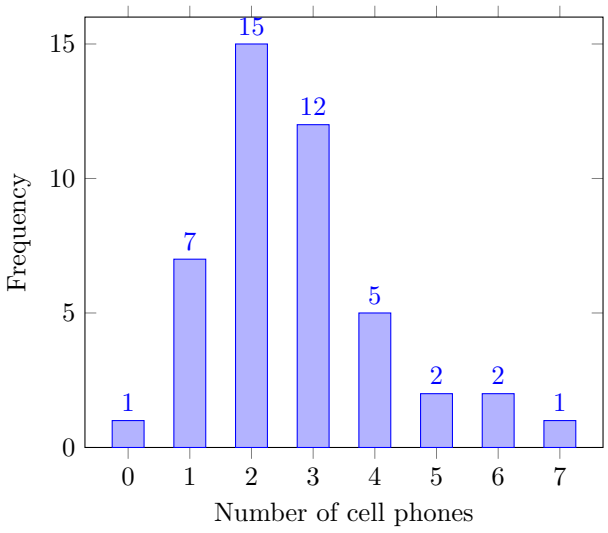

FIGURE 3 – Bar chart of the distribution of cell phones per family.