Correction Série 1

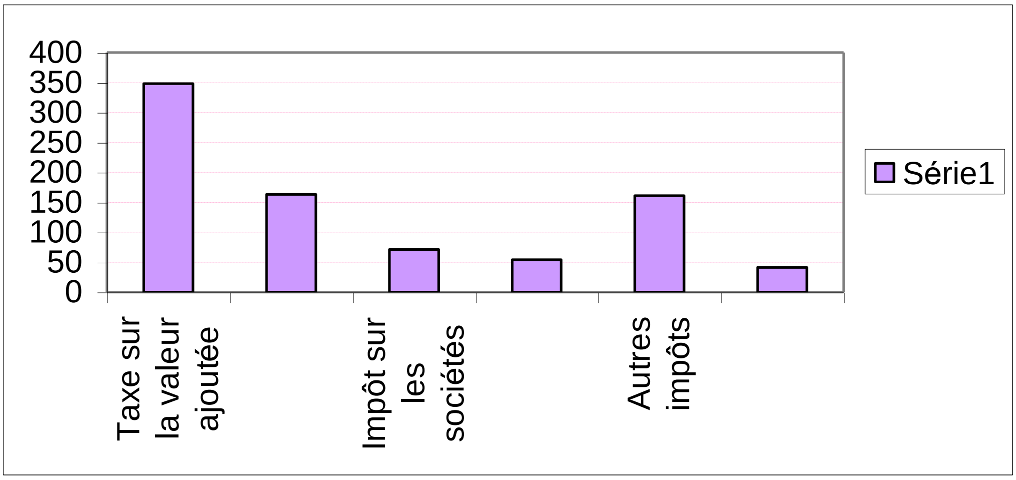
Exercice n° 1 :

1) La population est 838 milliards de dinars de recettes de l'Etat pour l'année « N ». Le caractère est nature des recettes de l'Etat. Il est qualitatif. Il n'existe pas, alors de variable statistique.

2) Pour pouvoir représenter le graphique en tuyaux d'orgue, nous ajoutons au tableau n°1 la colonne n° 3 :

(1) Recette de l'Etat	(2) Effectifs : n_i	(3) Fréquences : f_i (%)	(4) $\theta_i = 360 \times f_i$
Taxe sur la valeur ajoutée	348	41,527	149,497
Impôt sur le revenu	163	19,451	70,024
Impôt sur les sociétés	71	8,472	30,499
Taxe sur les produits pétroliers	54	6,444	23,198
Autres impôts	161	19,212	69,163
Recettes non fiscales	41	4,894	17,619
Total	838	100	360

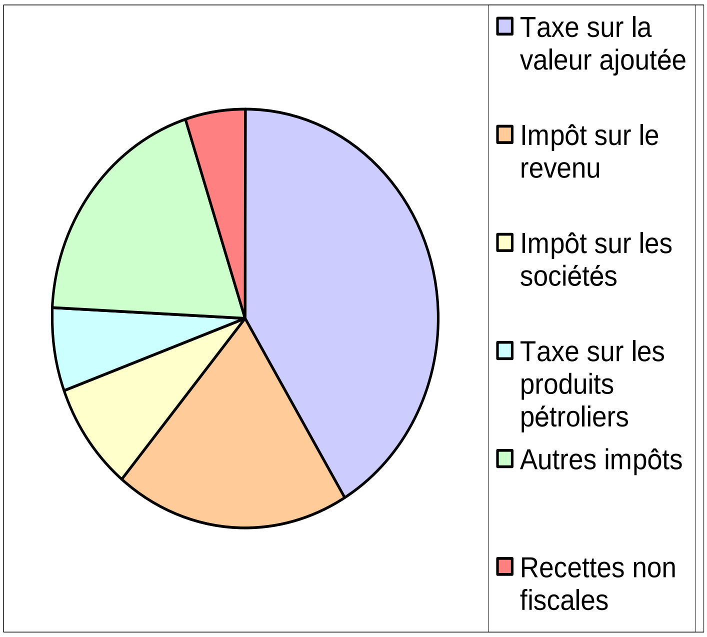
Titre : Tableau n° 4 : La répartition, les fréquences et les écarts (θ_i) des recettes de l'Etat suivant leurs natures.



Titre : Graphique n° 1 : Diagramme en tuyaux d'orgue : Représentation des effectifs (la somme des recettes de l'Etat) suivant leurs natures.

Interprétation : Les recettes de l'Etat pendant l'année « N » sont principalement constitués par la taxe sur la valeur ajoutée. Par contre, les recettes non fiscales y interviennent le moins.

Pour pouvoir représenter le graphique en secteurs circulaire, nous ajoutons au tableau n°1 la colonne n° 4 :



Titre : Graphique n° 2 : Diagramme en secteurs circulaires : Représentation des effectifs (la somme des recettes de l'Etat) suivant leurs natures.

Interprétation : La majorité des recettes de l'Etat pendant l'année « N » est principalement constituée par la taxe sur la valeur ajoutée. Par contre, la minorité est formée par les recettes non fiscales.

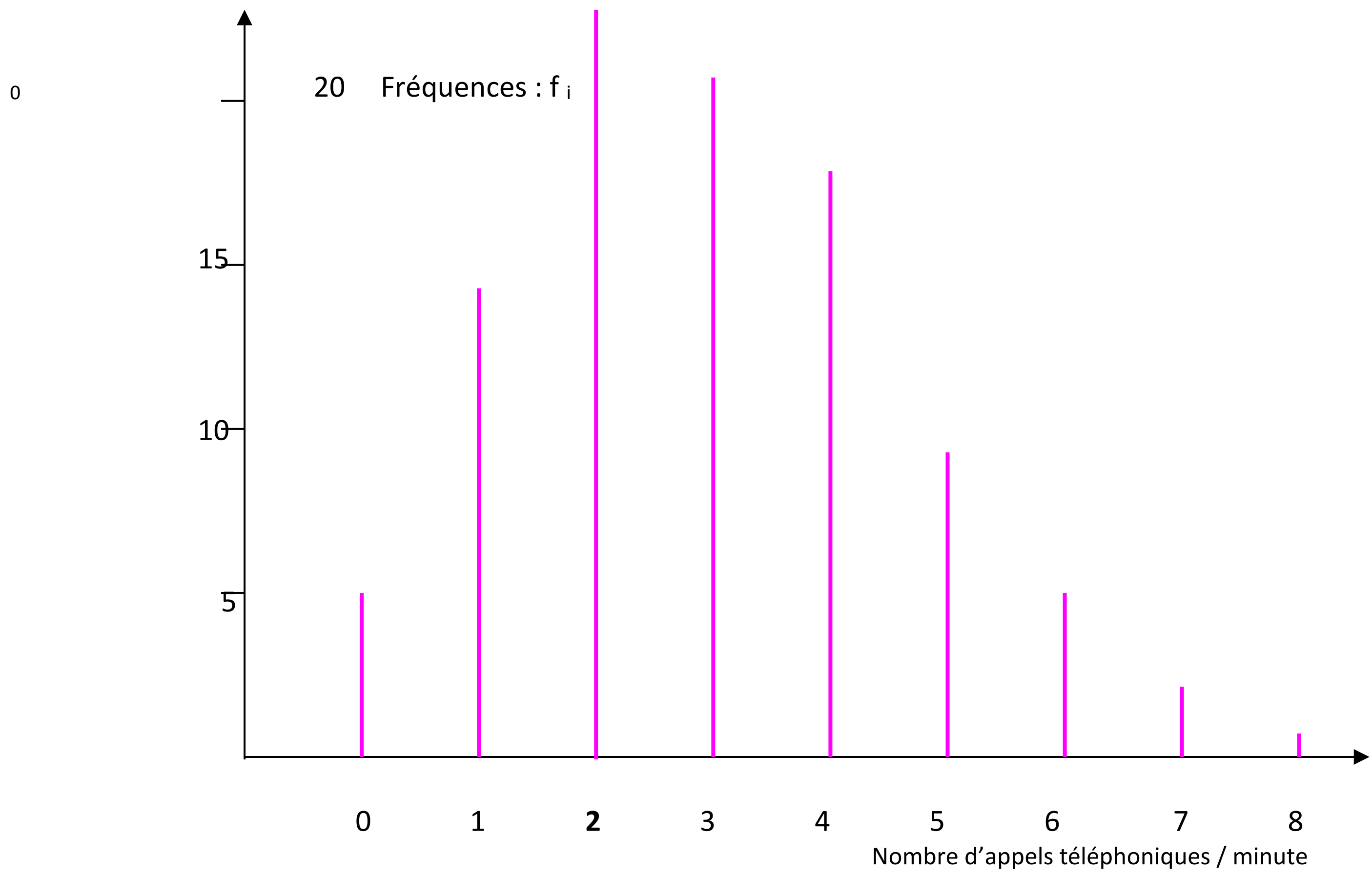
<u>Exercice n° 2 :</u>

1) La population est représentée à travers 1800 minutes. C'est le nombre de minutes compté entre 11 heures et midi chaque jour, pendant 30 jours. Le caractère est le nombre d'appels téléphoniques : il est quantitatif. Il existe, alors, une variable statistique qui est de nature discrète.

2) Pour pouvoir construire le diagramme en bâtons de la distribution des fréquences, il est nécessaire d'ajouter la colonne n°3 au tableau n°2 :

(1) Nombre d'appels téléphoniques par minute	(2) Nombre de minutes	(3) Fréquences f_i (%)	(4) FC↑ (%)
0	93	5,167	0
1	261	14,5	5,167
2	416	23,111	19,667
3	393	21,833	42,778
4	308	17,111	64,611
5	174	9,667	81,722
6	93	5,167	91,389
7	42	2,333	96,556
8	20	1,111	98,889 <100
Total	1800	100	

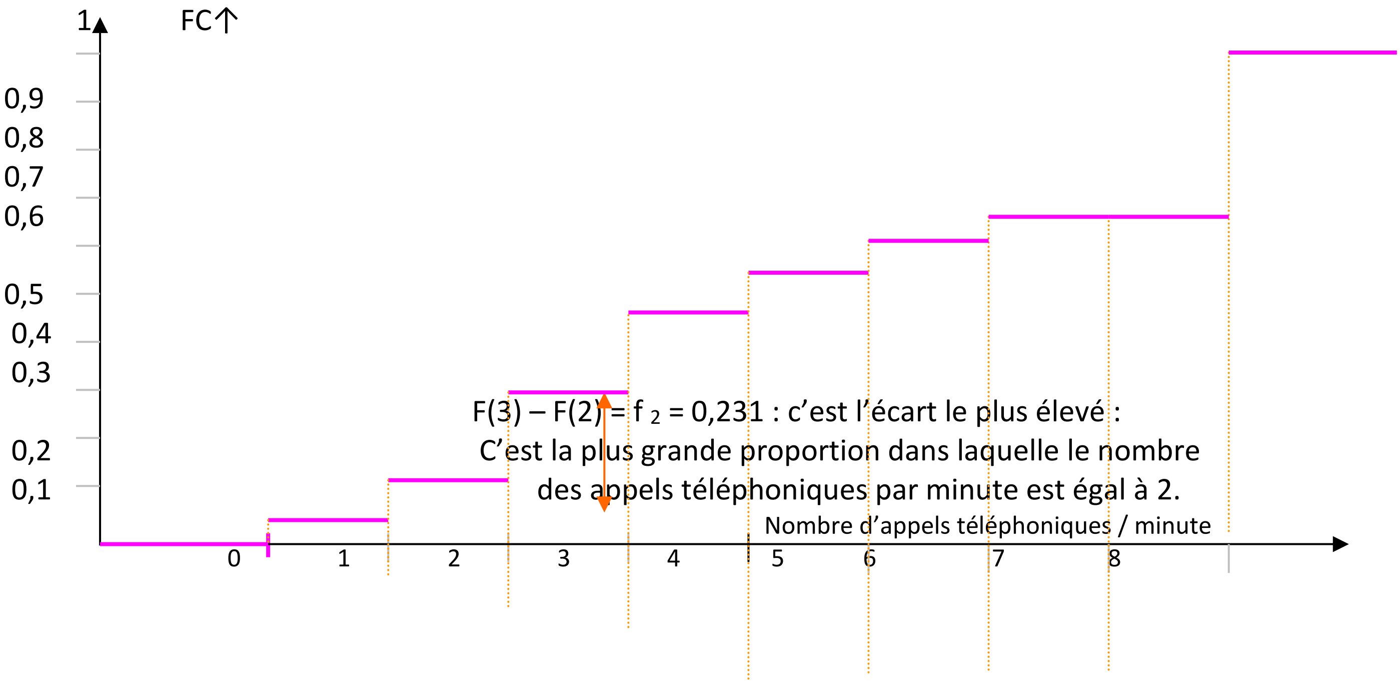
Titre : Tableau n° 5 : Répartition, fréquences et fréquences cumulées croissantes des 1800 minutes suivant le nombre d'appels téléphoniques par minute.



Titre : Graphique n° 3 : Diagramme en bâtons : Représentation des fréquences des 1800 minutes suivant le nombre d'appels téléphoniques par minute.

Interprétation : La fréquence la plus élevée (23,1%) correspond à un nombre d'appels téléphoniques par minute égal à 2. Pour décider du nombre de postes à ouvrir pour le poste de réceptionniste d'appels téléphoniques, et en prenant compte que les trois pourcentages les plus élevés (très proches) du nombre des minutes (23%, 21% et 17 %) correspondent respectivement à 2, 3 et 4 appels téléphoniques par minute, nous mène à conseiller l'entreprise d'employer en moyenne 3 réceptionnistes.

3) Pour pouvoir construire le diagramme en escaliers de la distribution des fréquences cumulées croissantes, il est nécessaire d'ajouter la colonne n°4 au tableau n°2 :



Titre : Graphique 4 : Représentation des FC ↑ : diagramme en escalier.

Même interprétation que le graphique précédent.

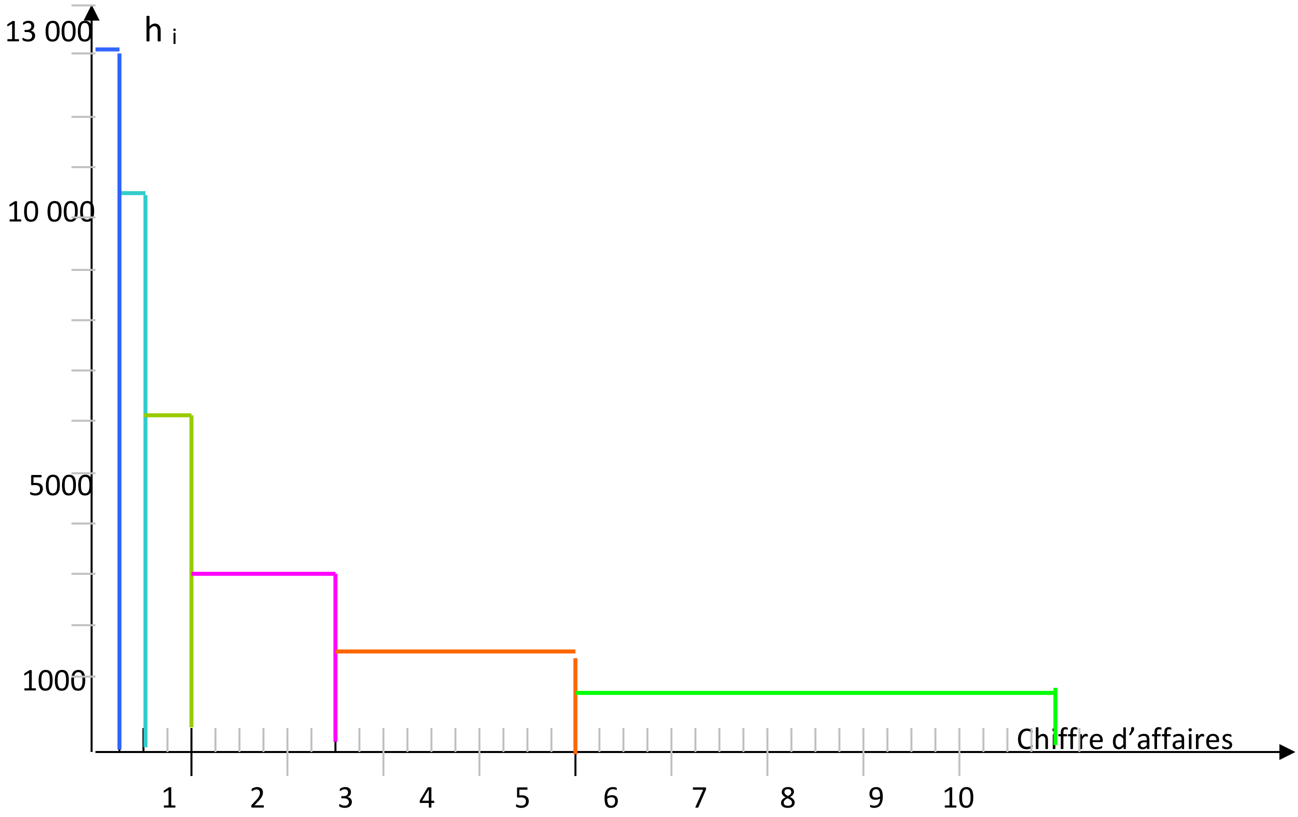
<u>Exercice n°3 :</u>

1) La population est formée de 67 640 entreprises adhérentes à la fédération nationale de la réparation et du commerce de l'automobile. Le caractère est le chiffre d'affaires de ces entreprises : il est quantitatif. Il existe, alors, une variable statistique et elle est de nature continue.

2) Pour pouvoir construire l'histogramme de la distribution des effectifs des entreprises suivant leurs chiffres d'affaires, il est nécessaire d'ajouter les colonnes n° 3 et 4 au tableau n° 3 :

(1) Chiffre d'affaires en million de D	(2) Nombre d'entreprises	(3) a_i	(4) $h_i = n_i^C = \alpha\, n_i / a_i$	(5) EC↑	(6) EC↓
[0 - 0,25[	13 712	0,25	13 712	13 712	67 640
[0,25 - 0,5[	10 674	0,25	10 674	24 386	53 928
[0,5 - 1[	11 221	0,5	5 610,5	35 607	43 254
[1 - 2,5[	15 496	1,5	2 582,67	51 103	32 033
[2,5 - 5[	10 043	2,5	1 004,3	61 146	16 537
[5 - 10[	3 347	5	167,35	64 493	6 494
[10 - 60[	3 147	50	15,735	67 640	3 147
Total	67 640				

Titre : Tableau n° 6: La répartition corrigée et effectifs cumulés des entreprises suivant leurs chiffres d'affaires.

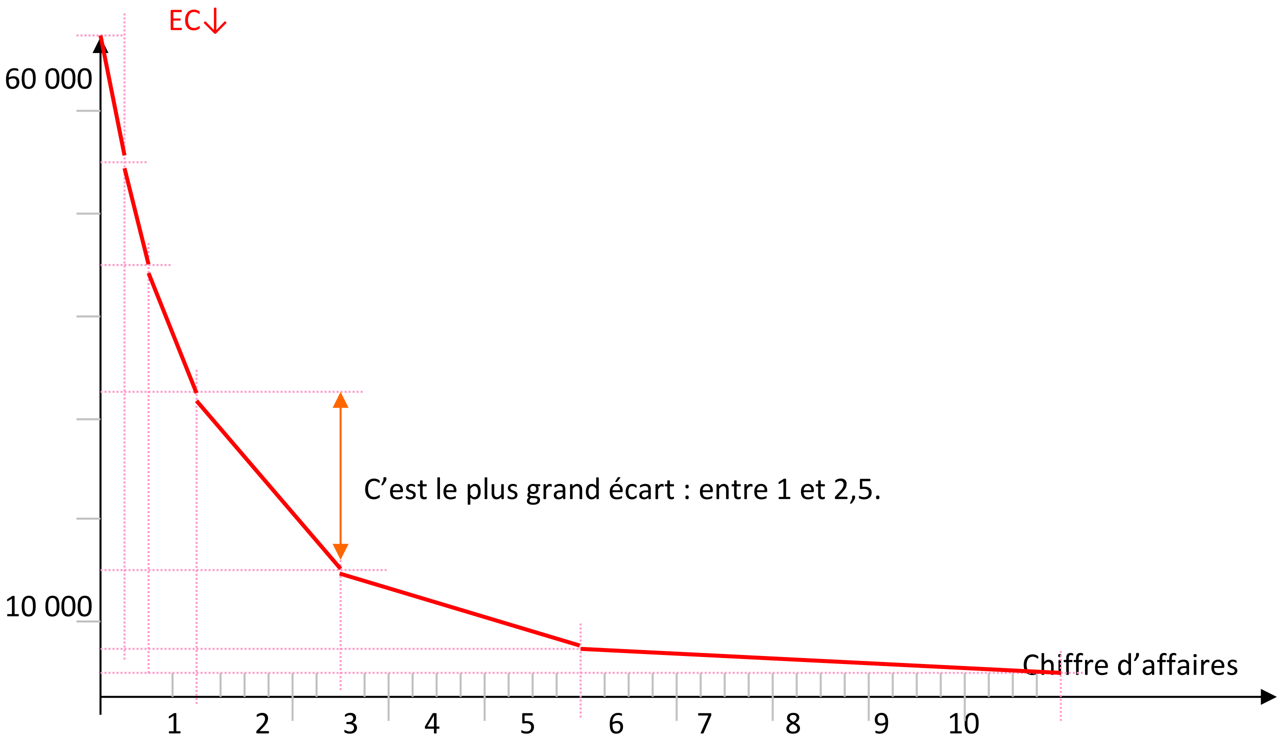


Titre : Graphique 5 : Histogramme : Représentation des effectifs corrigés des entreprises suivant leurs CA

N. B : La dernière classe n'était pas représentée graphiquement puisque son amplitude est très large (= 50) par rapport aux autres amplitudes des autres classes et, aussi, parce que la valeur de « h_i » correspondant (15,7) est très faible par rapport aux autres valeurs « h_i » des autres classes.

__Interprétation :__ la majorité des entreprises, adhérentes à la fédération nationale de la réparation et du commerce de l'automobile, réalisent un chiffre d'affaires inférieur à 0,25 millions de dinars. Elles sont, alors, des petites et moyennes entreprises.

3) Pour pouvoir construire la courbe cumulative de la distribution des EC↓ des entreprises suivant leurs chiffres d'affaires, il est nécessaire d'ajouter les colonnes n° 5 et 6 au tableau n° 3.



Titre : Graphique 6 : Courbe cumulative : Représentation des effectifs cumulés décroissants des entreprises suivant leurs CA.

N. B : La dernière classe n'était pas représentée graphiquement puisque son amplitude est très large (= 50) par rapport aux autres amplitudes des autres classes et, aussi, parce que la valeur de « EC↓ » correspondant (3 147) est très faible par rapport aux autres valeurs « EC↓» des autres classes.

Interprétation : la majorité des entreprises, adhérentes à la fédération nationale de la réparation et du commerce de l'automobile, réalisent un chiffre d'affaires entre 1 et 2,5 millions de dinars. Ce résultat diffère de celui trouvé dans la question précédente puisque dans cette question on a pris en compte les effectifs cumulés ordinaires, or dans la précédente on a pris en compte les effectifs corrigés. Donc, cette différence est justifiée.

Correction Série 2

Exercice n°1 :

a/ Le mode :

Analytiquement : M_0 = 2 appels téléphoniques par min, il correspond à n_3 = 416 : l'effectif le plus élevé dans le tableau statistique.

Graphiquement : le mode correspond au bâton le plus long du diagramme en bâton de la distribution.

Interprétation : la majorité du temps est répartis entre 2 appels téléphoniques par minute. Deux est le nombre d'appels téléphonique par minute le plus observé.

b/ La médiane :

Analytiquement : F (3) = 42,778 < F (M_e) < F (4) = 64,611.

Par convention, on va choisir M_e = 3 appels téléphoniques par minute.

Interprétation : Le temps pour lequel le nombre des appels téléphoniques par minute est < 3 est égal à celui pour lequel le nombre des appels téléphoniques par minute est > 3.

c/ La moyenne arithmétique : On va ajouter au tableau la colonne suivante (2) :

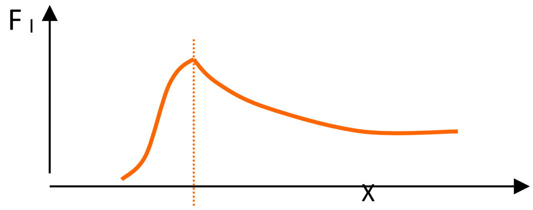
(1) Nombre d'appels téléphoniques par min	(2) $f_i \times x_i$ (%)	(3) $\frac{n_i}{x_i}$
0	0	----
1	14,5	261
2	46,222	208
3	65,499	131
4	68,444	77
5	48,335	34,8
6	31,002	15,5
7	16,331	6
8 et plus	8,888	2,5
Total	299,221	735,8

Titre : Tableau n°13 : Calcul du nombre d'appels téléphoniques moyen par minute.

Alors, $\overline{X} = \sum_{i=1}^{9} f_i$ = 2,992 appels téléphoniques par minute.

Interprétation : Le nombre moyen des appels téléphoniques par minute est de 2,992 $\approx$ 3.

Comparaison : M_0 = 2 < M_e = $\overline{X}$ = 3, donc, la distribution est asymétrique à gauche.



$M_0 \leq M_e \leq \overline{X}$.

Titre : Graphique n°7 : Distribution asymétrique à gauche

d/ La moyenne harmonique : On va ajouter la colonne (3) dans le tableau au-dessus.

$$H = \frac{n}{\sum_{i=1}^{k} n_i \left(\frac{1}{x_i}\right)} = \frac{1800}{735{,}8} = 2{,}446 \text{ appels téléphoniques par minutes.}$$

Comparaison : H = 2,4 < $\overline{X}$ = 3, donc, le résultat est normal et significatif.

Exercice n°2 :

a/ Le mode :

Analytiquement : La classe modale [$e_{i-1} - e_i$ [= [0 – 0,25 [, correspondant à h_i = 13 712 le plus élevé dans le tableau statistique (h_{i+1} = 10 674 et h_{i-1} = 0).

$$M_0 = \frac{e_{i-1}(h_i - h_{i+1}) + e_i(h_i - h_{i-1})}{(h_i - h_{i+1}) + (h_i - h_{i-1})} = \frac{0 \times (13712 - 10674) + 0{,}25(13712 - 0)}{(13712 - 10674) + (13712 - 0)} = 0{,}204 \text{ millions d.}$$

Interprétation : la majorité des entreprises ont un chiffre d'affaires égal à 0,2 millions de dinars. Ou bien, 0,2 est le chiffre d'affaires le plus observé.

b/ La médiane : n = 67 640 alors, n / 2 = 33 820 entreprises.

Analytiquement : La classe médiane est = [$e_{i-1} - e_i$ [= [0,5 – 1 [, on a :

F (0,5) = 24 386 < E (M_e) = 33 820 < F (1) = 35 607.

$$\text{Donc, } M_e = e_{i-1} + \frac{e_i - e_{i-1}}{F(e_i) - F(e_{i-1})}\left(n/2 - F(e_{i-1})\right) = 0{,}5 + \frac{1 - 0{,}5}{35607 - 24386} \times (33820 - 24386)$$

= 0,92 millions de dinars.

Interprétation : La moitié des entreprises ont un chiffre d'affaires < 0,92 millions de dinars ; l'autre moitié des entreprises ont un chiffre d'affaires > 0,92 millions de dinars.

c/ La moyenne arithmétique : on va ajouter les colonnes (1) et (2) au tableau :

Chiffre d'affaire (Millions D)	c_i : (1)	$n_i \times c_i$: (2)	n_i / c_i : (3)
[0 – 0,25 [	0,125	1 714	109 696
[0,25 – 0,5 [	0,375	4 002,75	28 464
[0,5 – 1 [	0,75	8 415,75	14 961,333
[1 – 2,5 [	1,75	27 118	8 854,857
[2,5 – 5 [	3,75	37 661,25	2 678,133
[5 – 10 [	7,5	25 102,5	446,267
[10 – 60 [	35	110 145	89,914
Total		214 159,25	165 190,504

Titre : Tableau n° 14: Calcul des moyennes des chiffres d'affaires des entreprises.

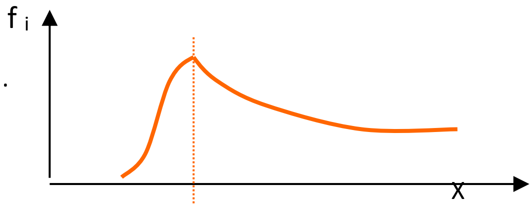
Avec $c_i = \dfrac{e_{i-1}+e_i}{2}$

$$\overline{X} = \frac{\sum_{i=1}^{k} n_i c_i}{n} = \frac{214159{,}25}{67640} = 3{,}16 \text{ millions de dinars.}$$

Interprétation : le chiffre d'affaires moyen par entreprise est de 3,16 millions d.

Comparaison : $M_0 = 0{,}2 < M_e = 0{,}92 < \overline{X} = 3{,}16$; donc, la distribution est asymétrique à gauche.

Titre : Graphique n°8 : Distribution asymétrique à gauche .



$M_0 \leq M_e \leq \overline{X}$.

d/ La moyenne harmonique : On va ajouter la colonne (3) dans le tableau au-dessus.

$$H = \frac{n}{\sum_{i=1}^{k} n_i \left(\frac{1}{c_i}\right)} = \frac{67640}{165190{,}504} = 0{,}409 \text{ millions de dinars.}$$

Comparaison : $H = 2{,}4 < \overline{X} = 3$, donc, le résultat est normal et significatif.

<u>Exercice n°3 :</u>

A/ - 1/

Nb de yaourts consommés / an	n_i	$x_i - x_0$	x'_i	$n_i \times x'_i$	$n_i \times (x'_i)^2$
1000	200	- 534	- 1,074	- 214,8	230,695
1287	130	- 247	- 0,497	- 64,61	32,1
1534	89	0	0	0	0
1756	53	222	0,447	23,691	10,59
2000	67	466	0,938	62,846	58,95
2031	61	**497**	1	61	61
Total	600			- 131,873	393,334

Titre : Tableau n°15 : Calcul de moyenne (du nombre de yaourts consommés par an et par ménage) et variance par changement de variable x'$_i$ = $\frac{x_i - x_0}{a}$, avec x$_0$ = $\frac{x_1 + x_6}{2}$ = $\frac{1000+2031}{2}$ = 1 515,5 $\approx$1 534.

a = 497 : le plus grand (x$_i$ – x$_0$). Donc, x'$_i$ = $\frac{x_i - 1534}{497}$

$$\overline{X}^{/} = \frac{\sum_{i=1}^{k} n_i x_i^{/}}{n} = \frac{-131,873}{600} = -0,22$$

Par application de la propriété de la linéarité, on déduit $\overline{X}$ de $\overline{X}^{/}$, de la manière suivante : Puisque : x'$_i$= $\frac{x_i - x_0}{a}$, alors, x$_i$ =a x'$_i$ +x$_0$ $\Leftrightarrow$ $\overline{X}$ = a $\overline{X}^{/}$ + x$_0$ = 497 × (- 0,22) + 1534 = 1424,765 yaourts. Le nombre de yaourts consommés par an et par ménage est à peu près égal à 1425.

A/ - 2/ V (X') = $\frac{1}{n}\sum_{i=1}^{k} n_i x_i^{/2} - \left(\overline{X}^{/}\right)^2 = \frac{393,334}{600} - (-0,22)^2$ = 0,607 et σ (X') = 0,779.

V(X) = a ^{2} V(X') = (497)^2 × 0,607 = 149 934,463 (yaourt)^2 ,

et σ (X) = $|a|$ σ (X') = $|497|$ × 0,779 = 387,2 yaourts.

B/ -1/

Litres de lait consommés / an	n$_i$	x$_i$ – x$_0$	x'$_i$	n$_i$ × x'$_i$	n$_i$ × (x'$_i$)^2
360	450	- 160	- 2	- 900	1800
520	230	0	0	0	0
543	211	23	0,287	60,662	17,41
567	174	47	0,587	102,225	60,006
600	135	**80**	1	135	135
Total	1200			- 602,112	2012,416

Titre : Tableau n°16 : Calcul de moyenne (du nombre de litres de lait consommés par an et par ménage) et variance par changement de variable

x'$_i$ = $\frac{x_i - x_0}{a}$, avec x$_0$ = $\frac{x_1 + x_5}{2}$ = $\frac{360+600}{2}$ = 480 $\approx$520.

a = 80 : le plus grand (x$_i$ – x$_0$). Donc, x'$_i$ = $\frac{x_i - 520}{80}$

$$\overline{X}^{/} = \frac{\sum_{i=1}^{k} n_i x_i^{/}}{n} = \frac{-602,112}{1200} = -0,5$$

Par application de la propriété de la linéarité, on déduit $\overline{X}$ de $\overline{X}^{/}$, de la manière suivante : Puisque : x'$_i$= $\frac{x_i - x_0}{a}$, alors, x$_i$ =a x'$_i$ +x$_0$ $\Leftrightarrow$ $\overline{X}$ = a $\overline{X}^{/}$ + x$_0$ = 80 × (- 0,5) + 520 = 479,92 litres de lait. Le nombre moyen de litres de lait consommés par an et par ménage est à peu près égal à 480 litres.

V (X') = $\frac{1}{n}\sum_{i=1}^{k} n_i x_i^{/2} - \left(\overline{X}^{/}\right)^2 = \frac{2012,416}{1200} - (-0,5)^2$ = 1,427 et σ (X') = 1,194.

$V(X) = a^2 V(X') = (80)^2 \times 1{,}427 = 9\,132{,}8$ (litres de lait)^2,

et $\sigma(X) = |a|\,\sigma(X') = |80| \times 1{,}194 = 95{,}566$ litres de lait.

La deuxième distribution présente une dispersion plus faible (95,566 < 387,2), donc les ménages (consommateurs de lait) sont plus homogènes que les consommateurs de yaourts.

B/ - 2/ La variance totale de l'union des deux distributions :

P_1 : les ménages consommateurs de yaourts ; et P_2 : les ménages consommateurs de lait.

Leurs effectifs respectifs $n_1 = 600$ et $n_2 = 1200$.

Leurs moyennes arithmétiques respectifs $\overline{X}_1 = 1424{,}765$ et $\overline{X}_2 = 479{,}92$. Alors, la population totale :

$P = P_1 \cup P_2$, d'effectif total $n = n_1 + n_2 = 1800$, et ayant pour moyenne arithmétique totale :

$$\overline{X} = \frac{n_1 \overline{X}_1}{n} + \frac{n_2 \overline{X}_2}{n} = \frac{(600 \times 1424{,}765) + (1200 \times 479{,}92)}{1800} = 794{,}86.$$

et, possède la variance totale suivante :

$$V(X) = \left[\frac{1}{n} \sum_{i=1}^{2} n_i \times V(X_i) \right] + \left[\frac{1}{n} \sum_{i=1}^{2} n_i \times \left(\overline{X}_i\right)^2 - \left(\overline{X}\right)^2 \right]$$

$$= \frac{(600 \times 149934{,}463) + (1200 \times 9132{,}8)}{1800} + \frac{\left(600(1424{,}765)^2\right) + \left(1200(479{,}92)^2\right)}{1800} - (794{,}86)^2$$

$= 56\,066{,}687 + 830\,200{,}572 - 631\,802{,}419 = 254\,464{,}839.$

<u>Exercice n° 4 :</u>

A/ - 1/

Superficie en km^2	n_i	a_i	c_i	$c_i - c_0$	c'_i	$n_i \times c'_i$	$n_i \times (c'_i)^2$
[0 – 200 [	175	200	100	- 350	- 3,5	-612,5	2143,75
[200 – 400 [	280	200	300	- 150	- 1,5	-420	630
[400 – 500 [	300	**100**	450	0	0	0	0
[500 – 600 [	255	100	550	100	1	255	255
[600 – 800 [	190	200	700	250	2,5	475	1187,5
Total	1200					-302,5	4216,25

Titre : Tableau n°17 : Calcul de moyenne (de la superficie par km^2 et par exploitation) et variance par changement de variable

$c'_i = \dfrac{C_i - C_0}{a}$, avec $c_0 = \dfrac{C_1 + C_5}{2} = \dfrac{100 + 700}{2} = 400 \approx 450.$

a = 100 : le plus petit a_i .Donc, $c'_i = \dfrac{C_i - 450}{100}$

$$\overline{X}' = \frac{\sum_{i=1}^{k} n_i c'_i}{n} = \frac{-302,5}{1200} = -0,252$$

Par application de la propriété de la linéarité, on déduit $\overline{X}$ de $\overline{X}'$, de la manière suivante :

Puisque : $c'_i = \frac{C_i - C_0}{a}$, alors, $c_i = a\, c'_i + c_0 \Leftrightarrow \overline{X} = a\, \overline{X}' + c_0 = 100 \times (-0,252) + 450 = 424,791$ km². La superficie moyenne par exploitation est égale à 424,8 km².

A/ - 2/ $V(X') = \frac{1}{n} \sum_{i=1}^{k} n_i {c'_i}^2 - (\overline{X}')^2 = \frac{4216,25}{1200} - (-0,252)^2 = 3,449$ et $\sigma(X') = 1,857$.

$V(X) = a^2 V(X') = (100)^2 \times 3,449 = 34\ 490$ (km)^4,

et $\sigma(X) = |a|\ \sigma(X') = |100| \times 1,857 = 185,715$ km².

B/ - 1/

Kg d'engrais par km²	n_i	a_i	c_i	$c_i - c_0$	c'_i	$n_i \times c'_i$	$n_i \times (c'_i)^2$
[20 – 120 [	120	100	70	- 115	- 3,833	-459,96	1763,027
[120 – 150 [	205	**30**	135	- 50	- 1,667	-341,735	569, 672
[150 – 220 [	287	70	185	0	0	0	0
[220 – 260 [	343	40	240	55	1,833	628,719	1152,442
[260 – 300 [	45	40	280	95	3,167	142,515	451,345
Total	1000					-30,461	3936,486

Titre : Tableau n° 18: Calcul de moyenne (de la consommation d'engrais en kg par km² et par exploitation) et variance par changement de variable

$c'_i = \frac{C_i - C_0}{a}$, avec $c_0 = \frac{C_1 + C_5}{2} = \frac{70+280}{2} = 175 \approx 185$.

a = 30 : le plus petit a_i .Donc, $c'_i = \frac{C_i - 175}{30}$

$$\overline{X}' = \frac{\sum_{i=1}^{k} n_i c'_i}{n} = \frac{-30,461}{1000} = -0,03$$

Par application de la propriété de la linéarité, on déduit $\overline{X}$ de $\overline{X}'$, de la manière suivante :

Puisque : $c'_i = \frac{C_i - C_0}{a}$, alors, $c_i = a\, c'_i + c_0 \Leftrightarrow \overline{X} = a\, \overline{X}' + c_0 = 30 \times (-0,03) + 185 = 184,1$ kg / km². Le nombre moyen de kg d'engrais par km² et par exploitation est égal à 165,94.

$V(X') = \frac{1}{n} \sum_{i=1}^{k} n_i {c'_i}^2 - (\overline{X}')^2 = \frac{3936,486}{1000} - (-0,03)^2 = 3,935$ et $\sigma(X') = 1,984$.

$V(X) = a^2 V(X') = (30)^2 \times 3,935 = 3\ 541,5$ (kg / km²)²,

et $\sigma(X) = |a|\ \sigma(X') = |30| \times 1,984 = 59,51$ kg / km².

La deuxième distribution présente une dispersion plus faible (59,51 < 185,715), donc les exploitations concernant les superficies sont plus homogènes que celles concernant les engrais.

B/ - 2/ La variance totale de l'union des deux distributions :

P_1 : les exploitations concernant la superficie ; et P_2 : les exploitations concernant les engrais.

Leurs effectifs respectifs $n_1 = 1200$ et $n_2 = 1000$.

Leurs moyennes arithmétiques respectifs $\overline{X}_1 = 424{,}791$ et $\overline{X}_2 = 184{,}10$. Alors, la population totale :

$P = P_1 \cup P_2$, d'effectif total $n = n_1 + n_2 = 2200$, et ayant pour moyenne arithmétique totale :

$$\overline{X} = \frac{n_1 \overline{X}_1}{n} + \frac{n_2 \overline{X}_2}{n} = \frac{(1200 \times 424{,}791) + (1000 \times 184{,}1)}{2200} = 315{,}386.$$

et, possède la variance totale suivante :

$$V(X) = \left[\frac{1}{n} \sum_{i=1}^{2} n_i \times V(X_i) \right] + \left[\frac{1}{n} \sum_{i=1}^{2} n_i \times \left(\overline{X_i}\right)^2 - \left(\overline{X}\right)^2 \right]$$

$$= \frac{(1200 \times 34490) + (1000 \times 3541{,}5)}{2200} + \frac{\left(1200\left(424{,}791\right)^2\right) + \left(1000\left(184{,}1\right)^2\right)}{2200} - (315{,}386)^2$$

$= 20\,422{,}5 + 113831{,}674 - 99468{,}329 = 34\,785{,}785.$

Exercice n° 5 :

1) Pour le salaire des femmes : $e = x_{max} - x_{min} = 100\,000 - 25\,000 = 75\,000^{D}$.

Ce paramètre ne donne pas un résultat précis concernant la dispersion de la distribution puisqu'il ne prend en compte que les valeurs extrêmes de la variable.

2) Les quartiles de la distribution des salaires des femmes :

Salaire annuel (× 1000)	Nb de personnel (n_i)	f_i	FC ↑
[25 – 30 [	960	0,08	**0,08**
[30 – **35** [	1704	0,142	**0,222**
[35 – **40** [	2640	0,22	**0,442**
[40 – **45** [	2196	0,183	**0,625**
[45 – **55** [	2808	0,234	**0,859**
[55 – **65** [	996	0,083	0,**942**
[65 – 85 [	516	0,043	0,985
[85 – 100 [	180	0,015	1
Total	12 000	1	___

Titre : Tableau n°19 : effectifs, fréquences et fréquences cumulées croissante du personnel suivant leurs salaires annuels.

Q_1 ? tel que $F(Q_1) = 0,25$ avec : $\begin{cases} 35 \rightarrow 0,222 \\ Q_1 \rightarrow 0,25 \\ 40 \rightarrow 0,442 \end{cases}$

En appliquant la méthode d'interpolation linéaire, on obtient :

$\frac{Q_1 - 35}{40-35} = \frac{0,25-0,222}{0,442-0,222} \Leftrightarrow Q_1 = 35636^D$.

Q_2 ? tel que $F(Q_2) = 0,5$ avec : $\begin{cases} 40 \rightarrow 0,442 \\ Q_2 \rightarrow 0,25 \\ 45 \rightarrow 0,625 \end{cases}$

En appliquant la méthode d'interpolation linéaire, on obtient :

$\frac{Q_{21} - 40}{45-40} = \frac{0,5-0,442}{0,625-0,442} \Leftrightarrow Q_2 = 41585^D$.

Q_3 ? tel que $F(Q_3) = 0,75$ avec : $\begin{cases} 45 \rightarrow 0,625 \\ Q_3 \rightarrow 0,75 \\ 55 \rightarrow 0,859 \end{cases}$

En appliquant la méthode d'interpolation linéaire, on obtient :

$\frac{Q_3 - 45}{55-45} = \frac{0,75-0,625}{0,859-0,625} \Leftrightarrow Q_3 = 50342^D$.

$IIQ = Q_3 - Q_1 = 50,342 - 35,636 = 14706^D$.

3) Pour le salaire des femmes :

D_1= ? tel que $F(D_1) = 0,1$ avec : $\begin{cases} 30 \rightarrow 0,08 \\ D_1 \rightarrow 0,1 \\ 35 \rightarrow 0,222 \end{cases}$

En appliquant la méthode d'interpolation linéaire, on obtient :

$\frac{D_1 - 30}{35-30} = \frac{0,1-0,08}{0,222-0,08} \Leftrightarrow D_1 = 30704^D$.

D_9= ? tel que $F(D_9) = 0,9$ avec : $\begin{cases} 55 \rightarrow 0,859 \\ D_1 \rightarrow 0,9 \\ 65 \rightarrow 0,942 \end{cases}$

En appliquant la méthode d'interpolation linéaire, on obtient :

$\frac{D_9 - 55}{65-55} = \frac{0,9-0,859}{0,942-0,859} \Leftrightarrow D_9 = 59940^D$.

$IID = D_9 - D_1 = 59940 - 30704 = 29\ 236^D$.

4) Comparaison:

Les femmes		*Les hommes*
$e = 75\ 000^D$	<	$e = x_{max} - x_{min} = 105000 - 25000 = 80000^D$.

Apparemment, la dispersion chez les femmes est plus faible.

$IIQ = 14706^D$.	<	$IIQ = 67500-45300 = 22\ 200^D$.

La dispersion des salaires chez les femmes est plus faible que celle entre les hommes.

$IID = 29\ 236^D$.	<	$IID = 83500 - 37250 = 46250^D$.

La dispersion des salaires chez les femmes est plus faible que celle entre les hommes.

Les trois paramètres nous donne le même résultat en matière de dispersion des deux distribution (ce qui n'est pas toujours le cas).

Exercice n°6 :

Cons° / 100 km	n_i	f_i	c_i	$n_i \times c_i$	$n_i \times (c_i)^2$	EC ↑	$\lvert c_i - M_e \rvert$	$\lvert c_i - \overline{X} \rvert$
[5 – 5,5 [	80	0,16	5,25	420	2205	80	0,757	0,819
[5,5 – 5,75 [	81	0,162	5,625	455,706	2563,346	161	0,382	0,444
[5,75 – 6 [	87	0,174	5,875	511,125	3002,859	248	0,132	0,194
[6 – 6,25 [	75	0,15	6,125	459,375	2813,672	323	0,118	0,056
[6,25 – 6,5 [	65	0,13	6,375	414,375	2641,641	388	0,368	0,306
[6,5 – 6,75 [	44	0,088	6,625	291,5	1931,187	432	0,618	0,556
[6,75 – 7 [	28	0,056	6,875	192,5	1323,437	460	0,868	0,806
[7 – 7,5 [	40	0,08	7,25	290	2102,5	500	1,243	1,181
Total	500	1		3034,581	18583,643		4,486	

Cons° / 100 km	$f_i \lvert c_i - M_e \rvert$	$f_i \lvert c_i - \overline{X} \rvert$
[5 – 5,5 [	0,121	0,131
[5,5 – 5,75 [	0,062	0,072
[5,75 – 6 [	0,023	0,034
[6 – 6,25 [	0,018	0,008
[6,25 – 6,5 [	0,048	0,04
[6,5 – 6,75 [	0,054	0,049
[6,75 – 7 [	0,048	0,045
[7 – 7,5 [	0,1	0,094
Total	0,474	0,473

Titre : Tableau n° 20: Calcul des caractéristiques de tendance centrale (M_e et $\overline{X}$) et de dispersion (V(X), CV et EAM) des tests effectués sur les automobiles suivant leur consommation de carburant.

1) $\overline{X} = \dfrac{\sum_{i=1}^{k} n_i c_i}{n} = \dfrac{3034{,}581}{500}$ = 6,069 l / km. La consommation moyenne de carburant de cette nouvelle marque de voiture est de 6,069 l / km et par test.

2) CV = $\sigma_X / \overline{X}$?

V(X) = $\dfrac{1}{n} \sum_{i=1}^{k} n_i c_i^2 - (\overline{X})^2$ = 37,167 – (6,069)^2 = 0,334 (l / km)^2

σ_X = 0,578 l / km : indique la dispersion absolue.

CV = $\sigma_X / \overline{X}$ = 0,578 / 6,069 = 0,095 (sans unité) : indique la dispersion relative.

M_e ? tel que $\begin{cases} 6 \rightarrow 248 \\ M_e \rightarrow 250 \\ 6{,}25 \rightarrow 323 \end{cases}$

Par interpolation linéaire, on obtient : $\dfrac{M_e - 6}{6{,}25 - 6} = \dfrac{250 - 248}{323 - 248} \Leftrightarrow M_e$ = 6,007 litres par 100 km.

EAM / M_e = e $_{(Me)}$ = $\frac{1}{n}\sum_{i=1}^{k} n_i |c_i - M_e|$ = 0,474.

EAM / $\overline{X}$ = e $(\overline{X})$ = $\frac{1}{n}\sum_{i=1}^{k} n_i |c_i - \overline{X}|$ = 0,473.

Exercice n°7 :

1) Les prix annuels des postes de télévision:

Année	Taux de croissance (%)	Prix des poste de TV
85	___	794,159
86	2	810,042
87	2	826,243
88	2	842,768
89	2	859,623
90	2	876,816
91	5	920,657
92	8	994,309
93	8	1073,854
94	- 3,5	1036,269
95	- 3,5	1000

Titre : Tableau n°21 : Prix des TV et leurs taux de croissance annuels.

Suivant les données, on a : $P_{95} = P_{85}$ (1 + t), alors, $P_{85} = P_{95}$ /(1 + t) = 1000/(1 – 0,035)

= 1036,269^D.

Et ainsi de suite.

2) Le taux de croissance du prix des TV pour toute la période de 85 à 95 :

$P_{95} = P_{85} (1 + t) \Leftrightarrow t = 25{,}919\%$.

3) Le taux de croissance annuel moyen du prix des TV pour toute la période de 85 à 95 : (notre cas est assimilé à des observations individuelles : x_i =1+taux correspondant à la période i).

G = $\sqrt[10]{x_1 \times x_2 \times \ldots \times x_n}$ - 1

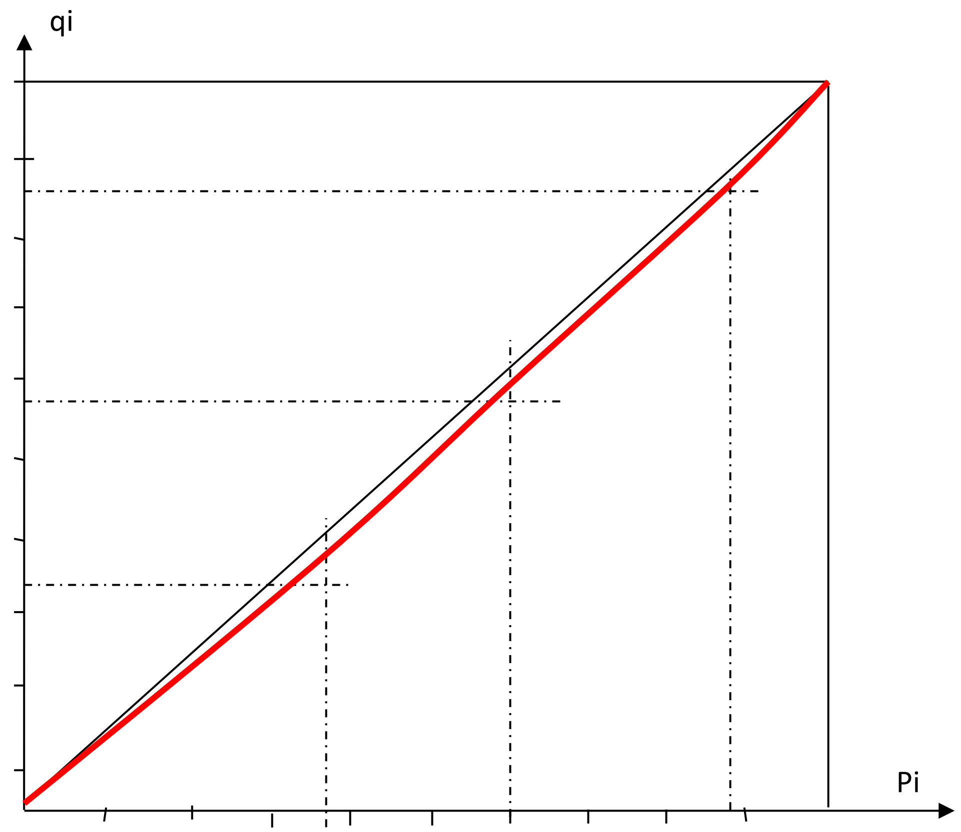
G = $\sqrt[10]{1{,}02 \times 1{,}02 \times 1{,}02 \times 1{,}02 \times 1{,}02 \times 1{,}05 \times 1{,}08 \times 1{,}08 \times (0{,}965) \times (0{,}965)}$ - 1 = 1,023 – 1 = 2,3 %.

Correction Série n° 3

Exercice n° 1 :

1)

Tranches Dépense en D	ni	fi(%)	Pi=FC	ci	nici	$\sum_{j=1}^{n} n_j c_j$	qi (%)	$q_{i-1} + q_i$	$f_i(q_{i-1} + q_i)$
[300-380[	170	37,280	37,280	340	57800	57800	33,545	0,335	0,124
[380-390[	105	23,027	60,307	385	40425	98225	57,00	0,905	0,208
[390-410[	125	27,413	87,72	400	50000	148225	86,02	1,43	0,391
[410-450[	56	12,280	100,00	430	24080	172305	1	1,860	0,226
Total	456	100	------	-----	172305	--------	------	-------	0,949



2) $E = M_l - M_e$

$$\begin{cases} 380 \longrightarrow 37,280 \\ M_e \longrightarrow 50 \\ 390 \longrightarrow 57 \end{cases}$$

$$\frac{Me - 380}{390 - 380} = \frac{50 - 37,280}{60,307 - 37,280}$$

$$\frac{Me - 380}{10} = \frac{12,72}{23,027}$$

Me = 385,523.

$$\begin{cases} 380 \longrightarrow 33,545 \\ M_l \longrightarrow 50 \\ 390 \longrightarrow 57 \end{cases}$$

$$\frac{Ml - 380}{10} = \frac{50 - 33,545}{57 - 33,545} = \frac{16,455}{23,455}$$

M_l = 387,015.

E = 1,492 ; E est petit, la concentration est faible, la répartition est égalitaire.

3) i = 2 Sc = 1 - $\sum_{i=1}^{k}$ fi($q_{i-1} + q_i$) = 0,051 : proche de 0 : il y'a une faible concentration, la répartition des touristes selon les tranches de dépenses par personne est légèrement inégalitaire.

Exercice n° 2 :

1)

Nombre d'années d'existence	ni	fi(%)	Pi=FC ↗	ci	nici	nici^2	$(ci - \overline{X})^3$	$fi(ci - \overline{X})^3$
[0-2[	18	6,27	6,27	1	18	18	-389,017	-24,39
[2-5[	62	21,605	27,875	3,5	217	759,5	-54,872	-11,85
[5-10[	119	41,463	69,338	7,5	892,5	6693,75	0,008	0,0033
[10-12[	88	30,662	100	11	968	10648	50,653	15,53
Total	287	100	------	-----	2095,5	18119,25	------	-20,706

$$\begin{cases} 2 \longrightarrow 6,27\% \\ Q1 \longrightarrow 25\% \\ 5 \longrightarrow 27,875\% \end{cases}$$

$$\frac{Q1 - 2}{3} = \frac{25 - 6,27}{27,875 - 6,27} = \frac{18,73}{21,605}$$

Q1 = 4,6.

$$\begin{cases} 10 \longrightarrow 69,338\% \\ Q3 \longrightarrow 75\% \\ 12 \longrightarrow 100\% \end{cases}$$

$$\frac{Q3 - 10}{5} = \frac{75 - 69,338}{100 - 69,338} = \frac{5,662}{1,442}$$

Q3 = 17,852.

IIQ = Q3 – Q1 = 13,252.

2) $\overline{X} = 7,3$

$V(X) = 63,1088 - (7,3)^2 = 9,818.$

$\sigma(X) = 3,133.$

3)

- Le coefficient de Pearson : $p = 3(\overline{X} - M_e) / \sigma(X)$.

$$\begin{cases} 5 \longrightarrow 27,875\% \\ M_e \longrightarrow 50\% \\ 10 \longrightarrow 69,338\% \end{cases}$$

$$\frac{Me - 5}{5} = \frac{50 - 27,875}{69,338 - 27,875} = \frac{22,125}{41,463}$$

Me = 7,668.

$P = \frac{3(7,3-7,668)}{3,133} = -0,352 < 0$: la distribution est asymétrique négative : $M_0 > Me > \overline{X}$.

- Le coefficient de Fisher :

$\delta = \mu_3 / (\sigma(X))^3 = \sum_{i=1}^{k} f_i (c_i - \overline{X})^3 / (\sigma(X))^3 = -20,7067 / 30,752 = -0,673 < 0,$

alors, μ_3 est < 0 : distribution est asymétrique négative.

- Le coefficient de Yule :

$Y = \frac{[(Q3-Me)-(Me-Q1)]}{[(Q3-Me)+(Me-Q1)]} = \frac{(17,852-7,668)-(7,668-4,6)}{10,184+3,068} = \frac{7,116}{13,252} = 0,537 > 0.$

La distribution est oblique à gauche (étalée vers la droite).

- Le coefficient d'aplatissement de Fisher :

$\beta = \frac{\mu 4}{(\sigma(X))}4 - 3$

$\mu 4 = \sum_{i=1}^{k} f_i (c_i - \overline{X})^4 = 178,057 + 45,049 + 0,000663 + 57,465 = 280,572.$

$\beta = -0,088 < 0$: Le polygone de la variable est moins aplati qu'une courbe en cloche (distribution normale), on a une distribution aigue (leptokurtique).

Exercice n° 3 :

1)

Tranches de dépenses en D	fi (%)	Pi=FC ↗ %	ci	fici	$f_ic_i^2$	$f_i\lvert c_i - Me\rvert$	$f_i\lvert c_i - \overline{X}\rvert$
[2200-3600[	39	39	2900	1131	3279900	466,742	724,035
[3600-5000[	31	70	4300	1333	5731900	63	141,515
[5000-8000[	21	91	6500	1365	8872500	504,677	366,135
[8000-9000[	2,5	93,5	8500	212,5	1806250	110,08	93,587
[9000-13000[	6,5	100	11000	715	7865000	448,71	405,827
Total	100	------	-----	4756,5	27555550	1953,209	1731,1

$\overline{X} = \sum f_i c_i = 4756{,}5$ D

$$\begin{cases} 3600 \longrightarrow 39\% \\ M_e \longrightarrow 50\% \\ 5000 \longrightarrow 70\% \end{cases}$$

$$\frac{Me - 3600}{5000 - 3600} = \frac{50 - 39}{70 - 39} = \frac{11}{31}$$

Me = 4096,774 D.

2)

$$\begin{cases} 2200 \longrightarrow 0 \\ Q1 \longrightarrow 25\% \\ 3600 \longrightarrow 39\% \end{cases}$$

$$\frac{Q1 - 2200}{3600 - 2200} = \frac{25 - 0}{39 - 0}$$

Q1 = 3097,436D.

Q2 = Me = 4096,774

$$\begin{cases} 5000 \longrightarrow 70\% \\ Q3 \longrightarrow 75\% \\ 8000 \longrightarrow 91\% \end{cases}$$

$$\frac{Q3 - 5000}{3000} = \frac{5}{21}$$

Q3 = 5714,286D.

$V(X) = \sum f_ic_i^2 - (\overline{X})^2 = 4\,931\,257{,}75\ D^2$.

$\sigma(X) = 2220{,}643$ D.

$CV = \frac{\sigma(X)}{\overline{X}} = 0{,}467$.

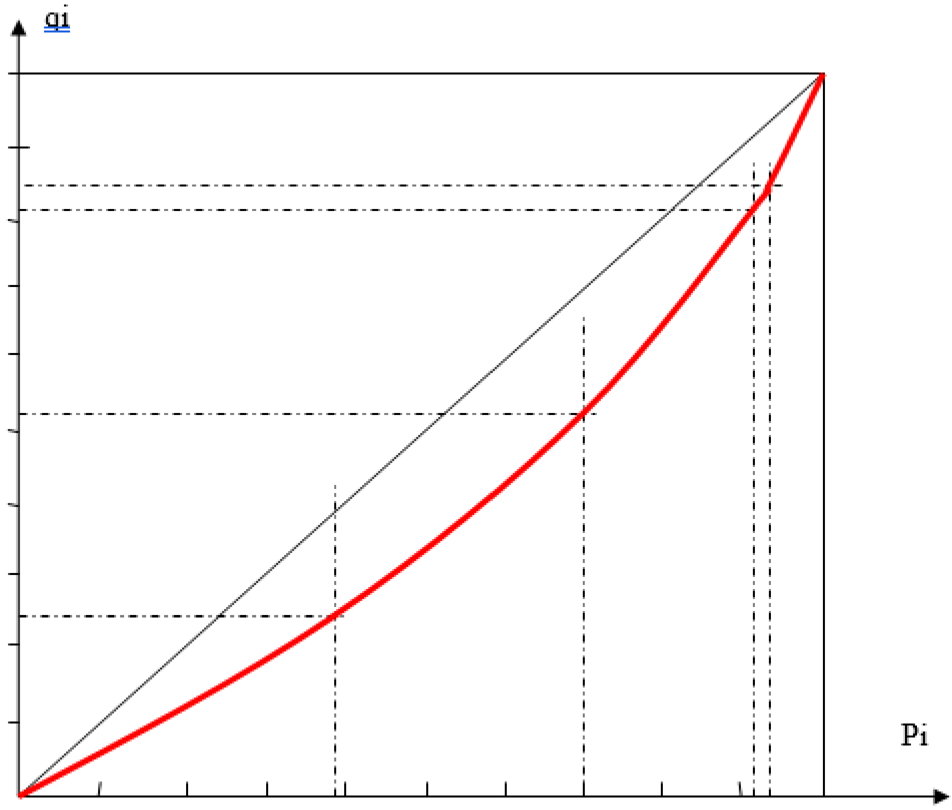
e (Me) = $\frac{1}{n}\sum_{i=1}^{k} n_i \left| c_i - M_e \right|$ = 1593,209.

e ($\overline{X}$) = $\frac{1}{n}\sum_{i=1}^{k} n_i \left| c_i - \overline{X} \right|$ = 1731,1.

3)

Tranches de dépenses en D	fici	∑fj cj	qi (%)	fi(q i-1 + qi)
[2200-3600[	1131	1131	23,778	0,093
[3600-5000[	1333	2464	51,803	0,234
[5000-8000[	1365	3829	80,5	0,278
[8000-9000[	212,5	4041,5	84968	0,041
[9000-13000[	715	4756,5	1	0,12
Total	4756,5	------	---------	0,766

i = 1 - ∑ f_i (q_{i-1} + q_i) = 0,234 : proche de 0, faible concentration, répartition légèrement inégalitaire.



$$\begin{cases} 3600 \longrightarrow 23{,}8\% \\ M_l \longrightarrow 50\% \\ 5000 \longrightarrow 751{,}8\% \end{cases}$$

$$\frac{Ml - 3600}{5000 - 3600} = \frac{50 - 23{,}8}{51{,}8 - 23{,}8} = \frac{26{,}2}{28}$$

Me = 4910.

E = Ml – Me = 813,226.

E est petit, la concentration est faible, la répartition est égalitaire.

Exercice n° 4 :

Fortune des ménages	ni	EC↗	fi	Pi=FC↗	ci	nici	nici ^{2}	$\sum_{j=1}^{i} njcj$	qi
[0-1[	30	30	0,3	0,3	0,5	15	7,5	15	0,006
[1-10[	30	60	0,3	0,6	5,5	165	907,5	180	0,072
[10-50[	30	90	0,3	0,9	30	900	27000	1080	0,435
[50-200[	9	99	0,09	0,99	125	1125	140625	2205	0,889
[200-350[	1	100	0,01	1	275	275	75625	2480	1
Total	100	----	1	------	-----	2480	244165	--------	------

1)

$$\begin{cases} 1 \longrightarrow 30\% \\ M_e \longrightarrow 50\% \\ 10 \longrightarrow 60\% \end{cases}$$

$$\frac{\text{Me} - 1}{9 - 30} = \frac{20}{30}$$

Me = 7 D.

$$\begin{cases} 0 \longrightarrow 0 \\ Q1 \longrightarrow 25\% \\ 1 \longrightarrow 30\% \end{cases}$$

$$\frac{\text{Q1} - 0}{1} = \frac{25 - 0}{30 - 0}$$

Q1 = 0,833D.

$$\begin{cases} 10 \longrightarrow 60\% \\ Q3 \longrightarrow 75\% \\ 50 \longrightarrow 90\% \end{cases}$$

$$\frac{\text{Q3} - 10}{50 - 10} = \frac{75 - 60}{90 - 60}$$

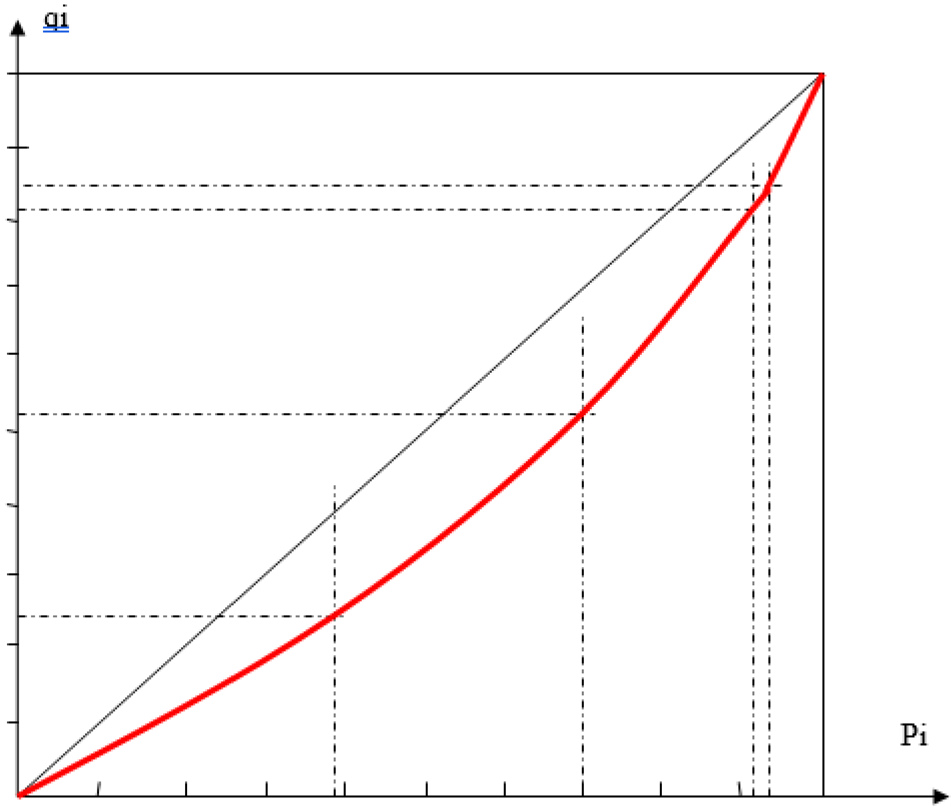
Q3 = 30D.

IIQ = Q3 – Q1 = 29,167.

2) $\overline{X} = \frac{\sum ni\, ci}{n}$ = 2480 / 100 = 24,8

V(X) = $\sum$ ni ci ^{2} – $(\overline{X})$ = 244165 – 615,04 = 1826,61.

σ (X) = 42,739.



Fortune des ménages	qi	$q_{i-1} + q_i$	$f_i(q_{i-1} + q_i)$
[0-1[	0,006	0,006	0,0018
[1-10[	0,072	0,078	0,0234
[10-50[	0,435	0,507	0,152
[50-200[	0,889	1,324	0,119
[200-350[	1	1,889	0,019
Total	------	-------	0,3152

i = 1 - $\sum$ f_i $(q_{i-1} + q_i)$ = 1 – 0,315 = 0,685 : proche de 1, assez élevé concentration, répartition fortement inégalitaire.

3)

$$\begin{cases} 50 \longrightarrow 43{,}5\% \\ M_l \longrightarrow 50\% \\ 200 \longrightarrow 88{,}90\% \end{cases}$$

$$\frac{\mathrm{Ml} - 50}{200 - 50} = \frac{0{,}5 - 0{,}43520}{0{,}889 - 0{,}435} = \frac{0{,}065}{0{,}454}$$

Ml = 92,739 > Me = 7.

Ml – Me = 85,739 > 0 : concentration forte.

Correction Série 4

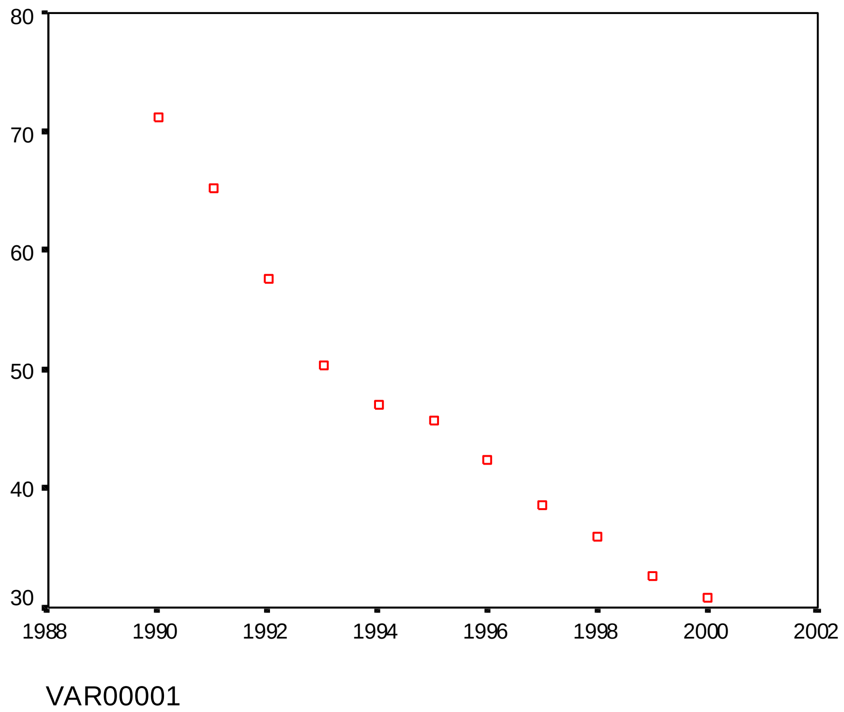
Exercice n°1:

1) Dans ce tableau se présente deux chroniques : l'évolution des effectifs des employés (les colonnes 1 et 2) et celle de la production nette (les colonnes 1 et 3).

(1) Années	(2) Effectifs des employés en milliers	(3) Production nette en millions
1990	71,3	40,1
1991	65,3	35,8
1992	57,6	32,7
1993	50,4	28,4
1994	47,1	25,7
1995	45,8	25,6
1996	42,4	25,1
1997	38,6	24,4
1998	35,9	22,4
1999	32,7	21,1
2000	30,8	20,7

Titre : Tableau n°22 : L'évolution des effectifs des employés et de la production nette.

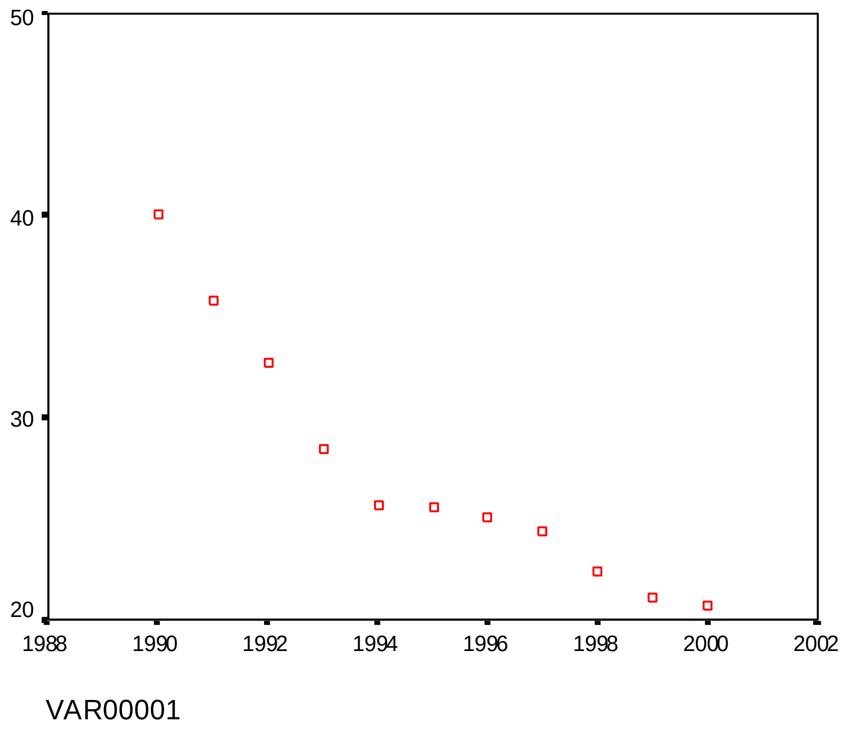
La représentation de la première chronique se fait de la manière suivante:



Titre : Graphique n° 9: L'évolution des effectifs des employés.

Légende : VAR0001 désigne les années. VAR00002 désigne les effectifs des employés (en milliers).

La représentation de la deuxième chronique se fait de la manière suivante :



Titre : Graphique n° 10: L'évolution de la production nette.

Légende :

VAR0001 désigne les années.

VAR00003 désigne la production nette (en millions).

Interprétation : les deux courbes suivent, presque, la même allure (décroissante) avec deux rythmes différents. La première chronique descend beaucoup plus vite, la deuxième tend à stagner. On peut, donc, expliquer de cette manière : suite à la stagnation de la production nette, le chef d'entreprise a diminué les effectifs de ses employés. Pour confirmer ses résultats, nous devons mieux analyser l'éventuelle relation qui pourrait exister entre ces deux variables.

2) Voir graphique au-dessous.

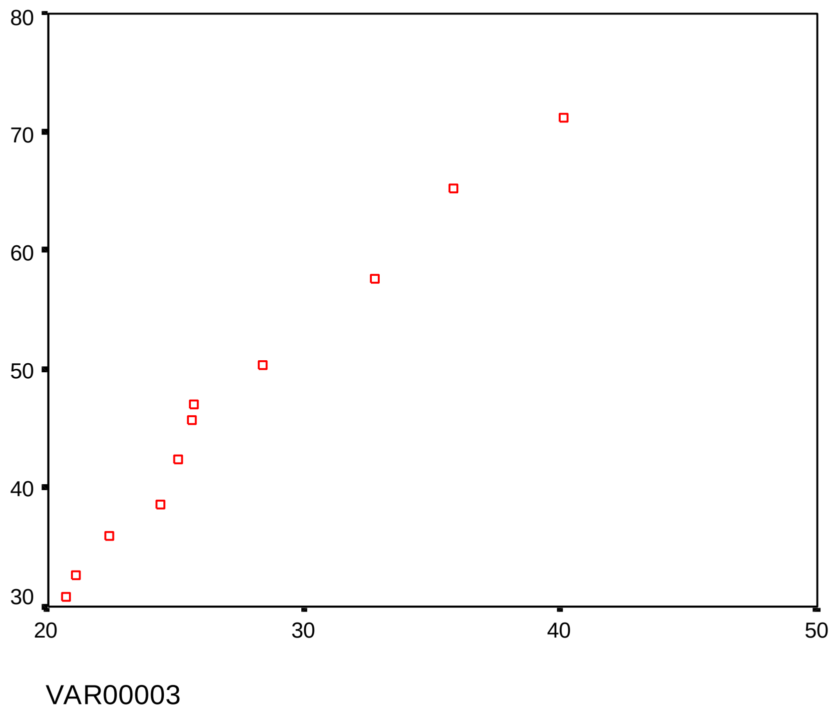
Légende :

VAR0002 désigne les effectifs des employés (en milliers).

VAR00003 désigne la production nette (en millions).

Commentaire : Le nuage de point est très serré, on peut alors faire passer, entre un nombre élevé de points, une droite. Alors, il existe une relation fonctionnelle entre la production nette et les effectifs employés : plus les effectifs employés augmentent, plus la production nette augmente (et vice versa). D'où, notre raisonnement, dans la première question, est confirmé.

Ainsi, on peut conclure qu'il y'a une forte corrélation entre ces deux variables.



Titre : Graphique n° 11: Graphique de corrélation.

Exercice n°2:

La population étudiée est 90 sportifs. Les deux caractères sont : la catégorie socio-professionnelle et la discipline sportive.

1) La distribution marginale des effectifs :

(1) CS \ DS	(2) Athlétisme	(3) Aviron	(4) Cyclisme	(5) Escrime	(6) Yaching	(7) Total
Petits artisans	1	3	6	1	2	13
Employés-ouvriers	8	4	4	0	1	17
Cadres moyens	5	5	2	1	1	14
Cadres supérieurs	1	1	0	3	0	5
Professeurs d'éducation physique	10	2	1	5	0	18
Etudiants	8	5	0	3	7	23
Total	33	20	13	13	11	90

Titre : Tableau n°26 : Les distributions marginales des sportifs selon les deux caractères : la catégorie socio-professionnelle (CS) et la discipline sportive (DS).

La colonne (7) présente la distribution marginale des sportifs selon le caractère « catégorie socio-professionnelle ». La dernière ligne (8) présente la distribution marginale des sportifs selon le caractère « discipline sportive ».

2) La distribution marginale des fréquences : avec $f_{ij} = n_{ij} / n$ (les valeurs sont en pourcentage).

(1) CS \ DS	(2) Athlétisme	(3) Aviron	(4) Cyclisme	(5) Escrime	(6) Yaching	(7) Total
Petits artisans	1,111	3,333	6,667	1,111	2,222	**14,444**
Empoyés-ouvriers	8,89	4,444	4,444	0	1,111	18,889
Cadres moyens	5,556	5,555	2,222	1,111	1,111	15,555
Cadres supérieurs	1,111	1,111	0	3,334	0	5,556
Professeurs d'éducation physique	11,111	2,222	1,111	5,556	0	20
Etudiants	8,89	5,555	0	3,333	7,778	25,556
Total	**36,669**	22,22	14,444	14,445	12,222	100

Titre : Tableau n°27 : Les distributions marginales des fréquences des sportifs selon les deux caractères : la catégorie socio-professionnelle (CS) et la discipline sportive (DS).

La colonne (7) présente la distribution marginale des fréquences des sportifs selon le caractère « catégorie socio-professionnelle ». La dernière ligne (8) présente la distribution marginale des fréquences des sportifs selon le caractère « discipline sportive ».

Interprétation : 14,444% des sportifs sont des petits artisans (la discipline sportive n'intervient plus). 36,669% des sportifs pratique l'athlétisme (sans prendre en compte leurs catégories socio-professionnelles).

3) Les distributions conditionnelles des sportifs selon « la catégorie socio-professionnelle » pour chaque modalité de « discipline sportive » : $f_{i/j} = n_{ij} / n_{.j}$ (les valeurs sont en pourcentage)

(1) CS \ DS	(2) Athlétisme	(3) Aviron	(4) Cyclisme	(5) Escrime	(6) Yaching	(7) Ensemble
Petits artisans	**3,03**	15	40,154	7,692	18,182	14,444
Empoyés-ouvriers	24,243	20	30,769	0	9,091	18,889
Cadres moyens	15,152	25	15,385	7,692	9,091	15,555
Cadres supérieurs	3,03	5	0	23,077	0	5,556
Professeurs d'éducation physique	30,303	10	7,692	38,462	0	20
Etudiants	24,242	25	0	23,077	63,636	25,556
Total	100	100	100	100	100	100
	⇓	⇓	⇓	⇓	⇓	

Distribution1- Distribution2 -Distribution3-Distribution4-Distribution 5

Titre : Tableau n°28 : Les distributions conditionnelles des sportifs selon « la catégorie socio-professionnelle » pour chaque modalité de « discipline sportive »

Interprétation : On a cinq distributions conditionnelles des sportifs selon « la catégorie socio-professionnelle » : C'est le nombre des modalités du caractère « discipline sportive ».

3,03% des sportifs qui pratiquent l'athlétisme sont des petits artisans.

4) Les distributions conditionnelles des sportifs selon la « discipline sportive » pour chaque modalité de « la catégorie socio-professionnelle » : $f_{j/i} = n_{ij} / n_{i.}$ (les valeurs sont en pourcentage)

(1) CS \ DS	(2) Athlétisme	(3) Aviron	(4) Cyclisme	(5) Escrime	(6) Yaching	(7) Total
Petits artisans	**7,692**	23,077	46,154	7,622	15,385	100 ⇒ D 1
Empoyés-ouvriers	47,059	23,529	23,529	0	5,883	100 ⇒ D 2
Cadres moyens	35,714	35,714	14,286	7,143	7,143	100 ⇒ D 3
Cadres supérieurs	20	20	0	60	0	100 ⇒ D 4
Professeurs d'éducation physique	55,555	11,111	5,556	27,778	0	100 ⇒ D 5
Etudiants	34,783	21,739	0	13,043	30,435	100 ⇒ D 6
Ensemble	36,669	22,22	14,444	14,445	12,222	100

Titre : Tableau n°29: Les distributions conditionnelles des sportifs selon la « discipline sportive » pour chaque modalité de « la catégorie socio- professionnelle »

Interprétation : On a six distributions conditionnelles des sportifs selon la « discipline sportive »: C'est le nombre des modalités du caractère « la catégorie socio-professionnelle ».

7,692 % des petits artisans pratiquent l'athlétisme.

Exercice n°3 :

La population étudiée est 40 étudiants. Les deux caractères sont : le genre et l'âge.

1) La distribution marginale des effectifs :

(1) Age \ Genre	(2) Hommes	(3) Femmes	(4) Total
[20 – 22[	5	4	9
[22 – 24[	10	5	15
[24 – 26[	5	5	10
[26 – 28[	4	2	6
Total	24	16	40

Titre : Tableau n°30 : Les distribution marginales des étudiants selon les deux caractères : le genre et l'âge.

La colonne (4) présente la distribution marginale des étudiants selon le caractère « âge ». La dernière ligne (6) présente la distribution marginale des étudiants selon le caractère « genre ».

2) La distribution marginale des fréquences : avec $f_{ij} = n_{ij} / n$

(1) Age \ Genre	(2) Hommes	(3) Femmes	(4) Total
[20 – 22[	0,125	0,1	**0,225**
[22 – 24[	0,25	0,125	0,375
[24 – 26[	0,125	0,125	0,25
[26 – 28[	0,1	0,05	0,15
Total	**0,6**	0,4	1

Titre : Tableau n°31 : Les distributions marginales des fréquences des étudiants selon les deux caractères âge et genre.

La colonne (4) présente la distribution marginale des fréquences des étudiants selon le caractère « âge ». La dernière ligne (6) présente la distribution marginale des fréquences des étudiants selon le caractère « genre ».

Interprétation : 22,5 % des étudiants sont âgés entre 20 et 22 ans (le genre n'intervient plus). 60% des étudiants sont de genre des hommes (sans prendre en compte leurs classes d'âge).

3) Les distributions conditionnelles des étudiants selon « l'âge » pour chaque modalité de la variable « genre » : $f_{i/j} = n_{ij} / n_{.j}$

(1) Age \ Genre	(2) Hommes	(3) Femmes	(4) Ensemble
[20 – 22[	**0,208**	0,25	0,225
[22 – 24[	0,417	0,3125	0,375
[24 – 26[	0,208	0,3125	0,25
[26 – 28[	0,167	0,125	0,15
Total	1	1	1
	$\Downarrow$ Distribution 1	$\Downarrow$ Distribution 2	

Titre : Tableau n°32 : Les distributions conditionnelles des étudiants selon l'âge pour chaque genre.

Interprétation : On a deux distributions conditionnelles des étudiants selon « l'âge » : C'est le nombre des modalités du caractère « genre ».

20,8 % des hommes ont un âge entre 20 et 22 ans.

4) Les distributions conditionnelles des étudiants selon la « le genre » pour chaque modalité « d'âge » : $f_{j/i} = n_{ij} / n_{i.}$

(1) Age \ Genre	(2) Hommes	(3) Femmes	(4) Total
[20 – 22[	0,556	0,444	$1 \Rightarrow$ D 1
[22 – 24[	0,667	0,333	$1 \Rightarrow$ D 2
[24 – 26[	0,5	0,5	$1 \Rightarrow$ D 3
[26 – 28[	0,667	0,333	$1 \Rightarrow$ D 4
Ensemble	0,6	0,4	1

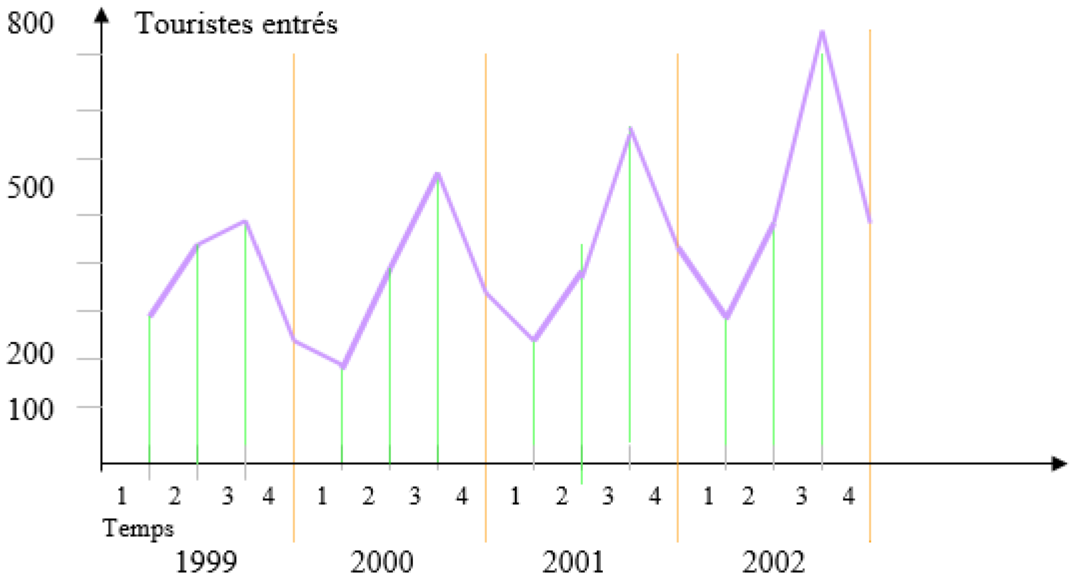
Titre : Tableau n°33 : Les distributions conditionnelles des étudiants selon le genre pour chaque classe d'âge.

<u>Exercice 4 :</u>

1) La population est le nombre total des touristes entrés, trimestriellement.

Les deux caractères sont : la période et le niveau des entrées des touristes.

1) Cette chronique peut être représentée par un diagramme cartésien.



Titre : Diagramme cartésien : L'évolution des touristes entrés en Tunisie.

On a supposé que le nombre de touristes est calculé en fin de chaque période, d'où, on a représenté chaque point en ce moment précis.

On remarque que la pointe de la chronique, en chaque année, se situe dans la troisième période.

Correction Série 5

Question de cours:

- Le caractère quantitatif discret : présente des modalités sous forme de valeurs entières et numériques.
- Le caractère quantitatif continu : présente des modalités sous forme d'intervalle ou de classe. Il peut prendre une infinité de valeurs entre deux valeurs données.

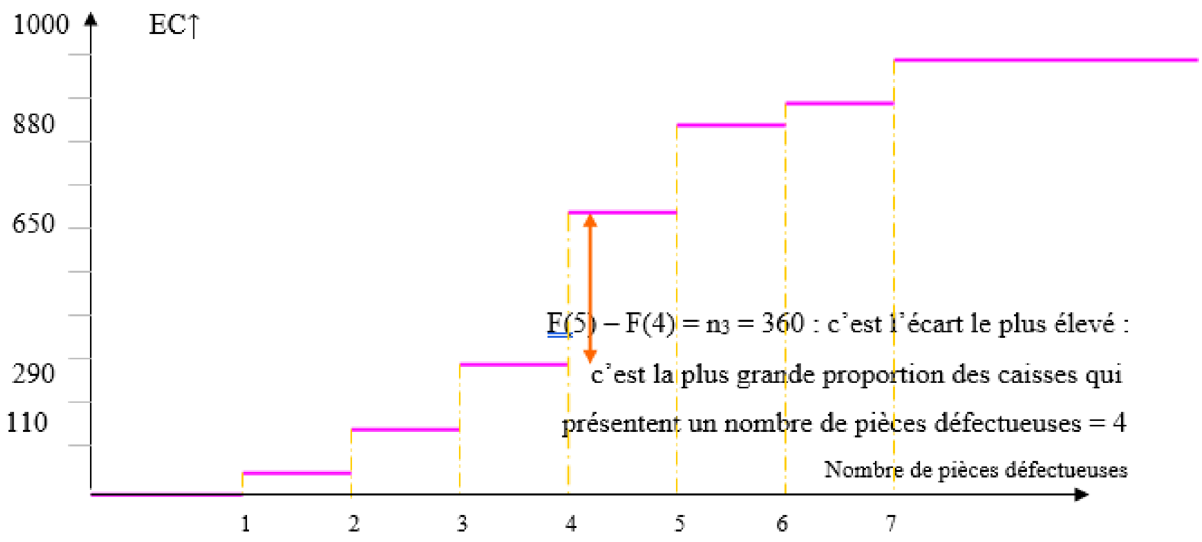
Exercice :

1) La population : 1000 caisses fabriquées par cet industriel.
Le caractère : Le nombre de pièces défectueuses dans chaque caisse. Il est quantitatif discret.

2) Avant de représenter graphiquement les effectifs cumulés croissant, il faut calculer leurs valeurs :

(1)	(2)	(3)
Nombre de pièces défectueuses	*Effectifs*	*Effectifs cumulés croissants*
1	20	0
2	90	20
3	180	110
4	360	290
5	180	650
6	50	830
7	120	880
Total	1000	-----

Titre tableau 2 : La distribution des effectifs cumulés croissant des caisses selon le nombre des pièces défectueuses.



Titre graphique 1 : Diagramme intégral : Courbe en escalier : La représentation des effectifs cumulés croissant des caisses suivant le nombre des pièces défectueuses.

Correction Série 6

1) Le mode : est déterminé par les deux premières colonnes :

(1) Nombre de pannes	(2) Nombre de voitures	(3) E C croissants	(4) $n_i \times x_i$
0	35	0	0
1	30	**35**	30
2	20	**65**	40
3	15	85	45
Total	100	----	115

Titre tableau 4 : Le calcul de caractéristiques de tendance centrale du nombre des voitures selon le nombre de pannes.

Le nombre de voiture le plus élevé est 35 et associé à la variable particulière 0. M_o = 0 panne.

La majorité des voitures sont acceptés lors de la visite : ils présentent 0 pannes.

- Pour déterminer la valeur de la médiane, on ajoute la troisième colonne : la valeur n / 2 n'existe pas dans la colonne 3, on utilise alors l'interpolation linéaire comme suit :

$$\begin{cases} 1 \rightarrow 35 \\ M_e \rightarrow 50 \\ 2 \rightarrow 28 \end{cases}$$

$\frac{M_e - 1}{2-1} = \frac{50-35}{65-35}$; d'où M_e = 1,5 pannes.

La moitié des voitures présente un nombre de pannes supérieur à 1,5 ; l'autre moitié présente un nombre de pannes inférieur à 1,5.

- Pour déterminer la valeur de la moyenne, on ajoute la troisième colonne :

$\overline{X} = \frac{\sum_{i=1}^{4} n_i x_i}{n} = \frac{115}{100}$ = 1,15 pannes par voiture.

Le nombre moyen de pannes par voiture est = 1,15.

Correction Série 7

Questions de cours :

1) La variance totale : Soit des populations P_1 , P_2 ,…, P_k d'effectifs respectifs n_1, n_2,…, n_k, de moyennes arithmétiques respectifs $\overline{X}_1$, $\overline{X}_2$,…, $\overline{X}_k$. Alors, la population totale $P = P_1 \cup P_2 \ldots \cup P_k$, d'effectif total $n= n_1+n_2+\ldots+n_k$, et ayant pour moyenne arithmétique : $\overline{X} = \frac{n_1 \overline{X}_1}{n} + \frac{n_2 \overline{X}_2}{n} + \ldots + \frac{n_k \overline{X}_k}{n}$

et, possède la variance totale suivante :

$$V(X) = \left[\frac{1}{n} \sum_{i=1}^{k} n_i \times V(X_i) \right] + \left[\frac{1}{n} \sum_{i=1}^{k} n_i \times \left(\overline{X_i}\right)^2 - \left(\overline{X}\right)^2 \right]$$

V(X) = variance intra-population (interne dans chaque population à part) + variance inter-population (entre les populations).

2) Les composantes d'une chronique :

- La tendance (T) : est l'allure générale de la chronique à $\pm$ LT.
- La composante cyclique (C) : est périodique à $\pm$ LT.

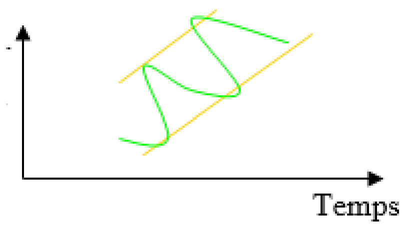
La tendance et la composante cyclique forme le mouvement conjoncturel ou extra saisonnier.

- La composante saisonnière (S) : est un mouvement périodique de durée moyenne ($\leq$ à 1 an).
- La composante accidentelle (E) : tient compte des faits accidentels et imprévisible.

3) Les modèles d'une série temporelle :

- Le modèle additif : pour lequel l'équation de la courbe s'écrit en additionnant les quatre composantes : $y_t = T_t + C_t + S_t + E_t$. Elles sont alors indépendantes.

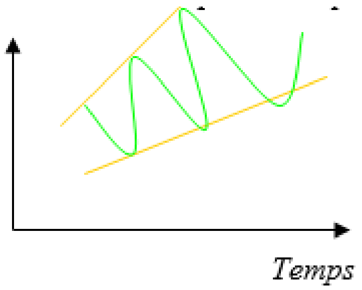
La droite qui lie les creux et celles qui lie les pointes sont parallèles.



Titre : Graphique 2 : Chronique de modèle additif.

- Le modèle multiplicatif : pour lequel l'équation de la courbe s'écrit en multipliant les quatre composantes : $y_t = T_t \times C_t \times S_t \times E_t$. Elles sont alors dépendantes.

La droite qui lie les creux et celles qui lie les pointes sont sécante.



Titre : Graphique 3 : Chronique de modèle multiplicatif.

Exercice :

1) a/ La population : 25 entreprises.

b/ L'unité statistique : une entreprise.

c/ Le caractère : le nombre de représentants commerciales régionales, il est quantitatif.

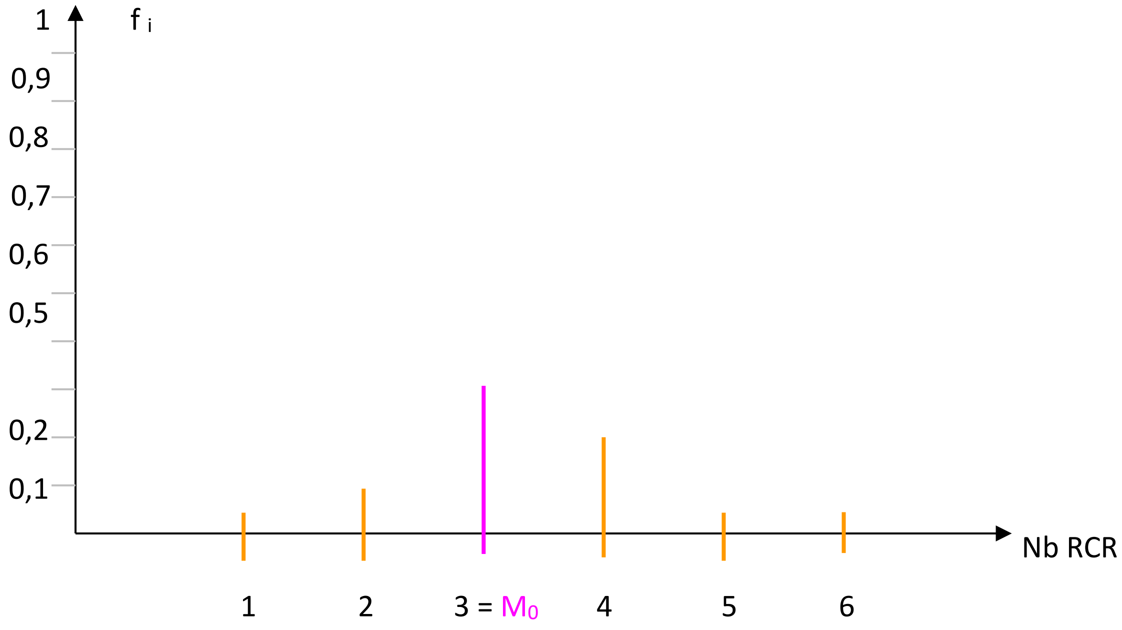
d/ Il existe une variable statistique et elle est discrète.

2) a/ La colonne (1) et (2) :

(1) Nb R C R	(2) n_i	(3) f_i	(4) FC↑	(5) $f_i \times x_i$	(6) $f_i \times (x_i)^2$
1	3	0,12	0	0,12	0,12
2	4	0,16	0,12	0,32	0,64
3	8	0,32	0,28	0,96	2,88
4	6	0,24	0,6	0,96	3,84
5	3	0,12	0,84	0,6	3
6	1	0,04	0,96	0,24	1,44
Total	25	1		3,2	11,92

Titre : Tableau 5 : La répartition des entreprises selon le nombre de représentants commerciaux régionaux et calcul de paramètres.

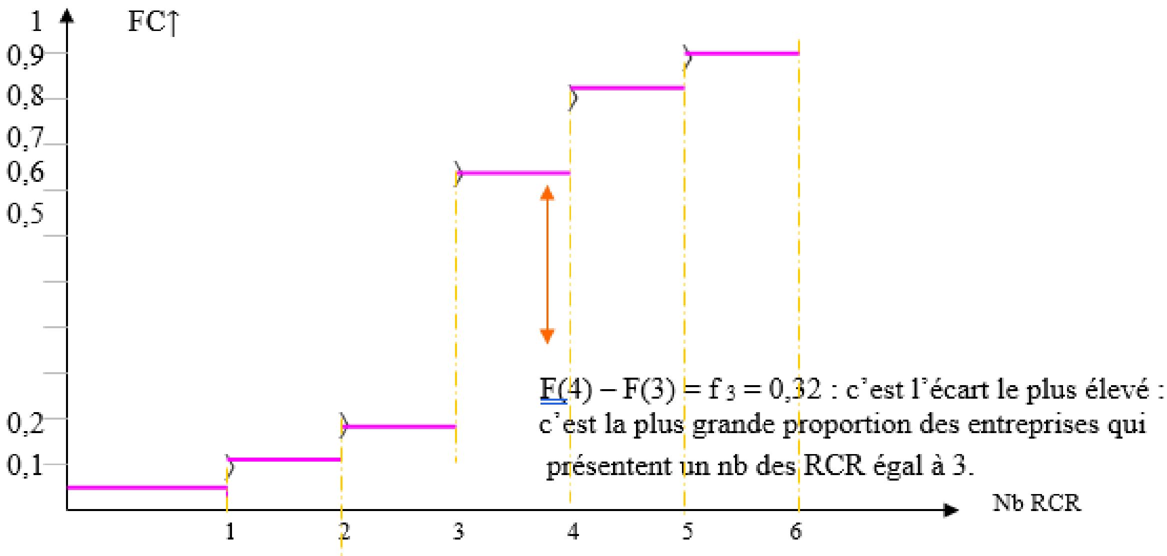
b/ Les fréquences : (colonne n°3) : Diagramme en bâtons.



Titre : Graphique 4 : Diagramme en bâtons : Représentation des fréquences des entreprises selon le nombre de représentants commerciaux régionaux.

Interprétation : La majorité des entreprises (32%) présente un nombre de représentants commercial régional égal à 3.

La représentation des fréquences cumulées croissantes : Courbe en escaliers :



Titre : Graphique 5 : Diagramme en escalier : Représentation des FC ↑ des entreprises selon le nombre de représentants commerciaux régionaux.

3) a/ Le mode:

Analytiquement, M_0 correspond au f_i le plus élevé, il est égal à 3 représentants commerciaux régionaux.

Graphiquement, suivant le diagramme en bâton, M_0 correspond au bâton le plus élevé.

Interprétation : le nombre de RCR le plus observé est 3. Ou bien, la majorité des entreprises ont 3 RCR.

b/ La médiane : M_e = ? tel que $F(M_e) = 0,5$. On a : $\begin{cases} 3 \rightarrow 0,28 \\ M_e \rightarrow 0,5 \\ 4 \rightarrow 0,6 \end{cases}$

Par convention, on prend M_e = 3 RCR.

Interprétation : La moitié des entreprises a un nombre de RCR < 3 ; l'autre moitié a un nombre de RCR > 3 .

c/ La moyenne : $\overline{X} = \dfrac{\sum_{i=1}^{6} n_i x_i}{n} = \sum_{i=1}^{6} f_i \times x_i$ = 3,2 RCR par entreprise.

Interprétation : Le nombre moyen de RCR par entreprise est égal à 3,2.

$M_0 \leq M_e < \overline{X}$, alors la distribution est asymétrique à gauche.

4) a/ $V(X) = \dfrac{1}{n} \sum_{i=1}^{k} n_i \times x_i^2 - \left(\overline{X}\right)^2 = \sum_{i=1}^{k} f_i x_i^2 - \left(\overline{X}\right)^2 = 11,92 - (3,2)^2 = 1,68$ $(RCR)^2$.

$\sigma(X) = \sqrt{V(X)}$ = 1,296 RCR (indique la dispersion absolue de la variable).

b/ $CV = \sigma_X / \overline{X}$ = 1,296 / 3,2 = 0,405 (indique la dispersion relative de la variable).

5) IIQR ?

Q_1 = ? tel que $F(Q_1) = 0,25$. On a : $\begin{cases} 2 \rightarrow 0,12 \\ Q_1 \rightarrow 0,25 \\ 3 \rightarrow 0,28 \end{cases}$

En appliquant l'interpolation linéaire, on obtient : $\dfrac{Q_1 - 2}{3-2} = \dfrac{0,25-0,12}{0,28-0,12}$, d'où, Q_1 = 2,812 RCR : 25% des entreprises ont un nombre de RCR inférieur à 2,812 ans.

Q_2 = ? tel que $F(Q_2) = 0,5$. On a : $\begin{cases} 3 \rightarrow 0,28 \\ Q_2 \rightarrow 0,5 \\ 4 \rightarrow 0,6 \end{cases}$

En appliquant l'interpolation linéaire, on obtient : $\dfrac{Q_2 - 3}{4-3} = \dfrac{0,5-0,28}{0,6-0,28}$, d'où, Q_2 = 3,687 RCR : 50% des entreprises ont un nombre de RCR inférieur à 3,687 RCR.

Q_3 = ? tel que $F(Q_3) = 0,75$. On a : $\begin{cases} 4 \rightarrow 0,16 \\ Q_3 \rightarrow 0,75 \\ 5 \rightarrow 0,84 \end{cases}$

En appliquant l'interpolation linéaire, on obtient : $\frac{Q_3-4}{5-4} = \frac{0,75-0,6}{0,84-0,6}$, d'où, Q_3 = 4,625 RCR : 75% des entreprises ont un nombre de RCR inférieur à 4,625 RCR.

IIQR = IIQR = $\frac{IIQ}{M_e} = \frac{Q_3-Q_1}{Q_2} = \frac{4,625-2,812}{3,687}$ = 0,492.

Correction Série 8

Questions de cours :

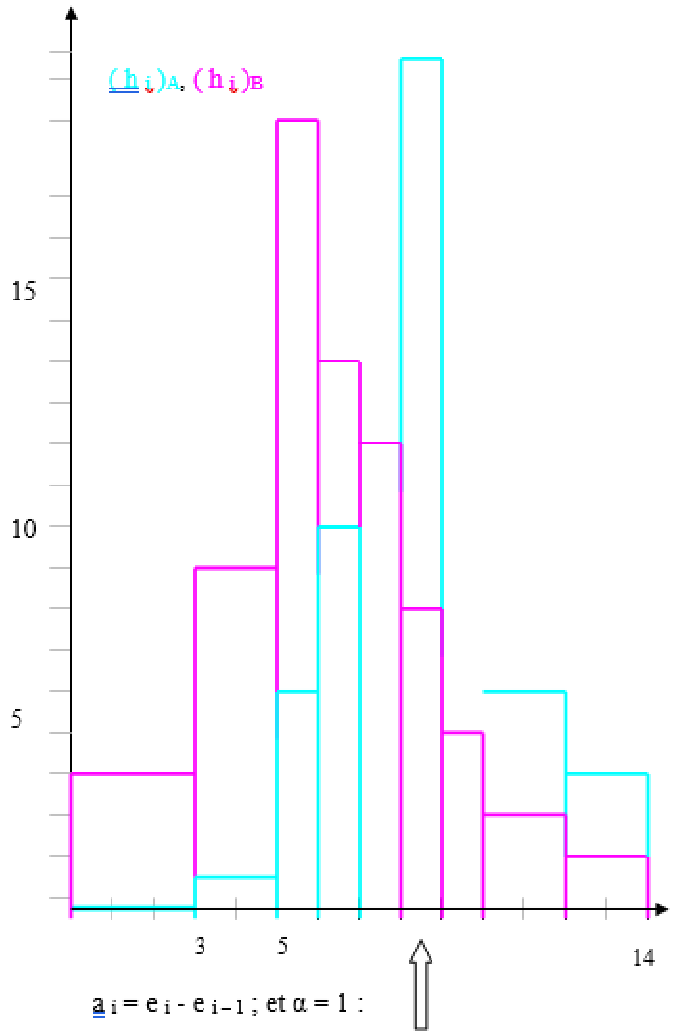
1) $P[a < X < b] = P[X < b] - P[X < a] = F(b) - F(a)$
2) $E(aX + b) = aE(X) + b$.
3) $V(a) = 0$.

Problème :

1) Pour représenter les graphiques des deux distributions, on va construire le tableau suivant :

(1)	(2)		(3)	(4)		(5)	
Temps d'hospitalisation	n_i		a_i	$h_i = n_i^c = \alpha n_i / a_i$		EC ↑	
	TA	TB		TA	TB	TA	TB
[0 – 3 [	2	12	3	0,66	4	2	12
[3 – 5 [	3	18	2	1,5	9	5	30
[5 – 6 [	6	21	1	6	21	11	51
[6 – 7 [	10	14	1	10	14	21	65
[7 – 8 [	18	12	1	18	12	39	77
[8 – 9 [	23	8	1	23	8	62	85
[9 – 10 [	18	5	1	18	5	80	90
[10 – 12 [	12	6	2	6	3	92	96
[12 – 14 [	8	4	2	4	2	100	100
Total	100	100					

Titre : Tableau 7 : Distribution des effectifs simples, corrigés et cumulés des malades selon le temps d'hospitalisation



$a_i = e_i - e_{i-1}$; et $\alpha = 1$:
c'est l'amplitude la plus petite ou la plus fréquente.

Titre : Graphique 6 : histogramme : Représentation des effectifs corrigés des malades selon le temps d'hospitalisation.

Conclusion : Selon les graphiques, on peut dire que le TB est meilleur que le TA car il guérit plus de malades en moins de temps d'hospitalisation.

2) *TA* || *TB*

- Le mode M_0 = ?

La classe modale = [8 – 9 [|| La classe modale = [5 – 6 [

$$M_0 = [\, e_{i-1} (h_i - h_{i+1}) + e_i (h_i - h_{i-1}) \,] \, / \, [\, (h_i - h_{i+1}) + (h_i - h_{i-1}) \,]$$

TA :

$$M_0 = \frac{8 (23 - 18) + 6 (23 - 18)}{(23 - 18) + (23 - 18)} = \mathbf{8,5} \text{ mois} \;>$$

Interprétation :

La majorité des malades qui ont suit le TA ont resté dans l'hôpital 8,5 mois avant de guérir.

TB :

$$M_0 = \frac{5 (21 - 14) + 6 (21 - 9)}{(21 - 14) + (21 - 9)} = \mathbf{5,631} \text{ mois}$$

Interprétation :

La majorité des malades qui ont suit leTB ont resté 5,631mois avant de guérir

- La médiane M_e = ? On ajoute la colonne (5) :

La classe médiane = [8 – 9 [| La classe médiane = [5 – 6 [

$$M_e = e_{i-1} + (e_i - e_{i-1}) \frac{(50 - E(e_{i-1}))}{E(e_i) - E(e_{i-1})}$$

$M_e = 8 + (9-8)\dfrac{(50-39)}{62-39}$ = **8,478** mois **>** $M_e = 5 + (6-5)\dfrac{(50-33)}{51-30}$ = **5,952** mois.

Interprétation :

La moitié des malades qui ont suit le TA, ont guérit pendant un temps inférieur à 8,478 mois. L'autre moitié, ont guérit pendant un temps supérieur à 8,478 mois.

Interprétation :

La moitié des malades qui ont suit le TB, ont guérit pendant un temps inférieur à 5,652 mois. L'autre moitié, ont guérit pendant un temps supérieur à 8,478 mois.

- La moyenne $\overline{X}$ = ?

	(6)	(7)	(8)		(9)	
Temps d'hospitalisation	c_i	c'_i	$n_i \times c'_i$		$n_i \times c'^2_i$	
			TA	TB	TA	TB
[0 – 3 [	1,5	- 5,75	- 11,5	- 69	66,125	396,75
[3 – 5 [	4	- 3,25	- 9,75	- 58,5	31,687	308,948
[5 – 6 [	5,5	- 1,75	- 10,5	- 36,75	18,375	64,312
[6 – 7 [	6,5	- 0,75	- 7,5	- 10,5	5,625	7,875
[7 – 8 [	7,5	0,25	4,5	3	1,125	0,75
[8 – 9 [	8,5	1,25	28,75	10	35,937	12,5
[9 – 10 [	9,5	2,25	40,5	11,25	91,125	25,312
[10 – 12 [	11	3,75	45	22,5	168,75	84,375
[12 – 14 [	13	5,75	46	23	164,5	132,25
Total			125,5	- 105	683,249	1033,072

Titre: Tableau 8 : Calcul de moyenne et de variance par changement de variable.

$c_i = (e_i + e_{i-1}) / 2$

$c'_i = (c_i - c_0) / a$, avec a = 1 (l'amplitude la plus petite) et $c_0 = (c_1 + c_k) / 2 = 7,25$

$$\overline{X'} = 1/n \sum_{i=1}^{9} n_i \, c'_i$$

$\overline{X'} = 125{,}5 / 100 = 1{,}255$ | $\overline{X'} = -105 / 100 = -1{,}05$

Par application de la propriété de linéarité, on a :

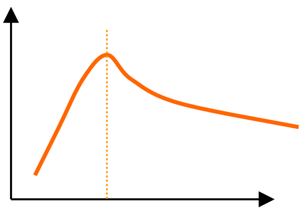
$$\overline{X} = a\,\overline{X'} + c_0$$

$\overline{X} = 1 \times 1{,}255 + 7{,}25 =$ **8, 505** mois. **>** $\overline{X} = 1 \times (-1{,}255) + 7{,}25 =$ **5, 995** mois.

Le temps d'hospitalisation moyen nécessaire à la guérison par malade, qui a suit le TA, est égal à 8, 505 mois.

Comparaison :

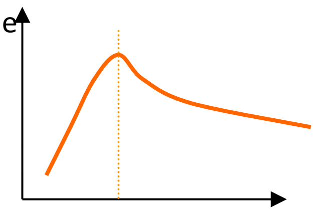
$M_0 \leq M_e < \overline{X}$, donc, la distribution est asymétrique à gauche :



Le temps d'hospitalisation moyen de guérison par malade, qui a suit le TB, est égal à 5, 995 mois.

Comparaison :

$M_0 \leq M_e < \overline{X}$, donc, la distribution est asymétrique à gauche



Titre : Graphique 7 : Distribution asymétrique à gauche.

Selon les caractéristiques de tendance centrale, le TB est meilleur que le TA car le mode, la médiane et la moyenne associés au TA excèdent ceux associés au TB. Donc, la conclusion de la question (1) est confirmée : le TB guérit plus de personnes et plus rapidement que le TA.

2) On ajoute la colonne (9) au tableau de calcul :

$$V(X') = 1/n \sum_{i=1}^{9} n_i \times c'^2_i - (\overline{X'})^2$$

$V(X') = 683{,}249 / 100 - (1{,}255)^2 = 5{,}26$ | $V(X') = 1033{,}072 / 100 - (-1{,}05)^2 = 9{,}228$

$$\sigma(X) = \sqrt{V\langle X \rangle}$$

$\sigma(X') = \sqrt{5{,}26} = 2{,}293$ | $\sigma(X') = \sqrt{9{,}228} = 3{,}038$

Par application à la propriété de linéarité, on obtient :

$$V(X) = a^2\, V(X')$$

$V(X) = 5{,}26\ (\text{mois})^2$ | $V(X) = 9{,}228\ (\text{mois})^2$

$$\sigma(X) = \sqrt{V\langle X\rangle}$$

$\sigma(X) = \sqrt{6{,}26} = 2{,}293$ mois $<$ $\sigma(X) = \sqrt{9{,}228} = 3{,}038$ mois.

Les malades guéris qui ont suit le TA sont moins dispersés que ceux qui ont suit le TB. Les premiers sont, alors plus homogènes que les seconds.

4) Calcul de la variance totale de l'union des deux distributions (les malades qui ont suit les deux traitements). On a, alors :

P_A : $n_A = 100$, $\overline{X}_A = 8{,}505$, $V(X_A) = 5{,}26$

P_B : $n_B = 100$, $\overline{X}_B = 5{,}995$, $V(X_B) = 9{,}228$

$n = n_A + n_B = 200$

$$\overline{X} = \frac{n_A\overline{X}_A + n_B\overline{X}_B}{n} = \frac{100 \times 8{,}505 + 100 \times 5{,}995}{200} = 7{,}25 \text{ mois.}$$

$$V(X) = \frac{1}{n}\sum_{i=1}^{2} n_i\, V(X_i) + \frac{1}{n}\sum_{i=1}^{2} n_i\, \left(\overline{X}\right)^2 - \left(\overline{X}\right)^2$$

$V(X) = 7{,}244 + 54{,}137 – 52{,}562 =$ **8,818** (mois)^2.