STATISTIQUE

Première Leçon

GÉNÉRALITÉS

1. Définition de la statistique. — Le mot « statistique » évoque des ensembles de nombres, groupés en tableaux, traduisant des observations relatives à des faits économiques ou démographiques. Le statisticien s'intéresse, dans une première phase, à l'enregistrement des documents chiffrés et à leur coordination pour « obtenir des rapports numériques sensiblement indépendants des anomalies du hasard et qui dénotent l'existence des causes régulières ».

La statistique est donc la science qui a pour objet l'étude, l'analyse et l'interprétation des observations relatives à un même phénomène.

EXEMPLE. — On peut se proposer d'étudier :

— l'évolution de la population d'un pays de 1936 à 1970;

— la variation de la production de la consommation d'électricité dans une région déterminée, au cours d'une période bien définie;

— la répartition des résidences principales d'après le nombre de pièces et le nombre d'occupants dans la région parisienne.

2. Aperçu historique de la statistique. — On trouve dès l'Antiquité des exemples de *dénombrement* et de *collecte de renseignements :* le recensement de la population guerrière ordonné par l'empereur César. Jusqu'à la Révolution, l'Église entreprend des enquêtes analogues pour collecter des renseignements statistiques. Citons Quetelet (Belge), Galton (Anglais), Charlier (Scandinave), Pearson (Anglais), Neyman (Américain) qui créèrent et développèrent la statistique mathématique.

3. Vocabulaire de la statistique. — ***On appelle population, tout ensemble faisant l'objet d'une étude statistique.***

Les éléments de cette population présentent un trait commun, bien défini, appelé ***caractère.***

On considère par exemple :

— la population des élèves d'une classe ayant obtenu une *moyenne annuelle supérieure à dix* (caractère);

— ou la population des automobiles d'un parc à voitures *immatriculées en France* (caractère).

4. Caractère quantitatif. Caractère qualitatif.

1° *Le caractère d'une population est quantitatif lorsqu'on peut le mesurer.* C'est le cas de l'âge, du poids, de la taille d'individus; de la rémunération mensuelle d'un fonctionnaire; de la consommation d'électricité d'une usine, etc.

Certains d'entre eux ne prennent que des valeurs entières (nombre d'enfants d'une famille, nombre de livrets de la Caisse d'épargne déposés à date fixe). Un tel caractère est dit ***discret*** ou ***discontinu.***

D'autres, au contraire, peuvent prendre des valeurs quelconques dans un intervalle fini ou infini (taille d'un enfant, poids d'un objet). Un tel caractère est dit ***continu.***

2° Le caractère d'une population peut être *qualitatif* : sexe d'une personne, profession, couleur d'une fleur dans une population florale. Un ordre de classement n'est pas imposé, mais on établit une *nomenclature* des catégories; par exemple : la liste des départements pour une étude géographique.

ANALYSE STATISTIQUE

5. Observation des faits. — Les sources d'information nécessaires pour faire l'étude statistique d'une population constituent des éléments de base fondamentaux. On peut obtenir les renseignements concernant *tous* les membres d'une population (recensement du nombre de naissances d'une ville), ou une *partie* de la population (étude de la population active par sexe et âge d'un pays). Dans ce dernier cas, on procède à un ***sondage*** : les unités statistiques ainsi étudiées constituent un ***échantillon.***

De la collecte des observations relatives à un phénomène bien défini, on obtient un ensemble de nombres qui constitue une ***série statistique*** présentée généralement sous forme de *tableaux descriptifs.*

6. Enregistrement des observations. — Supposons qu'on doive définir une statistique portant sur les notes de composition de mathématiques de cent élèves. Les notes figurent dans le tableau ci-dessous et ont été distribuées au fur et à mesure de la correction.

7	6	10	12	14	1	8	9	10	15	14	12	12	13	2	4	6
5	1	9	11	9	6	11	7	4	12	8	13	10	12	12	18	10
9	5	10	6	5	11	12	13	14	8	11	7	15	8	5	13	8
6	11	10	18	12	14	16	6	13	15	11	18	8	19	17	8	15
10	7	11	9	11	9	9	10	13	15	16	9	13	15	17	13	9
9	10	11	7	8	11	16	7	17	10	7	10	10	8	10		

Si certaines caractéristiques apparaissent facilement : peu de notes inférieures à 3 et supérieures à 17, il est difficile de tirer des indications générales de ces données présentées sans ordre. On commencera donc par les *classer* suivant le tableau qui présente le nombre d'élèves n_i ayant obtenu la note x_i.

Notes x_i	n_i	Notes x_i	n_i	Notes x_i	n_i	Notes x_i	n_i
1	2	6	6	11	10	16	3
2	1	7	7	12	8	17	3
3	0	8	9	13	8	18	3
4	2	9	10	14	4	19	1
5	4	10	13	15	6	20	0

Ce tableau présente l'inconvénient d'être volumineux. On peut effectuer un regroupement des valeurs de la variable (note) suivant 5 *classes*. Pour cela on partage l'intervalle [0,20] en 5 intervalles partiels :

$$[0,4[\quad [4,8[\quad [8,12[\quad [12,16[\quad [16,20].$$

Dans ces conditions, le dépouillement se présente suivant le tableau :

Classes des notes	Nombre d'élèves
moins de 4	3
4 à moins de 8	19
8 — 12	42
12 — 16	26
16 à 20	10

CONVENTION. — La limite supérieure de la classe ne fait pas partie de la classe. Ainsi la note 12 est comptée dans la classe de 12 à 16 qui ne comprend que les notes x telles que $12 \leqslant x < 16$; classe notée [12,16[.

7. Tableaux statistiques. — ***On appelle effectif total d'une population le nombre d'éléments de cette population.***

L'effectif d'une valeur ou le *nombre des répétitions* représente le nombre d'unités qui possèdent, dans la population, la valeur correspondante de la variable. On dit aussi *fréquence absolue.*

La fréquence d'une valeur x_i de la variable est le rapport de l'effectif de cette valeur à l'effectif total. On la note $f(x_i)$.

$$f(x_i) = \frac{\text{Effectif de } x_i}{\text{Effectif total}} = \frac{n_i}{n}.$$

On dit aussi fréquence relative.

EXEMPLE. — Dans l'étude précédente, l'effectif de la population d'élèves est 100. L'effectif de la note 13 est 8.

La fréquence de la note 5 est $\frac{4}{100}$ ou 4 %.

On peut compléter le tableau précédent en figurant dans la colonne 3 les fréquences relatives et dans les colonnes suivantes les *résultats cumulés* dont les explications font suite au tableau :

Résultats				Résultats cumulés			
Classe des notes	Effectifs	Fréquences	Fréquences %	Effectifs cumulés croissants	Effectifs cumulés décroiss.	Fréquences cumulées croissantes	Fréquences cumulées décroiss.
moins de 4	3	0,03	3	3	100	0,03	1
[4,8[	19	0,19	19	22	97	0,22	0,97
[8,12[	42	0,42	42	64	78	0,64	0,78
[12,16[	26	0,26	26	90	36	0,90	0,36
[16,20]	10	0,10	10	100	10	1	0,10

8. Interprétation des résultats cumulés. — Dans la colonne 5, figurent les effectifs cumulés croissants. Par définition, *l'effectif cumulé croissant d'une classe est la somme des effectifs de cette classe et des classes antérieures.* Ainsi :

1re ligne : il y a 3 élèves ayant une note inférieure à 4.
2e ligne : il y a $3 + 19 = 22$ élèves ayant une note inférieure à 8.
3e ligne : il y a $22 + 42 = 64$ élèves ayant une note inférieure à 12.
4e ligne : il y a $64 + 26 = 90$ élèves ayant une note inférieure à 16.
5e ligne : 100 élèves ont une note inférieure ou égale à 20.

Dans la colonne 6, figurent les effectifs cumulés décroissants.
L'effectif cumulé décroissant d'une classe est la somme des effectifs de cette classe et des classes postérieures. Ainsi :

5e ligne : 10 élèves ont une note supérieure ou égale à 16.
4e ligne : $10 + 26 = 36$ élèves ont une note supérieure ou égale à 12.
3e ligne : $36 + 42 = 78$ élèves ont une note supérieure ou égale à 8 et ainsi de suite.

Les colonnes relatives aux fréquences cumulées s'en déduisent immédiatement.

9. Série statistique à caractère discontinu. — Présentons une série statistique à *caractère discontinu*, traduisant la répartition des résidences principales d'après le nombre de pièces, à Paris en 1963.

Unité : 1 000 logements

Nombre de pièces	Effectifs	Effectifs cumulés croissants	Effectifs cumulés décroissants
1	392	392	1 169,20
2	403,20	795,20	777,20
3	220,32	1 015,52	374,00
4	78,30	1 093,82	153,68
5	36,88	1 130,70	75,38
6 et plus	38,50	1 169,20	38,50
	1 169,20		

Nous interprétons partiellement ce tableau, en dénombrant :

220 320 logements de 3 pièces;
1015 520 logements comprenant *au plus* 3 pièces;
374 000 logements comprenant *au moins* 3 pièces.

10. Séries chronologiques.

Les séries chronologiques présentent les grandeurs statistiques dans le temps.

EXEMPLE : *Trafic de l'aéroport de Paris-Orly.*

Année	Mouvement d'avions	Passagers (arrivées et départs) *en milliers*
1950	23 241	352,2
1955	48 955	1 306,7
1962	89 235	3 539,2
1963	95 272	4 031,7
1964	99 550	4 451,7

11. Séries doubles. — Dans les séries précédentes, à chaque unité statistique correspondait une seule variable :

— une note (variable) à chaque composition (unité statistique);
— le nombre de pièces (variable) à chaque logement (unité statistique).

On peut observer deux variables sur une même unité statistique. Le résultat des observations sera mentionné dans un tableau à double entrée.

EXEMPLE. — *Répartition des résidences principales d'après le nombre de pièces et le nombre d'occupants à Paris en 1963.*

Unité : 1 000 logements

Nombre d'occupants du logement	Logements composés de ___ pièces d'habitation						Total des logements
	1	2	3	4	5	6 et plus	
1	233,44	129,26	35,96	7,36	2,26	2,52	410,80
2	105,52	150,38	78,52	22,36	7,38	4,30	368,46
3	34,90	75,70	52,44	18,50	8,22	6,12	195,88
4	12,22	38,18	31,86	14,94	8,06	6,94	106,20
5	3,76	10,62	13,46	8,44	5,44	7,06	48,78
6 et plus	2,16	5,06	8,08	6,70	5,32	11,56	38,88
TOTAL	392,00	403,20	220,32	78,30	36,68	38,50	1 169,00

Interprétation des nombres portés dans le tableau :

233 440 logements de 1 pièce sont occupés par 1 personne.
8 220 logements de 5 pièces sont occupés par 3 personnes.
Il existe 403 200 logements de 2 pièces et 38 880 logements contenant 6 personnes et plus.

SYSTÈME DE NOTATIONS.

12. Signe de sommation $\sum$. — Dans les séries statistiques, les valeurs du caractère quantitatif (nº 4) pourront être désignées par les lettres $x_1, x_2, \dots x_p$ et les effectifs correspondants de la population (nº 7) par $n_1, n_2, \dots n_p$.

L'effectif total est $n = n_1 + n_2 + \cdots + n_p$ que l'on exprime en utilisant le signe $\sum$ (lire : grand sigma) par la formule

$$\boxed{\sum_{i=1}^{p} n_i = n}$$

L'expression $\sum_{i=1}^{p} x_i$ se lit : « somme de la variable x_i, pour l'indice i variant de 1 à p ».

Avec les mêmes notations $\sum_{i=1}^{5} x_i = x_1 + x_2 + x_3 + x_4 + x_5$.

13. Expressions usuelles.

1º $\sum_{i=1}^{5} n_i x_i = n_1 x_1 + n_2 x_2 + \cdots + n_5 x_5$.

2º $\sum_{i=1}^{k} n_i(x_i + y_i) = n_1 (x_1 + y_1) + n_2 (x_2 + y_2) + \cdots + n_k (x_k + y_k)$.

soit, en développant :

$$\boxed{\sum_{i=1}^{k} n_i (x_i + y_i) = \sum_{i=1}^{k} n_i x_i + \sum_{i=1}^{k} n_i y_i}$$

3º Si $a_i = a$ pour tout i, (a constant)

$$\boxed{\sum_{i=1}^{n} a = na}$$

4º $\sum_{i=1}^{n} (x_i + a) = (x_1 + a) + \cdots + (x_n + a) = na + \sum_{i=1}^{n} x_i$.

5º $\sum_{i=1}^{n} ax_i = ax_1 + ax_2 + \cdots + ax_n = a (x_1 + x_2 + \cdots x_n)$

soit :

$$\boxed{\sum_{i=1}^{n} ax_i = a \cdot \sum_{i=1}^{n} x_i}$$

14. Applications. — *Calculer* $\sum_{i=1}^{n}(x_i + a_i)^2$

$$\sum_{i=1}^{n}(x_i + a_i)^2 = \sum_{i=1}^{n}(x_i^2 + 2a_i x_i + a_i^2) = \sum_{i=1}^{n} x_i^2 + 2\sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} a_i^2.$$

Si $a_i = a$, $\forall_i$, la relation s'écrit :

$$\sum_{i=1}^{n}(x_i + a)^2 = \sum_{i=1}^{n} x_i^2 + 2a\sum_{i=1}^{n} x_i + na^2.$$

REMARQUE. — Dans la suite, lorsqu'il n'y aura aucune ambiguïté possible, on simplifiera l'écriture du symbole $\sum_{i=1}^{n}$ par $\sum_{i}$, et parfois par $\sum$.

EXERCICES

Dépouillement d'une collecte d'observations.

1. On a relevé les deux derniers chiffres minéralogiques de 80 voitures, au fur et à mesure de leur passage.

34	01	09	34	11	75	75	94	90	75	08	35	36	45	41	34
19	75	75	06	13	23	25	19	15	11	66	75	11	34	38	45
15	12	34	66	69	75	94	92	34	08	54	45	75	12	34	19
03	26	45	75	66	33	33	75	34	75	12	29	11	67	38	36
69	75	03	06	75	34	34	15	12	11	66	45	35	29	67	59

Dépouiller les renseignements fournis et présenter les résultats du dépouillement sous forme d'un tableau statistique suivant les classes :

[00,10[[10,20[[20,30[[30,40[[40,50[[50,60[.... [90,99].

2. On a procédé au recensement de 46 familles d'une ville en relevant le nombre des enfants à la charge de chacune d'elles. Au hasard de ses déplacements, le statisticien dresse le tableau suivant où chaque nombre représente celui des enfants d'un ménage :

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

Regrouper ces résultats suivant un tableau à 2 colonnes contenant la variable (nombre d'enfants) et les effectifs (nombre de ménages).

Calcul des fréquences. Tableau des effectifs cumulés.

3. A Paris, la répartition des salles de cinéma suivant le prix moyen des places pratiqué au cours de l'année 1964 (format standard) est fourni par le tableau :

Prix	Nombre de salles	Prix	Nombre de salles
moins de 1,40	440	de 1,90 à 1,99	460
de 1,40 à 1,49	332	de 2 à 2,49	1 342
de 1,50 à 1,59	413	de 2,50 à 2,99	521
de 1,60 à 1,69	469	de 3 à 3,99	363
de 1,70 à 1,79	568	de 4 à 4,99	98
de 1,80 à 1,89	492	plus de 5	78

Établir un tableau où figureront les fréquences en % et les effectifs cumulés croissants et décroissants.

4. La répartition suivant l'âge des conducteurs de cycles victimes d'accidents corporels de la circulation routière en 1963 est fourni par le tableau :

Ages	Tués	Blessés	Ages	Tués	Blessés
0 à 4 ans	1	12	35 à 44 ans	70	1 675
5 à 14 ans	104	3 016	45 à 54 ans	130	2 004
15 à 24 ans	65	3 644	55 à 64 ans	212	2 390
25 à 34 ans	46	1 700	65 et plus	185	1 271

1° Déterminer les fréquences en % des effectifs tués ou blessés.

2° Établir le tableau des effectifs cumulés croissants et décroissants, ainsi que leurs fréquences (tués ou blessés).

5. La répartition des livrets de caisse d'épargne ordinaire, en métropole, suivant leur importance, au 31 décembre 1964, est fourni par le tableau :

Sommes versées F	Nombre de livrets *(en milliers)*	Sommes versées	Nombre de livrets *(en milliers)*
Moins de 1,01	1 825	5 000 à 7 500	722
1,01 à 30	5 377	7 500 à 10 000	557
30,01 à 1 000	4 571	10 000 à 12 500	601
1 000 à 2 500	1 427	12 500 à 15 000	261
2 500 à 5 000	1 084	au-dessus de 15 000	254

Déterminer les fréquences en % et établir le tableau des effectifs cumulés croissants et décroissants.

6. Tableau à double entrée. — Faire inscrire sur une feuille, par chaque élève de la classe, son poids et sa taille. Regrouper les résultats dans un tableau à double entrée (série double) après avoir choisi les classes de poids et les classes de taille.

7. Décomposition par âge, au 31 décembre 1963, de l'effectif ouvrier fond et jour, des houillères de France (Sources : Charbonnages de France).

Age	Effectif	Age	Effectif	Age	Effectif	Age	Effectif	Age	Effectif
14	57	24	3 484	34	5 552	44	4 355	54	2 483
15	610	25	4 039	35	6 000	45	3 699	55	1 158
16	1 449	26	4 490	36	6 155	46	2 871	56	566
17	1 876	27	4 359	37	6 290	47	2 729	57	412
18	1 541	28	4 289	38	6 576	48	3 371	58	290
19	924	29	4 765	39	6 440	49	4 732	59	215
20	670	30	4 579	40	6 379	50	3 640	60	48
21	1 882	31	5 173	41	6 469	51	3 126	61	7
22	2 151	32	5 398	42	6 431	52	2 712	62	8
23	3 194	33	5 744	43	6 518	53	2 694	63	3

1° Présenter les résultats sous forme d'un tableau statistique où les observations seront groupées par classes d'amplitude 5 : de 14 à 18; de 19 à 23, etc.

2° Indiquer dans le tableau les fréquences, les effectifs cumulés croissants et décroissants.

Signe $\sum$

8. Développer et calculer :

$$\sum_{i=1}^{8} (x_i + 5) \qquad \sum_{i=1}^{6} 3\, x_i \qquad \sum_{i=a}^{b} 8\, x_i \qquad \text{(a et b entiers naturels)}$$

$$\sum_{i=n}^{m} (x_i - 3) \qquad \sum_{k+1}^{2k} (x_i - m) \qquad \sum_{n}^{2n} m\,(x_i + b)$$

9. Développer et calculer :

$$\sum_{i=1}^{4} (x_i - 1)^2 \qquad \sum_{i=1}^{n} (x_i - 3)^2 \qquad \sum_{i=k}^{2k} (x_i + a)^2$$

REPRÉSENTATION GRAPHIQUE DES SÉRIES

15. Présentation graphique des séries statistiques. — Pour donner une idée synthétique des tableaux numériques, on utilise généralement une représentation graphique sur un repère cartésien. L'axe Ox représente les valeurs du caractère quantitatif et l'axe Oy les valeurs des effectifs ou des fréquences.

Les graphes des séries statistiques simples se présentent sous deux formes : *le diagramme en bâton* et *l'histogramme.*

16. Le diagramme en bâton. — Il est surtout utilisé pour schématiser les séries statistiques correspondant à un caractère discret (variable discontinue, nº 4). A chaque valeur x_i de la variable, faisons correspondre un segment parallèle à l'axe Oy dont la *longueur* représente *l'effectif* n_i ou la *fréquence* $f(x_i)$.

Exemple. — *Répartition d'après le nombre d'enfants du personnel d'un lycée.*

Nombre d'enfants x_i	Effectifs n_i	Effectifs cumulés croissants
1	13	13
2	25	38
3	20	58
4	12	70
5	6	76
6	3	79
7	1	80

Nous obtenons :

— le diagramme en bâtons de la série (fig. 1);

— le diagramme en bâtons de la série cumulée croissante (fig. 2).

17. Histogramme d'une série statistique. — C'est le diagramme qui correspond aux séries à *variable quantitative continue.*

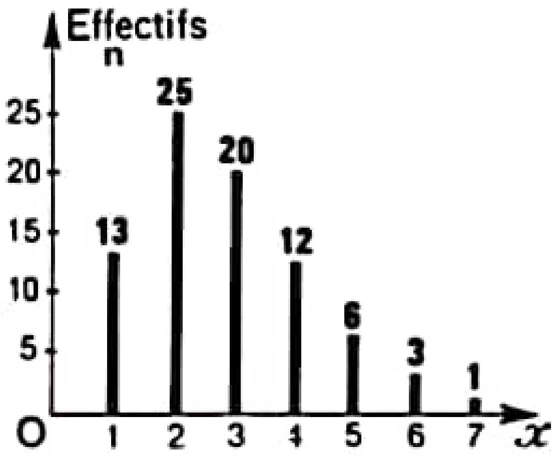

Fig. 1.

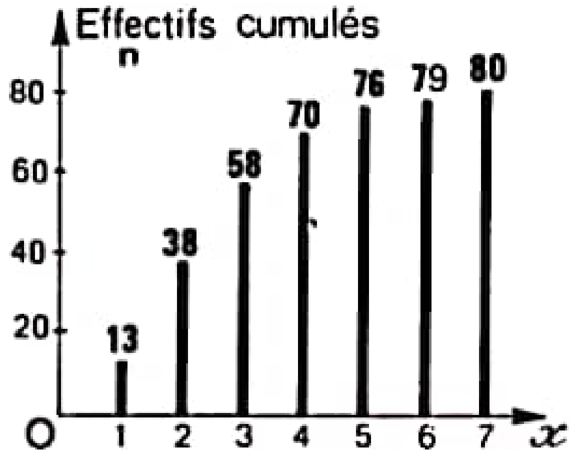

Fig. 2.

EXEMPLE. — *On donne la répartition des élèves d'un lycée d'après leurs tailles.*

Tailles *(cm)* x_i	[150,154[	[154,158[	[158,162[	[162,166[	[166,170]	TOTAL
Effectifs n_i	25	50	200	175	50	500

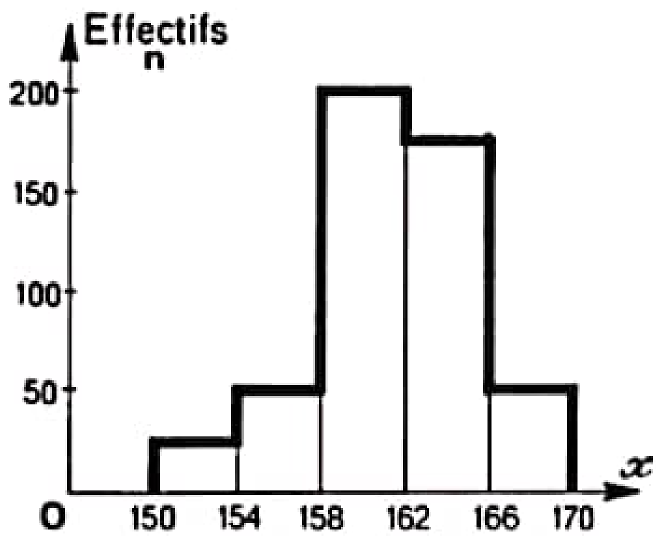

Fig. 3.

Sur la figure nº 3, une classe de tailles admet pour image un rectangle dont un côté porté par Ox représente l'amplitude de la classe (4 cm) et dont la hauteur parallèle à Oy représente l'effectif correspondant.

L'ensemble des cinq rectangles constitue l'histogramme de la série.

18. Remarques importantes.

1º Si l'unité d'aire (fig. 3) est celle du rectangle ayant pour dimensions les unités portées par Ox et Oy, l'aire limitée par l'ensemble des rectangles, lorsque *les classes ont même amplitude*, est le produit de l'intervalle de classe par la somme des longueurs des côtés parallèles à Oy. Cette somme représente l'effectif total. On déduit le résultat :

L'aire de l'histogramme est proportionnelle à l'effectif total.

2º *Les classes n'ont pas la même amplitude.*

Considérons une série statistique obtenue à partir de la précédente, en remplaçant la colonne

166 à 170
50

par la colonne

166 à 174
50

Le rectangle correspondant à la classe [166, 170[(fig. 4) de l'histogramme (fig. 3) est remplacé dans le nouveau diagramme par le rectangle ayant pour base l'intervalle [166, 174[et pour hauteur 25 et *non* 50 (fig. 5). Cela revient à admettre qu'il existe 25 élèves ayant une taille comprise entre 166 cm et 170 cm, et 25 élèves ayant une taille définie dans l'intervalle [170, 174[.

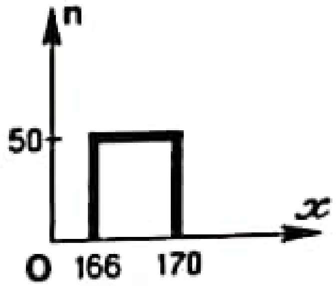

Fig. 4.

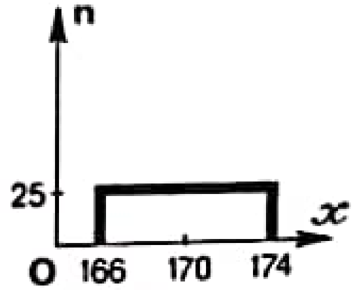

Fig. 5.

Plus généralement, pour la *construction de l'histogramme* d'une série, si l'amplitude d'une classe est *multipliée par k, l'effectif* correspondant doit être *divisé par k.*

Cette remarque met en évidence le résultat suivant :

Dans un histogramme, l'aire d'un rectangle est proportionnelle à son effectif.

19. Polygone des effectifs. — Considérons l'histogramme (fig. 6) correspondant à la série statistique du n° 17, dont les classes ont même amplitude.

On appelle centre de classe, la moyenne arithmétique des valeurs extrêmes de la classe.

Ainsi, le centre de la classe [154, 158[est : $\dfrac{154 + 158}{2} = 156$ cm.

Le polygone des effectifs s'obtient en joignant les points de l'histogramme admettant pour abscisses les centres des classes et pour ordonnées, les effectifs correspondants.

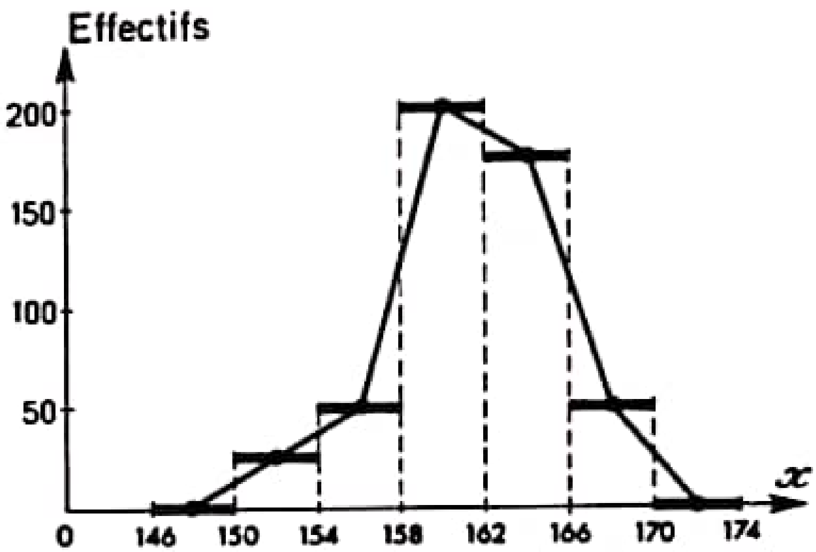

Fig. 6.

On peut le compléter aux extrémités par les points de l'axe Ox, ayant pour abscisses les milieux des segments relatifs aux classes précédant et suivant la série (fig. 6). Ainsi le polygone des effectifs et l'axe Ox limitent une surface dont l'aire est proportionnelle à l'effectif total.

Ce polygone donne une idée générale de la série étudiée. Si l'axe Oy est l'axe des fréquences, nous obtenons le *polygone des fréquences.*

Courbe de fréquence :

Si l'effectif d'une série est suffisamment nombreux pour qu'on puisse partager son étendue en classes dont le nombre est de plus en plus grand et l'intervalle de plus en plus petit, on obtient un histogramme composé de rectangles très étroits. Le polygone des fréquences correspondant tend vers une *courbe limite* lorsque le nombre des rectangles augmente indéfiniment, c'est-à-dire lorsque l'intervalle de classe tend vers zéro. Cette courbe limite est appelée *courbe de fréquence.*

20. Polygones des effectifs cumulés. Courbes cumulatives. — Reprenons la série du nº 17. Mentionnons, dans un tableau, les effectifs cumulés croissants et décroissants.

Tailles *(cm)* x_i	Effectifs n_i	Effectifs cumulés croissants	Effectifs cumulés décroissants	Fréquences $f(x_i)$
150 à moins de 154	25	25	500	0,05
154 — 158	50	75	475	0,10
158 — 162	200	275	425	0,40
162 — 166	175	450	225	0,35
166 à 170	50	500	50	0,10

1º La figure nº 7 représente le *polygone des effectifs cumulés croissants*, obtenu en joignant les points de coordonnées :

(150, 0) (154, 25) (158, 75)
(162, 275) (166, 450) (170, 500).

L'abscisse de l'un des points, sauf le premier, est la valeur de l'extrémité droite de la classe; l'ordonnée correspondante est l'effectif cumulé croissant.

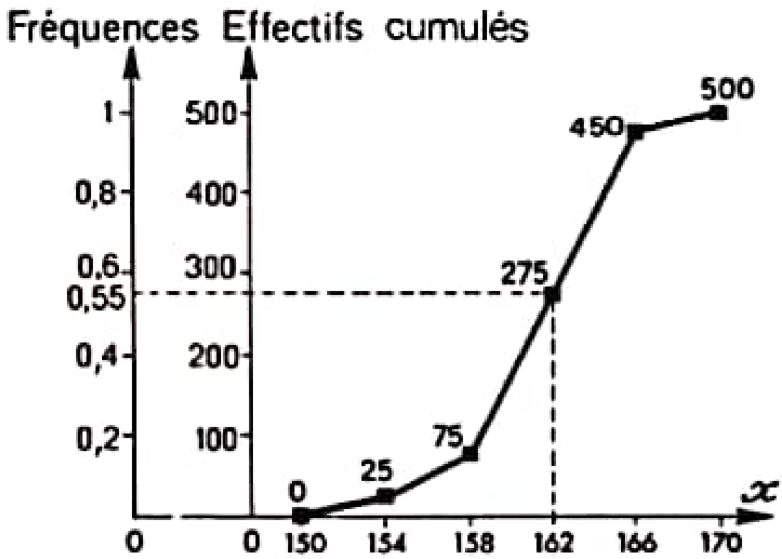

Fig. 7.

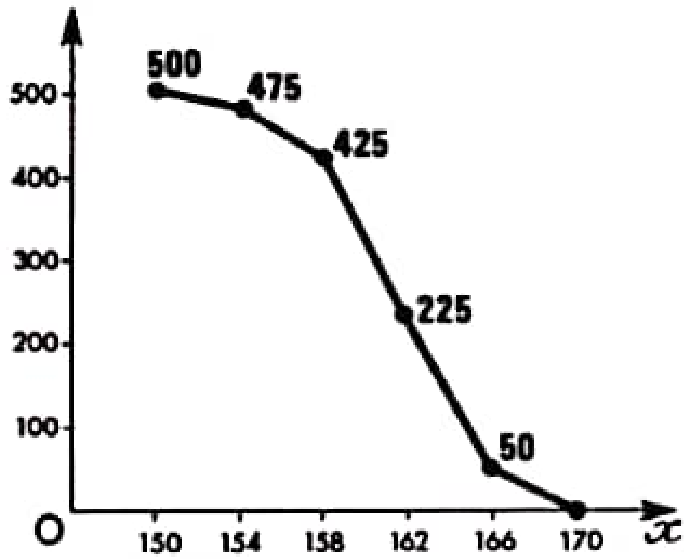

Fig. 8.

2º La figure nº 8 représente le *polygone des effectifs cumulés décroissants*, obtenu en joignant les points de coordonnées :

(150, 500) (154, 475) (158, 425)
(162, 225) (166, 50) (170, 0).

L'abscisse de l'un des points, sauf le dernier, est la valeur de l'extrémité gauche de la classe; l'ordonnée correspondante est l'effectif cumulé décroissant.

3º On vérifie facilement, en utilisant l'échelle des fréquences (fig. 7), que $\frac{275}{500} = 0{,}55$, soit 55 % des élèves, ont une taille inférieure à 162 cm.

REMARQUE. — En remplaçant les effectifs par les fréquences, on obtient le *polygone des fréquences cumulées* croissantes ou décroissantes.

Lorsque l'intervalle de classe tend vers zéro, dans les mêmes conditions qu'au nº 19, le polygone des fréquences cumulées tend vers une courbe dite *courbe des fréquences cumulées* et le polygone des effectifs cumulés tend vers la courbe des effectifs cumulés.

Les courbes précédentes sont encore appelées *courbes cumulatives.*

21. Diagramme des séries chronologiques. — Représentons sur un diagramme (fig. 9), la production de fonte brute en France, exprimée en millions de tonnes (source : Charbonnages de France).

Année	1956	1957	1958	1959	1960	1961	1962	1963
Production	11,5	11,9	12	12,5	14,1	14,6	14	14,3

Sur Ox, chaque intervalle représente une année dont le millésime est écrit sous l'intervalle correspondant. Les abscisses des points figuratifs du diagramme sont situées au milieu des intervalles.

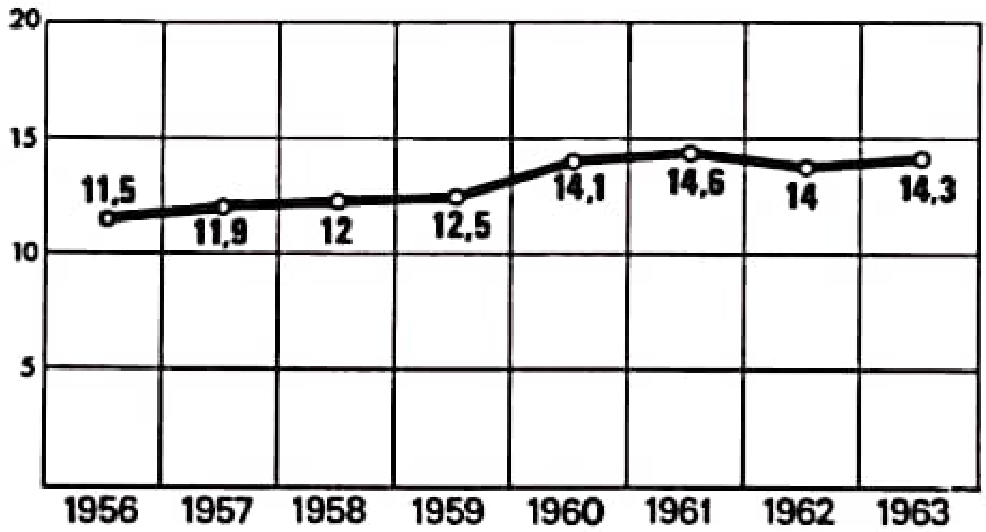

Fig. 9.

AUTRES DIAGRAMMES

22. Diagramme polaire.

1º *Coordonnées polaires d'un point.* — Dans un repère orthonormé xOy, considérons un point M et traçons l'axe OX admettant OM pour support (fig. 10). Posons :

$$(\overrightarrow{Ox}, \overrightarrow{OX}) = \theta \qquad \overline{OM} = r.$$

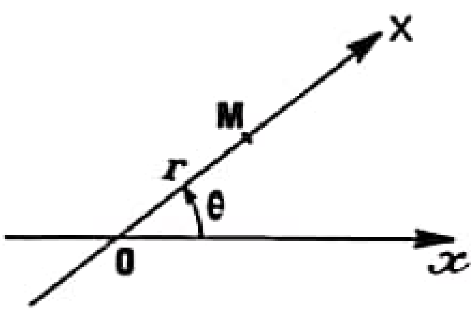

Fig. 10.

Le couple de nombres (θ, r) constitue un système de coordonnées polaires du point M.

2° Le diagramme polaire est utilisé pour représenter une distribution chronologique admettant un cycle périodique tel que la semaine, le mois, le trimestre, l'année.

EXEMPLE. — Représentons, sur un diagramme polaire, les températures moyennes mensuelles de l'air à la station de Montpellier en 1965 (Source : Météorologie nationale).

Mois	Janv.	Fév.	Mars	Avril	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.
$t°$	6,4	4,3	9,8	13,2	16,3	20	21,4	21,2	17,4	15,9	10,5	8,8

Le diagramme polaire (fig. 11) s'obtient en traçant douze rayons représentant les différents mois et sur lesquels on définit un point image de la température enregistrée.

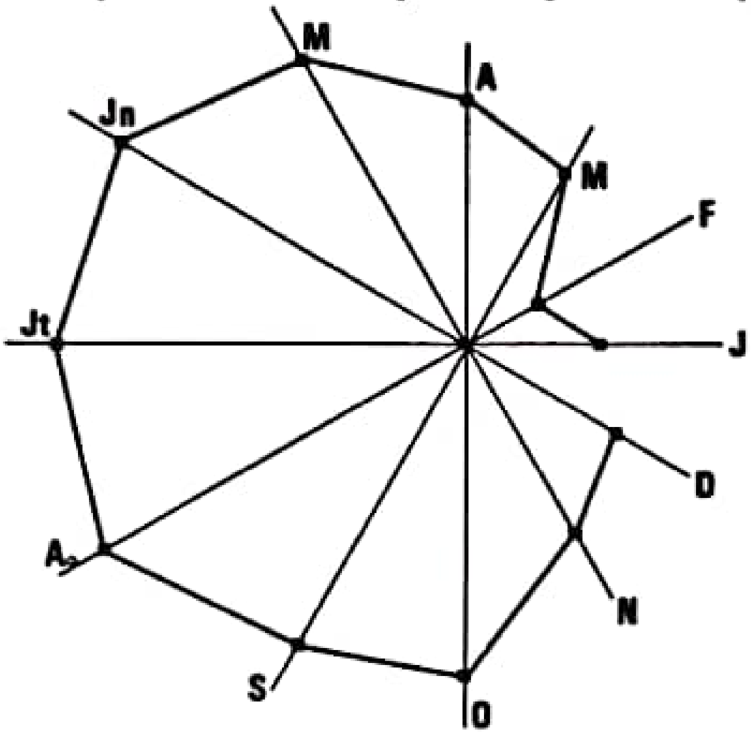

Fig. 11.

• 23. Diagrammes semi-logarithmiques. — Les échelles utilisées dans les diagrammes précédents étaient métriques. Supposons que l'on veuille tracer le graphique relatif à la série suivante :

Nombres de postes de télévision dans la Lozère.

Année	1957	1958	1959	1960	1961	1962	1963	1964	1965
Effectif	8	66	307	777	1 371	2 207	3 251	4 421	5 423

Il est difficile de représenter le graphique à cause de la différence importante entre le nombre de postes en 1957 et celui de 1965. Aussi utilise-t-on une échelle logarithmique.

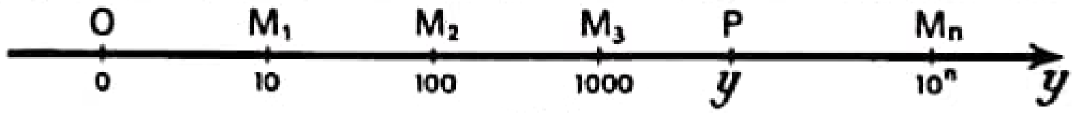

Fig. 12.

• L'étude de ce paragraphe ne figure pas au programme de 1re D.

Définition d'une échelle logarithmique.

Sur l'axe Oy (fig. 12), définissons les points M_1, M_2, M_3,... M_n, par les relations:

$$\overline{OM}_1 = \log 10 = 1 \qquad \overline{OM}_2 = \log 100 = 2 \qquad \overline{OM}_3 = \log 1\,000 = 3$$

$$\overline{OM}_n = \log 10^n = n.$$

Sur l'axe logarithmique Oy, les nombres 10, 10^2, 10^3,... 10^n marqués en regard des points M_1, M_2, M_3, ... M_n représentent leurs cotes. Ainsi, si P appartient à cet axe, la cote de P est le nombre y tel que

$$\overline{OP} = \log y.$$

Pour construire le diagramme de la série précédente (fig. 13), nous avons défini :

$\log 8 = 0{,}903\ 09$
$\log 66 = 1{,}819\ 54$
$\log 307 = 2{,}487\ 44$
$\log 777 = 2{,}890\ 42$
$\log 1371 = 3{,}137\ 04$
$\log 2207 = 3{,}343\ 80$
$\log 3251 = 3{,}511\ 28$
$\log 4421 = 3{,}645\ 52$
$\log 5423 = 3{,}734\ 24$

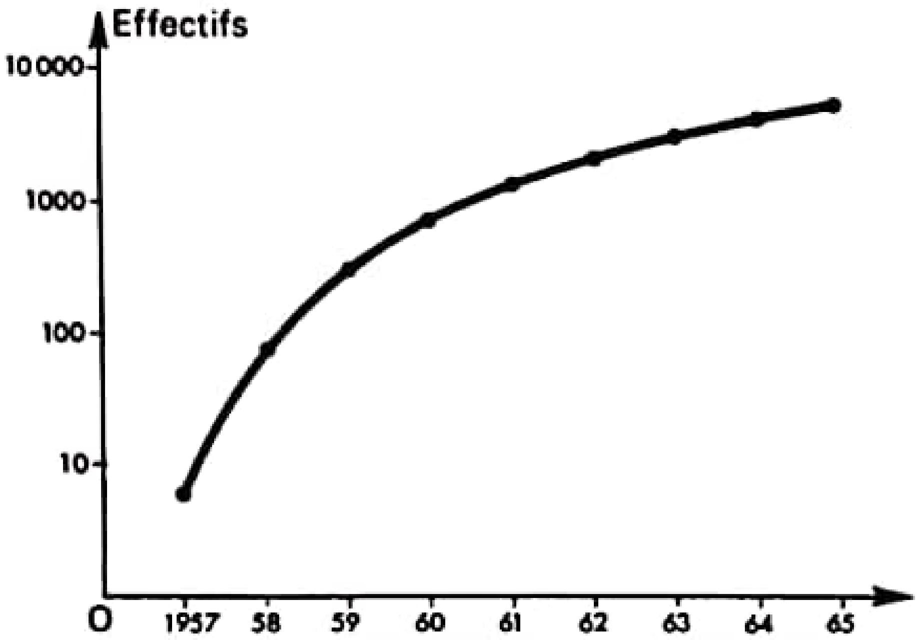

Fig. 13.

24. Diagramme à secteurs. — Il comprend un cercle, ou un demi-cercle divisé en secteurs proportionnels aux différents constituants de la population étudiée.

EXEMPLE. — En 1964, 1 535 millions de tonnes de marchandises ont été transportées en France. Les bateliers en ont acheminé 86 millions de tonnes, les chemins de fer 248 millions et les routiers 1 202 millions. Les pourcentages correspondants sont : 6 %, 16 %, 78 %, ce qui correspond sur le diagramme (fig. 14) à des angles au centre de 12 grades, 32 grades, 156 grades.

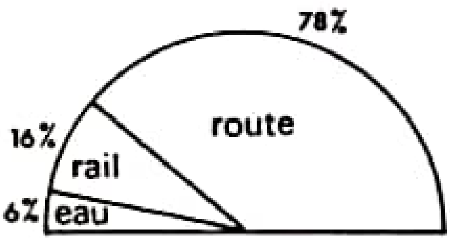

Fig. 14.

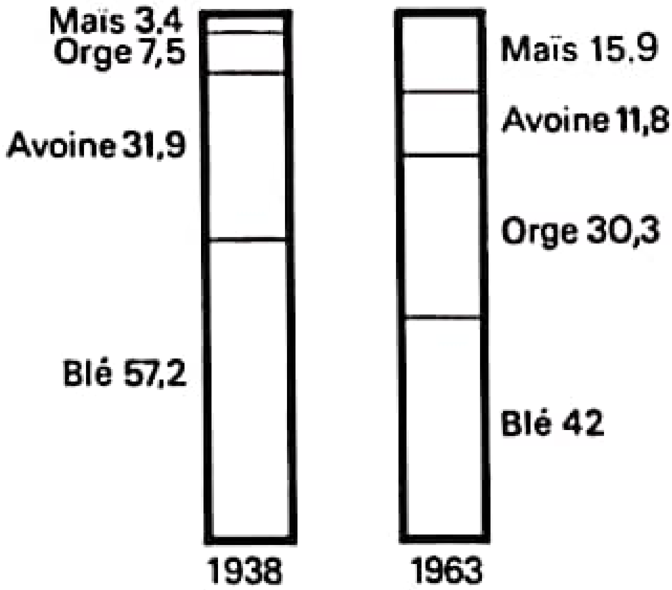

Fig. 15.

25. Diagramme en barres. — Ce diagramme, constitué de plusieurs rectangles accolés, permet de mettre en évidence les grandeurs relatives des différentes parties d'un ensemble.

Exemple. — Évolution de la production (en millions de quintaux) des principales céréales en France en 1938 et 1963 (source : min. de l'Agriculture).

Année	Blé	Orge	Avoine	Maïs
1938	98 010	12 908	54 574	5 786
1963	102 490	73 840	28 760	38 707

Le diagramme en barres (fig. 15) fait apparaître le pourcentage de la production des céréales en 1938 et 1963.

PROBLÈME RÉSOLU

Un groupe de 50 enfants a lancé un poids. Les distances en mètres sont indiquées dans le tableau :

Distance *m*	moins de 3	3 à < 4	4 à < 5	5 à < 7	7 à < 8	plus de 8
Effectifs	3	5	10	24	6	2

1° Tracer l'histogramme des fréquences relatives et le polygone des fréquences.
2° Tracer la courbe des fréquences cumulées.
3° Déterminer le nombre d'enfants qui lancent le poids à une distance inférieure à 6 m.

La série statistique présente des intervalles de classes inégaux. En choisissant *un* mètre pour unité des classes, nous introduisons [2 m, 3 m[pour la 1re classe et [8 m, 9 m,[pour la 6e classe. Pour tracer l'histogramme et le polygone des fréquences (fig. 16), nous rectifions les fréquences, en utilisant la remarque du n° 18.

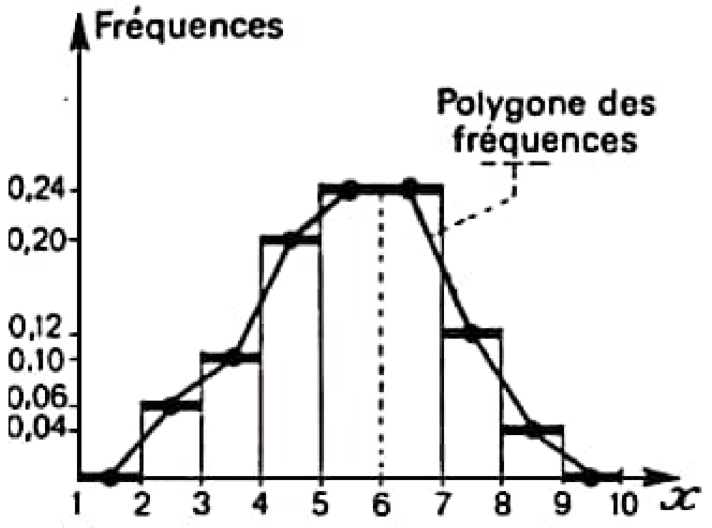

Fig. 16.

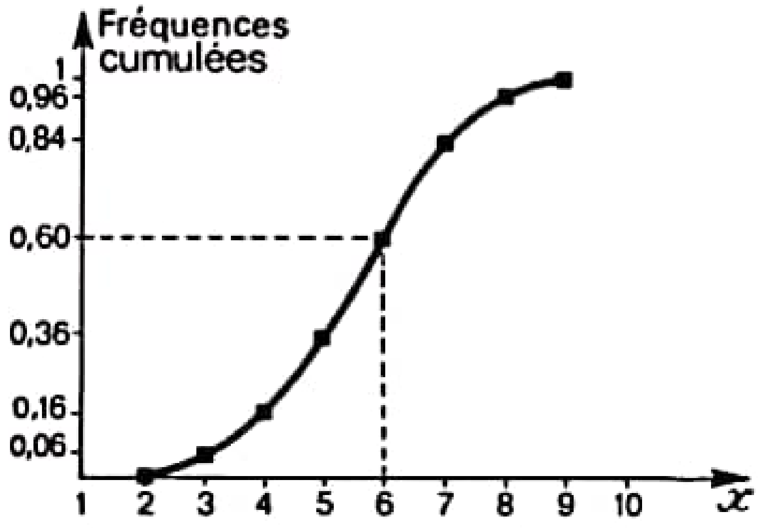

Fig. 17.

Les résultats sont consignés dans le tableau :

Classes	Intervalle de classe	Effectifs y_i	Fréquences relatives	Fréquences cumulées	Fréquence rectifiée
2 à 3	1	3	0,06	0,06	
3 à 4	1	5	0,10	0,16	
4 à 5	1	10	0,20	0,36	
5 à 7	2	24	0,48	0,84	0,48 : 2 = 0,24
7 à 8	1	6	0,12	0,96	
8 à 9	1	2	0,04	1,00	

Sur la courbe des fréquences cumulées croissantes (fig. 17), nous obtenons par lecture directe la fréquence cumulée correspondant à une distance de 6 m, soit 0,60 ou 60 %.

Le nombre d'enfants qui lancent le poids à une distance inférieure à 6 m est :

$$\frac{60}{100} \times 50 = 30 \text{ enfants.}$$

On déduit que 20 enfants le lancent à une distance supérieure à 6 m.

EXERCICES

Séries à caractère discontinu. Diagrammes en bâton.

10. D'après la distribution suivante du nombre d'enfants à la charge des familles :

Enfants	0	1	2	3	4	5	6	7	8
Familles	144	195	130	80	58	45	24	6	3

1° Établir un tableau où figureront : chaque classe, les effectifs, les fréquences, les effectifs cumulés croissants et décroissants.

2° Tracer les diagrammes en bâtons des effectifs, des effectifs cumulés croissants et décroissants.

11. D'après la répartition des résidences principales suivant le nombre d'occupants dans le département de la Seine (source I.N.S.E.E., recensement 1962) :

Unité : 1 000 logements

Occupants	1	2	3	4	5	6
Logements	612,16	661,94	408,16	248,58	124,68	102,96

1° Tracer le diagramme en bâtons, les diagrammes cumulés croissants et décroissants des effectifs.

2° Quelles sont les fréquences relatives cumulées des logements ayant un nombre d'occupants inférieur ou égal à n $(1 \leqslant n \leqslant 6)$? Représentation graphique.

Séries à caractère continu. Histogramme.

12. Répartition des naissances des enfants suivant l'âge de la mère en 1963 en France (Source I.N.S.E.E.) :

Moins de 20 ans	20 à 24 ans	25 à 29 ans	30 à 34 ans	35 à 39 ans	40 ans et plus
42 474	239 270	275 600	181 001	90 016	29 738

1° Dresser un tableau contenant les classes d'âge, les effectifs, les effectifs cumulés croissants et décroissants.

Indication : on admettra que l'effectif des moins de 20 ans correspond à la classe d'âge [15, 19] *et l'effectif des plus de 40 ans à la classe d'âge* [40, 45].

2° Tracer l'histogramme des effectifs.

3° Tracer les courbes des effectifs cumulés croissants et décroissants.

13. Structure de la population active par âge dans les Hauts-de-Seine (en millions d'individus).

Age au 1er janvier 1963	15-19	20-24	25-29	30-34	35-39	40-44	45-49
Nombre	30,54	51,04	79,80	87,10	80,08	71,92	58,62
Age	50-54	55-59	60-64	65-69	70-74	75 et plus	
Nombre	70,62	66,06	45,38	17,70	5,36	3,22	

1° Dresser un tableau contenant les effectifs, les fréquences relatives, les effectifs cumulés croissants et décroissants.

2° Tracer l'histogramme des effectifs et les courbes cumulatives des effectifs.

14. Reprendre les exercices nos 3, 4, 7 et tracer l'histogramme des effectifs et les courbes cumulatives des effectifs.

Diagrammes polaires.

Représenter, au moyen d'un diagramme polaire, les séries statistiques suivantes :

15. Nombre de mariages par trimestre en 1950 et 1951 dans le département de l'Aude (source I.N.S.E.E.).

Année	1er trim.	2e trim.	3e trim.	4e trim.
1950	324	497	397	603
1951	308	446	390	542

16. Nombre d'enfants nés vivants par mois, exprimé en milliers, en France (source I.N.S.E.E.).

Année	Janv.	Fév.	Mars	Avril	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.
1959	71,7	65,5	72,5	71,0	73,1	67,5	70,8	69,6	67,5	67,1	63,6	65,7
1960	67,7	65,7	70,4	68,5	73,0	68,7	71,6	69,7	67,9	65,7	61,4	65,9
1961	71,5	65,6	73,0	72,3	75,1	69,6	71,3	69,5	67,2	67,9	64,8	67,8

17. Étant donné l'évolution de l'indice des prix de détail dans l'agglomération parisienne de 1960 à 1963 (base 100 : juillet 1956).

	1960	1961	1962	1963
Janvier	130,1	133,1	139,2	146,6
Février	130,4	133,2	139,0	146,8
Mars	130,4	133,1	139,7	146,8
Avril	130,6	133,0	139,8	147,4
Mai	130,3	132,7	140,6	148,1
Juin	130,2	132,4	141,1	149,1
Juillet	130,7	133,4	141,8	150,0
Août	131,9	134,2	141,5	150,7
Septembre	132,1	134,9	142,0	151,9
Octobre	132,3	136,4	142,6	152,2
Novembre	132,7	137,8	143,9	153,1
Décembre	133,0	138,3	144,7	153,4

Représenter graphiquement la série statistique en coordonnées polaires.

Diagramme semi-logarithmique.

18. On considère la production d'une « Fabrication » lancée en 1962 :

Année x_i	1961	1962	1963	1964	1965	1966
Production (unités y_i)	9	101	1 089	10 480	99 890	1 001 028

1° Représenter dans un graphique la série précédente.

2° Si on pose $x = 1$ pour l'année 1961.
$x = 2$ pour l'année 1962, etc.

et y la production, montrer que $x \approx \log y$ (relation approximative qui lie y à x).

19. 1° Tracer deux axes rectangulaires gradués de la façon suivante : l'axe horizontal porte une échelle arithmétique, l'axe vertical une échelle logarithmique. Construire cette échelle de 1 à 12 à l'aide des renseignements suivants :

$\log 2 = 0{,}301\ 03 \qquad \log 3 = 0{,}477\ 12 \qquad \log 7 = 0{,}845\ 10. \qquad \log 11 = 1{,}041\ 39.$

2° Utiliser le tracé précédent pour représenter en coordonnées semi-logarithmiques les productions de deux usines (en millions de tonnes).

	1957	1961
Usine A..................	2	6
Usine B..................	4	12

En joignant les points représentatifs de la production de chaque usine, on obtient deux segments parallèles. Expliquer pourquoi.

(*Baccalauréat*).

ÉLÉMENTS CARACTÉRISTIQUES D'UNE SÉRIE

VALEURS TYPIQUES

26. Généralités. — Les diagrammes des séries statistiques étudiées précédemment (diagramme en bâton, histogramme) ont mis en évidence une accumulation des effectifs dans le voisinage d'une valeur particulière du caractère et un certain étalement des effectifs de part et d'autre de cette valeur.

Nous introduisons deux éléments caractéristiques dans l'étude d'une série statistique :

— *les valeurs typiques ou valeurs centrales :* le *mode*, la *médiane*, la *moyenne*.

— les *indices de dispersion*.

27. Le mode. — ***Le mode est la valeur du caractère correspondant à l'effectif le plus élevé.***

Dans la série statistique étudiée au nº 16 : « répartition du personnel d'un lycée d'après le nombre d'enfants », le mode est 2, correspondant au plus fort effectif 25.

Dans la série du nº 17 : « répartition des élèves d'un lycée d'après leur taille », on considère la *classe modale* (ou *classe dominante*), [158, 162[qui comprend le plus grand nombre d'élèves : 200. La valeur centrale de la classe, 160 cm, est le mode de la série.

REMARQUE. — Le mode est une valeur type, d'un usage limité, et perd toute signification dans une série dite *plurimodale* qui admettrait plusieurs maxima relatifs.

28. La médiane. — ***La médiane est la valeur du caractère qui partage la série suivant des effectifs égaux.***

Les notes obtenues par un élève, au cours d'une semaine, sont :

4, 7, 7, 11, 13, 15, 17.

La médiane de cette distribution est 11, car trois notes lui sont inférieures et supérieures.

Si le nombre des notes est pair ($n = 2p$), tout nombre compris entre la $p^{\text{ième}}$ et la $(p + 1)^{\text{ième}}$ valeur peut être médiane. On choisira la demi-somme. Ainsi, dans la répartition :

5 7 8 8 9 11 13 14 14 16,

la médiane est $$\frac{1}{2}(9 + 11) = 10.$$

REMARQUE. — Dans la distribution suivante : 9 9 9 $\boxed{9}$ 11 12 13, le nombre encadré 9 occupe la position de la médiane. Aucune note n'est inférieure à 9; trois notes lui sont supérieures. Il n'y a donc pas de médiane dans cette série.

29. Calcul de la médiane d'une série à variable continue. — Reprenons la répartition des élèves d'un lycée d'après leur taille (n° 17) et dressons le tableau des effectifs cumulés croissants et décroissants.

Tailles *(cm)* x_i	Effectifs n_i	Effectifs cumulés croissants	Effectifs cumulés décroissants
150 à moins de 154	25	25	500
154 à — 158	50	75	475
158 à — 162	200	275	425
162 à — 166	175	450	225
166 à — 170	50	500	50

Nous cherchons la taille de la $\frac{500}{2} = 250^e$ personne. Le tableau montre que 75 élèves ont une taille inférieure à 158 cm. La taille médiane se situe dans la classe [158, 162[. Admettons que la taille des 200 enfants qui se trouvent dans cette classe varie régulièrement. Nous effectuons une interpolation linéaire. La valeur médiane M est donc :

$$M = 158 + (162 - 158) \times \frac{250 - 75}{275 - 75} = 158 + 4 \times \frac{175}{200} = 161{,}5 \text{ cm.}$$

30. Détermination graphique de la médiane. — Le tracé de la courbe cumulative croissante C permet d'obtenir la médiane M en déterminant l'intersection de C avec la droite parallèle à Ox, « d'ordonnée 250 » (fig. 18).

On peut également utiliser la courbe cumulative C des effectifs décroissants (fig. 18). Il en résulte que l'intersection des deux courbes C et C′ est un point P dont l'abscisse est la médiane M.

31. Remarque. — Si la médiane se prête assez mal au calcul algébrique, elle présente l'avantage de ne pas tenir compte des valeurs anormalement grandes ou petites qui peuvent intervenir dans la série. La médiane serait inchangée dans l'étude précédente (n° 29), s'il y avait dans le lycée un nain et un géant.

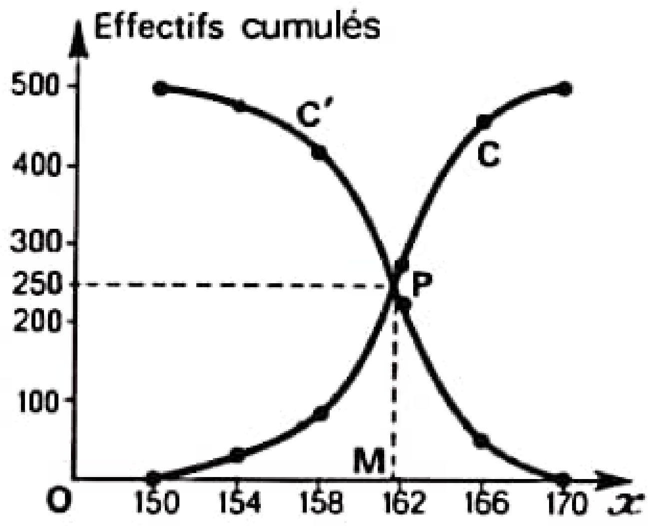

Fig. 18.

MOYENNE ARITHMÉTIQUE

32. Moyenne arithmétique simple. — Si $x_1, x_2, ..., x_n$ sont les valeurs de n observations, la moyenne arithmétique est :

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Exemple. — La moyenne arithmétique des notes 8, 10, 12, 14 obtenues par un élève est :

$$\bar{x} = \frac{8 + 10 + 12 + 14}{4} = 11.$$

33. Moyenne arithmétique pondérée. — Dans la distribution suivante, qui schématise une série à variable discontinue,

Variable x_i	x_1	x_2	$\cdots$	x_i		x_p
Effectif n_i	n_1	n_2	$\cdots$	n_i		n_p

la moyenne arithmétique pondéree est définie par l'expression :

$$x = \frac{n_1x_1 + n_2x_2 + \cdots + n_px_p}{n_1 + n_2 + \cdots + n_p} = \frac{\sum_{i=1}^{p} n_ix_i}{\sum_{i=1}^{p} n_i} = \frac{1}{n}\sum_{i=1}^{p} n_ix_i$$

en désignant par n l'effectif total.

Comme la fréquence $f(x_i) = \frac{n_i}{n}$ (n° 7), on déduit :

$$\bar{x} = \sum_{i=1}^{p} x_i f(x_i).$$

Les n_i sont appelés *coefficients de pondération.*

Exemple. — Les notes obtenues par un candidat à un examen dont les épreuves admettent pour coefficients 3, 5, 1, 1 sont respectivement 10, 12, 14, 8. Le coefficient 3 relatif à la note 10, correspond à l'effectif de cette note; autrement dit, la note 10 est obtenue trois fois. La série s'écrit :

Notes x_i	Effectifs n_i	Produits $n_i x_i$
10	3	30
12	5	60
14	1	14
8	1	8
	10	112

Moyenne : $x = \frac{112}{10} = 11,2.$

34. Exécution des calculs pour une série à variation continue. — Déterminons à titre d'exemple, la moyenne arithmétique de la série des tailles étudiée au n° 17. Dans l'application des formules précédentes (nº 32), x_i représente la valeur du centre de classe. Les calculs sont détaillés dans le tableau suivant :

Tailles (*cm*)	Centre de classes x_i	Effectifs n_i	Produits $n_i x_i$
150 — 154	152	25	3 800
154 — 158	156	50	7 800
158 — 162	160	200	32 000
162 — 166	164	175	28 700
166 — 170	168	50	8 400
		500	80 700

$$\bar{x} = \frac{80\,700}{500} = 161{,}4 \text{ cm.}$$

35. Simplification des calculs. — Le calcul précédent se simplifie en adoptant une moyenne provisoire x_0. Généralisons la méthode.

1º *Notations :*

x_i = valeur du caractère ou centre de classe.
n_i = effectif correspondant.
n = effectif total.
x_0 = moyenne provisoire.

2º *Calcul :*

Quel que soit i, $x_i = x_0 + (x_i - x_0)$.

Dans ces conditions :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i n_i = \frac{1}{n}\sum [x_0 n_i + (x_i - x_0)n_i]$$

Développons :

$$\bar{x} = \frac{1}{n}\sum x_0 n_i + \frac{1}{n}\sum (x_i - x_0)\, n_i.$$

Or :

$$\frac{1}{n}\sum x_0 n_i = \frac{x_0}{n}\sum n_i = \frac{x_0}{n}\cdot n = x_0.$$

Donc :

$$\bar{x} = x_0 + \frac{1}{n}\sum_{i=1}^{n} (x_i - x_0)\, n_i.$$

3º *Exemple :*

Reprenons l'exemple du nº 34 et choisissons comme moyenne provisoire le centre de la classe [158, 162[, soit $x_0 = 160$. Présentons les calculs dans le tableau suivant :

Tailles	Centres de classes x_i	Effectifs n_i	$x_i - x_0$	$n_i(x_i - x_0)$	
150 à 154	152	25	− 8	− 200	
154 à 158	156	50	− 4	− 200	
158 à 162	160	200	0	0	0
162 à 166	164	175	4		700
166 à 170	168	50	8		400
		500		− 400	1 100
				+ 700	

$$\bar{x} = x_0 + \frac{1}{n} \sum_{i=1} (x_i - x_0)\, n_i \qquad \text{soit} \qquad \bar{x} = 160 + \frac{700}{500} = 161{,}4 \text{ cm.}$$

AUTRES MOYENNES

36. Moyenne géométrique. — La moyenne géométrique de deux nombres positifs x_1 et x_2 est le nombre g tel que $\boxed{g = \sqrt{x_1 x_2}}$

Plus généralement, la moyenne géométrique de plusieurs nombres positifs $x_1, x_2, \ldots, x_n$ est le nombre g tel que $\boxed{g = \sqrt[n]{x_1 x_2 \cdots x_n}}$

Le calcul de la moyenne géométrique se fait en utilisant les logarithmes :

$$\log g = \frac{1}{n} (\log x_1 + \log x_2 + \cdots + \log x_n).$$

Remarque. — Dans la distribution suivante :

Variable x_i	x_1	x_2	$\cdots$	x_i	$\cdots$	x_p
Effectif n_i	n_1	n_2	$\cdots$	n_i	$\cdots$	n_p

la *moyenne géométrique pondérée* est définie par l'expression :

$$g = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdots x_p^{n_p}} \qquad \text{où} \qquad n = \sum_{i=1}^{p} n_i$$

La formule logarithmique s'écrit : $\log g = \dfrac{1}{n} \displaystyle\sum_{i=1}^{p} n_i \cdot \log x_i$.

37. Moyenne harmonique. — La moyenne harmonique de deux nombres x_1 et x_2 est le nombre h tel que

$$\frac{1}{h} = \frac{1}{2} \left(\frac{1}{x_1} + \frac{1}{x_2} \right).$$

Plus généralement, dans le cas d'une série d'observations $x_1, x_2, \dots, x_p$, d'effectifs $n_1, n_2, \dots, n_p$, la moyenne harmonique h est définie par l'expression

$$\boxed{\frac{1}{h} = \frac{1}{n}\sum_{i=1}^{p}\frac{n_i}{x_i}} \quad \text{en posant } n = \sum_{i=1}^{p} n_i.$$

Remarquons que l'inverse de la moyenne harmonique est égale à la moyenne arithmétique pondérée des inverses des valeurs observées.

Exemple. — Un cycliste parcourt la distance AB $= d$ km à la vitesse de 40 km/h à l'aller et 30 km/h au retour. Le temps t pour parcourir le trajet est :

$$t = \frac{d}{40} + \frac{d}{20} = d\left(\frac{1}{40} + \frac{1}{30}\right) \text{ heures.}$$

La vitesse moyenne du cycliste est :

$$V = \frac{d \times 2}{t} = \frac{2}{\frac{1}{40} + \frac{1}{30}} \text{ km/h}$$

soit :

$$\frac{1}{V} = \frac{1}{2}\left(\frac{1}{40} + \frac{1}{30}\right).$$

V est donc la moyenne harmonique des vitesses aller et retour et non leur moyenne arithmétique.

38. Relation entre les trois moyennes. — On démontre que :

$$\boxed{\bar{x} \geqslant g \geqslant h}$$

Ce résultat se vérifie facilement pour deux observations de valeurs x_1 et x_2 positives

$$\bar{x} = \frac{x_1 + x_2}{2}; \qquad g = \sqrt{x_1 x_2}; \qquad \frac{1}{h} = \frac{1}{2}\left(\frac{1}{x_1} + \frac{1}{x_2}\right).$$

En effet :

— L'inégalité $\bar{x} \geqslant g$ entraine :

$$\frac{x_1 + x_2}{2} \geqslant \sqrt{x_1 x_2} \longrightarrow (x_1 + x_2)^2 \geqslant 4x_1x_2 \longrightarrow (x_1 - x_2)^2 \geqslant 0$$

— L'inégalité $g \geqslant h$ entraine :

$$\sqrt{x_1 x_2} \geqslant \frac{2x_1x_2}{x_1 + x_2} \Longrightarrow x_1x_2 \geqslant \frac{4\,(x_1x_2)^2}{(x_1 + x_2)^2} \longrightarrow (x_1 + x_2)^2 \geqslant 4x_1x_2$$

qui est bien vérifié.

PROBLÈME RÉSOLU

Les salaires mensuels payés au personnel d'une entreprise se répartissent ainsi :

— de 800 à 900 F	4
— de 900 à 1 000 F	20
— de 1 000 à 1 100 F	107
— de 1 100 à 1 200 F	168
— de 1 200 à 1 300 F	122
— de 1 300 à 1 400 F	48
— de 1 400 à 1 500 F	21
— de 1 500 à 1 600 F	10

1° Établir la relation qui donne la moyenne arithmétique $\bar{x}$ des salaires en fonction de x_0 origine provisoire et k étendue d'une classe.

2° Appliquer cette relation au calcul du salaire moyen dans la série proposée.

3° Si on désigne par $\bar{x}$ la moyenne arithmétique, par Me la médiane, par Mo le mode, on montre que dans les séries de faible dissymétrie, comme celle proposée, il existe la relation $\bar{x} - Mo = 3(\bar{x} - Me)$.

Calculer Mo à partir de $\bar{x}$ et Me.

D'une façon générale posons :

x_0 moyenne provisoire.

k l'amplitude de l'intervalle de classe que nous choisissons pour unité.

x_i la valeur du centre de la classe d'ordre i.

La différence $x_i - x_0$ mesurée dans la nouvelle unité est k fois plus petite, soit :

$$u_i = \frac{x_i - x_0}{k} \iff x_i = x_0 + ku_i.$$

Calcul de la moyenne arithmétique $\bar{x}$.

$$\bar{x} = \frac{1}{n}\sum n_i x_i = \frac{1}{n}\sum n_i(x_0 + ku_i) = \frac{1}{n}\sum n_i x_0 + \frac{k}{n}\sum n_i u_i.$$

En posant $\bar{u} = \frac{1}{n}\sum n_i u_i$, il vient :

$$\boxed{\bar{x} = x_0 + k\bar{u}}$$

Appliquons ce résultat à la détermination du salaire moyen. Les calculs sont résumés dans le tableau suivant, dans lequel $x_0 = 1\,150$ et $k = 100$.

Classes	Centres de classes x_i	Effectifs n_i	$u_i = \frac{x_i - x_0}{k}$	$n_i u_i$		Effectif croissant
800 à 900	850	4	− 3	− 12		4
900 à 1 000	950	20	− 2	− 40		24
1 000 à 1 100	1 050	107	− 1	− 107		131
1 100 à 1 200	1 150	168	0	0	0	299
1 200 à 1 300	1 250	122	1		122	421
1 300 à 1 400	1 350	48	2		96	469
1 400 à 1 500	1 450	21	3		63	490
1 500 à 1 600	1 550	10	4		40	500
		500		− 159	+ 321	
				162		

$$\bar{u} = \frac{162}{500} = 0{,}324 \qquad \bar{x} = 1\,150 + 100 \times 0{,}324 = \underline{1\,182{,}4}\ \text{F}$$

Calcul de la médiane Me : Cherchons le salaire de l'effectif situé au 250[e] rang.

$$Me = 1\,100 + (1\,200 - 1\,100) \times \frac{250 - 131}{299 - 131} = \underline{1\,170{,}8} \text{ F}$$

Calcul du mode de la série :

$$\bar{x} - Mo = 3\,(\bar{x} - Me) \longrightarrow Mo = 3\,Me - 2\,\bar{x}.$$

$$Mo = 3 \times 1\,170{,}8 - 2 \times 1\,182{,}4 \quad \text{soit} \quad Mo = \underline{1\,147{,}6} \text{ F}$$

EXERCICES

Médiane.

20. Étant donné une série à variation continue, montrer en utilisant la série du n° 29, que la droite ayant pour abscisse la médiane, partage l'histogramme suivant deux surfaces équivalentes.

21. Déterminer la médiane de chacune des distributions :

1, 4, 5, 5, 0, 8, 2.
18, 0, 8, 16, 10, 4, 4, 14, 2, 14.

22. On considère la série statistique suivante :

Classes	Effectifs
0 à 5	2
5 à 10	4
10 à 20	6
20 à 40	8

Déterminer la médiane de cette série
a) par le calcul.
b) graphiquement.

23. Décès en 1963 des personnes du sexe masculin suivant l'âge, en France.

Moins de 5 ans	5 à 19	20 à 39	40 à 59	60 à 69	70 et plus
12 711	3 512	13 024	52 646	65 665	136 051

Déterminer la médiane de la distribution par le calcul et par le graphique.

Médiane. Moyennes.

24. Étant donné n observations de valeur x_i, de moyenne arithmétique x, calculer

$$\sum_{i=1}^{n} (x_i - \bar{x}).$$

25. Une série d'observations concernant les tailles d'un groupe d'adolescents de 11 à 14 ans a donné les résultats suivants :

Plus de 140 et au plus 144 cm	3
Plus de 144 et au plus 148 cm	17
Plus de 148 et au plus 152 cm	63
Plus de 152 et au plus 156 cm	82
Plus de 156 et au plus 160 cm	69
Plus de 160 et au plus 164 cm	31
Plus de 164 et au plus 168 cm	20
Plus de 168 et au plus 172 cm	4
Plus de 172 et au plus 174 cm	1
Plus de 174 et au plus 178 cm	1

1° Construire l'histogramme représentant la série. En déduire la classe dominante.
2° Déterminer la taille moyenne.

26. Les résultats du saut en hauteur au cours d'une séance de culture physique sont les suivants :

Hauteur cm	[90, 95[	[95, 100[	[100, 105[	[105, 110[	[110, 115[	[115, 120[	[120, 130[
Effectif	4	10	15	25	16	8	2

1° Construire l'histogramme et les courbes cumulatives croissantes et décroissantes.
2° Déterminer la classe modale et la médiane (calcul et méthode graphique).
3° Déterminer la hauteur moyenne du saut.

27. *Répartition des exploitations agricoles suivant la superficie (bois non compris).* Recensement de 1956.

Sources : ministère de l'Agriculture et I.N.S.E.E.

Superficie des exploitations	Nombre d'exploitations (Répartition en %)
Moins de 1 ha	6,6
1 ha à 1,99 ha	10,2
2 ha à 4,99 ha	18,2
5 ha à 9,99 ha	20,8
10 ha à 19,99 ha	23,5
20 ha à 49,99 ha	16,5
50 ha et plus	4,2
	100,0

1° Déterminer le mode de cette distribution.
2° Calculer la médiane (solution par le calcul et solution graphique).
3° Calculer la moyenne arithmétique.
4° Quelles conclusions peut-on tirer de la comparaison de ces trois paramètres de position ?

28. 17 copies d'examen notées de 0 à 20 ont donné les résultats suivants:

1, 2, 3, 3, 5, 7, 8, 8, 8, 10, 11, 12, 12, 13, 15, 16, 18

Calculer : la moyenne arithmétique; la moyenne géométrique; la moyenne harmonique. Écrire la relation entre ces trois moyennes.

29. Dans une usine, les salaires mensuels sont répartis en pourcentage et sous une forme simplifiée de la façon suivante :

Salaire	Hommes	Femmes
De 600 à 700 francs	30	45
De 700 à 800 francs	32	30
De 800 à 900 francs	24	19
De 900 à 1 000 francs	14	6

On demande :

1° La moyenne des salaires masculins : m;

2° La moyenne des salaires féminins : m';

3° La moyenne des salaires d'un ménage : M.

Pour résoudre cette dernière question, on démontrera que $M = m + m'$, de la façon suivante : étant donné une variable x pouvant prendre les valeurs x_1, x_2, x_3, x_4 de moyenne m, et une variable y pouvant prendre les valeurs y_1, y_2, y_3, y_4 de moyenne m', montrer que la variable $z = x + y$ (x et y se combinant de toutes les façons possibles) a pour valeur moyenne $M = m + m'$.

Peut-on généraliser dans le cas où les variables x et y prendraient n valeurs.

(*Baccalauréat*).

30. On achète :

I

Une première fois pour 1 000 F d'escudos au cours de 10 F l'escudo.

Une seconde fois pour 1 000 F d'escudos au cours de 12,5 F l'escudo.

Montrer que le *cours moyen pour l'ensemble des deux opérations* est la moyenne harmonique des deux cours pratiqués.

II

Soient deux données positives x_1 et x_2 distinctes.

a) Classer, en justifiant ce classement, les moyennes arithmétiques ($\bar{x}$), géométrique (G) et harmonique (h) de ces deux données.

b) Montrer que la moyenne géométrique des données est également la moyenne géométrique des deux autres moyennes considérées.

c) Établir que la moyenne harmonique h de x_1 et x_2 est également la moyenne harmonique des deux quantités $(h - x_1)$ et $(h - x_2)$.

Existe-t-il une propriété correspondante pour la moyenne arithmétique ?

(*Baccalauréat*).

INDICES DE DISPERSION

39. Généralités. — Les valeurs typiques d'une série (mode, médiane, moyenne) donnent une idée sommaire de la distribution des observations mais ne suffisent pas à la caractériser.

Considérons, par exemple, deux groupes de sept élèves d'une même classe qui obtiennent à un devoir les notes suivantes :

Groupe A :	9	10	10	11	12	12	13
Groupe B :	5	7	9	11	13	15	17

Ces deux séries ont même médiane et même moyenne arithmétique 11. Dans la première, les valeurs sont groupées autour de la valeur typique 11; dans la deuxième, elles sont plus étalées, plus *dispersées* de la valeur centrale. Nous disons que la distribution A a une *faible dispersion*, tandis que B a une *forte dispersion.*

40. Étendue d'une série ou intervalle de variation. — ***L'étendue d'une série est la différence entre ses deux valeurs extrêmes.***

EXEMPLES.

Groupe A : étendue = 13 − 9 = 4 points.
Groupe B : étendue = 17 − 5 = 12 points.

Dans la série des tailles de 500 enfants (nº 17), l'étendue de la série est 175 − 150 = 25 cm.

On utilise aussi les expressions : *éventail* ou *range.*

L'étendue d'une série est facile à déterminer, mais elle ne caractérise pas convenablement la dispersion lorsque les valeurs extrêmes sont accidentelles.

41. Les quartiles. — La médiane sépare la série des observations en deux groupes d'effectifs égaux. Déterminons la médiane de chacune des deux moitiés. On obtient :

1º le *premier quartile* Q_1 qui définit la valeur Q_1 du caractère tel que le quart des observations soit inférieur à Q_1, les trois quarts supérieurs à Q_1.

2º le *second quartile* ou médiane M de la série.

3º le *troisième quartile* Q_3 qui définit la valeur Q_3 du caractère telle que le quart des observations soit supérieur à Q_3 et les trois quarts inférieurs à Q_3.

Ces définitions sont schématisées par la figure nº 19.

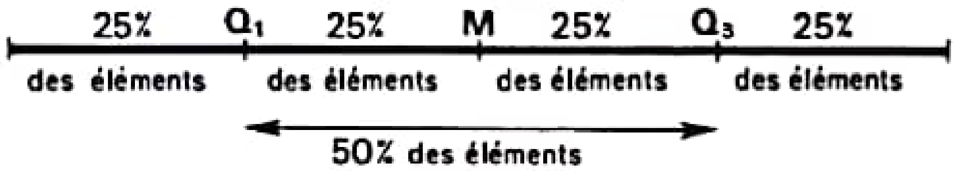

Fig. 19.

L'intervalle interquartile est la différence $Q_3 - Q_1$. Il caractérise la dispersion car *il contient la moitié* (50 %) *des effectifs.*

42. Exemple de détermination des quartiles. — Le calcul des quartiles, analogue au calcul de la médiane, s'obtient au moyen des séries cumulées. Considérons la répartition de 500 enfants d'après leur taille.

Tailles *(cm)*	Effectifs n_i	Effectifs cumulés
[150, 154[	25	25
[154, 158[	50	75
[158, 162[	200	275
[162, 166[	175	450
[166, 170]	50	500

Le quartile Q_1 est la taille qui correspond à la $500 \times \frac{25}{100} = 125^e$ observation, donc Q_1 appartient à la classe [158, 162[.

$$Q_1 = 158 + (162-158)\frac{125 - 75}{275 - 75} = 159 \text{ cm.}$$

Le quartile Q_3 est la taille qui correspond à la $500 \times \frac{75}{100} = 375^e$ observation. Q_3 appartient donc à la classe [162, 166[.

$$Q_3 = 162 + (166 - 162)\frac{375 - 275}{450 - 275} = 164{,}28 \text{ cm.}$$

L'intervalle interquartile est :

$$164{,}28 - 159 = 5{,}28 \text{ cm.}$$

Cette série, d'étendue 20 cm, contient 100 % des observations dont les 50 % se trouvent dans un intervalle d'amplitude 5,28 cm.

Remarque. — La courbe cumulative (fig. 20) permet de trouver graphiquement les résultats en traçant deux droites parallèles à Ox, d'ordonnées 125 et 375.

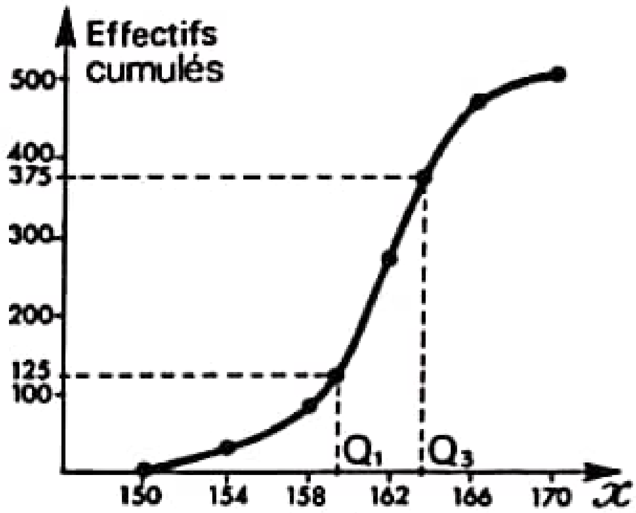

Fig. 20.

43. Déciles. — En généralisant la notion précédente de quartile, on définit les déciles qui, au nombre de 9, divisent la série en dix parties d'effectifs égaux.

Le $p^{ième}$ décile D_p d'une série, est la valeur du caractère tel que les $10\,p$ % des observations aient une valeur inférieure à D_p. Ainsi le premier décile D_1 est la valeur D_1 telle que 10 % des observations aient une valeur inférieure à D_1. De même, le neuvième décile D_9 est la valeur D_9, telle que 90 % des valeurs observées lui soient inférieures.

La différence $D_9 - D_1$ définit un intervalle contenant les 80 % des effectifs.

On définirait de la même façon des *centiles*, qui au nombre de 99, partageraient la série en 100 groupes d'effectifs égaux.

44. Écart absolu moyen.

1° On appelle *ecart* d'une variable x_i par rapport au nombre a, la valeur absolue de leur différence, c'est-à-dire $|x_i - a|$.

1° *On appelle écart absolu moyen d'un ensemble de données, la moyenne arithmétique des écarts de ces données par rapport à leur moyenne arithmétique.*

Si $\bar{x}$ est la moyenne arithmétique des n données $x_1, x_2, ..., x_i, ...x_n$, l'écart moyen est :

$$\boxed{e = \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}|}$$

Exemple. — Si 8, 10, 11, 12, 13, 14, 16 sont les notes hebdomadaires d'un élève, leur moyenne arithmétique est 12 et l'écart moyen a pour valeur :

$$e = \frac{1}{7}(4 + 2 + 1 + 0 + 1 + 2 + 4) = 2.$$

Cela signifie que les notes s'écartent en moyenne de 2 par rapport à leur moyenne.

3° Dans le cas d'observations groupées en classes, l'écart absolu moyen est défini par l'expression :

$$\boxed{e = \frac{1}{n} \sum_{i=1}^{c} n_i \, |x_i - \bar{x}|}$$

dans laquelle x_i est la valeur centrale de classe d'effectif n_i et c le nombre de classes.

Exemple. — Déterminer l'écart moyen de la série n° 42 sachant que la moyenne arithmétique est $\bar{x} = 161{,}4$ cm (n° 34).

Tailles	Centre de classes x_i	Effectifs n_i	$\lvert x_i - \bar{x}\rvert$	$n_i \lvert x_i - \bar{x}\rvert$
[150, 154[	152	25	9,4	235
[154, 158[	156	50	5,4	270
[158, 162[	160	200	1,4	280
[162, 166[	164	175	2,6	455
[166, 170]	168	50	6,6	330
		500		1 570

$$e = \frac{1\,570}{500} = 3{,}14 \text{ cm.}$$

Les tailles de ces enfants s'écartent en moyenne de 3,14 cm de la taille moyenne de l'ensemble des effectifs.

Remarque. — Dans cette étude, nous avons défini l'écart moyen d'une série par rapport à la moyenne arithmétique. On pourrait le définir par rapport à une valeur typique, la médiane par exemple. D'ailleurs, dans les distributions courantes, la médiane et la moyenne arithmétique ont des valeurs très voisines.

45. Fluctuation ou variance. — ***La fluctuation d'un ensemble de données est la moyenne arithmétique des carrés des écarts de ces données par rapport à leur moyenne arithmétique.***

La fluctuation, désignée par σ^2, s'exprime par la formule :

$$\boxed{\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

46. Écart-type ou écart quadratique moyen. — ***L'écart-type d'un ensemble de données est la racine carrée de leur fluctuation.***

$$\boxed{\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

L'unité de l'écart-type σ est celle du caractère de la série.

Dans le cas d'observations groupées en classes, la somme $\sum$ est étendue au nombre de classes et x_i représente la valeur centrale.

Exemple. — *Calcul direct de la fluctuation et de l'écart-type de la série n° 42, dans laquelle $\bar{x} = 161{,}4$ cm.*

Les calculs sont présentés dans le tableau suivant :

Tailles	Centre de classe x_i	Effectifs n_i	$x_i - x$	$(x_i - \bar{x})^2$	$n_i(x_i - x)^2$
150 à 154	152	25	− 9,4	88,36	2 209
154 à 158	156	50	− 5,4	29,16	1 458
158 à 162	160	200	− 1,4	1,96	392
162 à 166	164	175	2,6	6,76	1 183
166 à 170	168	50	6,6	43,56	2 178
		500			7 420

$$\sigma^2 = \frac{7\,420}{500} = 14{,}84 \qquad \text{donc} \qquad \boxed{\sigma \approx 3{,}8 \text{ cm}}$$

Remarque. — Les calculs se compliquent rapidement et deviennent erronés lorsque la moyenne arithmétique x n'est connue qu'avec une certaine approximation. Pour simplifier les calculs de l'écart-type on procède d'une façon analogue au calcul de la moyenne arithmétique (n° 35) en effectuant un changement d'origine.

47. — Simplification du calcul de l'écart-type : changement d'origine. — Choisissons une valeur arbitraire x_0 du caractère.

$$x_i - x_0 = (x_i - x) + (\bar{x} - x_0).$$

Calculons l'expression

$$E = \sum_{i=1}^{n} n_i(x_i - x_0)^2 = \sum n_i\,[(x_i - \bar{x}) + (\bar{x} - x_0)]^2.$$

Développons :

$$E = \sum n_i(x_i - \bar{x})^2 + 2n_i(x_i - \bar{x})(\bar{x} - x_0) + n_i(\bar{x} - x_0)^2$$

$$E = \sum n_i(x_i - \bar{x})^2 + 2\sum n_i(x_i - \bar{x})(\bar{x} - x_0) + \sum n_i(\bar{x} - x_0)^2$$

Or, $\bar{x} - x_0$ et $(\bar{x} - x_0)^2$ sont des constantes que nous mettons en facteur,

$$E = \sum n_i(x_i - \bar{x})^2 + 2(x - x_0)\sum n_i(x_i - \bar{x}) + (\bar{x} - x_0)^2\,n.$$

Comme $\quad \sum n_i(x_i - x) = \sum n_i x_i - \bar{x}\sum n_i = n\bar{x} - n\bar{x} = 0,$

il reste :

$$E = \sum n_i(x_i - \bar{x})^2 + n(\bar{x} - x_0)^2$$

soit
$$\sum n_i(x_i - \bar{x})^2 = \sum n_i(x_i - x_0)^2 - n(\bar{x} - x_0)^2. \qquad (1)$$

La fluctuation s'exprime donc par la formule :

$$\boxed{\sigma^2 = \frac{1}{n}\sum n_i(x_i - x_0)^2 - (x - x_0)^2}$$

Si $x_0 = 0$,
$$\boxed{\sigma^2 = \frac{1}{n}\sum n_i x_i^2 - x^2}$$

Exemple. — *Reprenons le calcul de l'écart-type de la série précédente en choisissant* $x_0 = 160$ *cm comme valeur arbitraire du caractère.*

Tailles	Centre de classe x_i	Effectifs n_i	$x_i - 160$	$(x_i - 160)^2$	$n_i(x_i - 160)^2$
150 à 154	152	25	− 8	64	1 600
154 à 158	156	50	− 4	16	800
158 à 162	160	200	0	0	0
162 à 166	164	175	4	16	2 800
166 à 170	168	50	8	64	3 200
		500			8 400

$$\sigma^2 = \frac{8\,400}{500} - (161{,}4 - 160)^2 = 16{,}8 - 1{,}96 = 14{,}84$$

$$\sigma \approx 3{,}8 \text{ cm.}$$

REMARQUE. — Le résultat (1) montre que

$$\sum n_i(x_i - \bar{x})^2 < \sum n_i(x_i - x_0)^2$$

autrement dit, l'expression $\sum n_i\,(x_i - x_0)^2$ est minimum si x_0 est la moyenne pondérée $\bar{x}$ des x_i.

48. Intérêt de l'écart-type. — Dans l'exemple précédent, déterminons les valeurs $\bar{x} \pm 2\sigma = 161{,}4 \pm 2 \times 3{,}8$ soit 153,8 cm et 169 cm. Un calcul très simple d'interpolation linéaire montre que l'intervalle [153,8; 169] couvre environ 92 % des effectifs de la série.

D'une façon générale, on démontre que l'intervalle $[x - 2\sigma, \bar{x} + 2\sigma]$ contient, quelle que soit la série, les 75 % des observations. Une faible valeur de l'écart-type indique une accumulation des effectifs au voisinage de la moyenne, tandis qu'une grande valeur de σ est l'indice d'un étalement des observations.

49. Coefficient de variation. — Dans la comparaison de plusieurs séries statistiques, on utilise le coefficient de variation ou coefficient de dispersion qui représente le rapport de l'écart-type à la moyenne arithmétique :

$$V = \frac{\sigma}{\bar{x}}.$$

c'est un coefficient sans dimension. Dans l'exemple précédent, le coefficient de variation est $\dfrac{3{,}8}{161{,}4} \approx 0{,}02$.

PROBLÈME RÉSOLU

Les séries statistiques se présentent souvent sous la forme du tableau ci-dessous, dont les valeurs centrales des classes forment une progression arithmétique.

Centres de classe	a	$a + b$	$a + 2b$	...	$a + (n - 1)b$	Total
Effectifs	y_0	y_1	y_2	...	y_{n-1}	N

1° Vérifier que la valeur moyenne, $\bar{x}$, et le carré de l'écart-type, σ^2, de cette série, ont pour valeur

$$\bar{x} = a + \frac{b\sum iy_i}{N}, \qquad \sigma^2 = \frac{b^2}{N}\left[\sum i^2 y_i - \frac{\left(\sum iy_i\right)^2}{N}\right].$$

2° Calculer à l'aide de ces formules le salaire moyen et l'écart-type de la série donnant la répartition du personnel spécialisé d'une entreprise suivant les salaires journaliers :

Salaires	30 à 40	40 à 50	50 à 60	60 à 70	70 à 80	80 à 90	90 à 100
Effectifs	11	26	63	81	35	21	13

1° Par définition : $\bar{x} = \frac{1}{N}\sum y_i(a + ib) = \frac{1}{N}\sum y_i a + \frac{1}{N}\sum y_i ib.$

Or :
$$\sum y_i a = a\sum y_i = aN$$

et :
$$\sum y_i ib = b\sum y_i i.$$

Donc :
$$\bar{x} = a + \frac{b}{N}\sum iy_i.$$

Pour le calcul de la fluctuation, utilisons la forme :

$$\sigma^2 = \frac{1}{N}\sum (x_i - x_0)^2\, n_i - (\bar{x} - x_0)^2$$

dans laquelle $x_0 = a$, , $n_i = y_i$. Il vient :

$$\sigma^2 = \frac{1}{N}\sum (a + ib - a)^2\, y_i - \left(a + \frac{b}{N}\sum iy_i - a\right)^2$$

soit :

$$\sigma^2 = \frac{b^2}{N}\sum i^2 y_i - \frac{b^2}{N^2}\left(\sum iy_i\right)^2 \quad \text{ou} \quad \sigma^2 = \frac{b^2}{N}\left[\sum i^2 y_i - \frac{\left(\sum iy_i\right)^2}{N}\right].$$

2° Les calculs sont résumés dans le tableau qui suit. On pose :

$$a = 35 \qquad b = 10.$$

N° de la classe i	Centre de classe	Effectif y_i	iy_i	i^2y_i
0	35	11	0	0
1	45	26	26	26
2	55	63	126	252
3	65	81	243	729
4	75	35	140	560
5	85	21	105	525
6	95	13	78	468
		250	718	2 560

$$\bar{x} = 35 + 10 \times \frac{718}{250} = 63{,}72F$$

$$\sigma^2 = \frac{100}{250}\left[2\,560 - \frac{(718)^2}{250}\right] = 199{,}16$$

$$\sigma \approx 14{,}1F.$$

EXERCICES

Range. Quartiles et Déciles.

31. Déterminer l'étendue des séries statistiques :

a) 8, 15, 5, 13, 1, 27, 38.

b) a, $\sqrt{a}$, $\sqrt{b}$, b, b^2 si $1 < a < b$.

32. Déterminer les quartiles des séries statistiques :

a) 7, 10, 22, 2, 0, 8, 13, 12, 20, 18, 21.

b) 2, 5, 0, 4, 6, 5, 9, 8, 11, 8.

33. Répartition de salariés d'après le salaire touché en 1961 (Source : I.N.S.E.E.).

Salaires milliers F	0 à 2	2 à 3	3 à 4	4 à 5	5 à 6	6 à 8	8 à 10	10 à 15	15 à 20	20 à 35	35 à 50	50 à 70
Effectifs milliers	188	216	654	855	973	1 673	975	869	250	209	56	35

Déterminer les quartiles.

Écart absolu.

34. Déterminer l'écart absolu moyen (ou écart arithmétique) des notes d'un élève :

9, 7, 10, 12, 18, 16.

35. Déterminer l'écart absolu moyen par rapport à la médiane, de la série :

5, 12, 3, 15, 11, 11, 8.

36. Déterminer, en utilisant la formule de définition, l'écart-type de la distribution

6, 18, — 24, 8, 24, — 14, 4, 12, 6, 20, — 16.

37. Écart-type : formule développée. Étant donné une série de n observations de valeurs $x_1, x_2, x_3, \dots x_n$, on désigne par $\bar{x}$ leur moyenne arithmétique.

1° Démontrer que :
$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2.$$

2° Déduire que l'écart-type de la série est :
$$\sigma = \left[\frac{1}{n}\sum x_i^2 - \bar{x}^2\right]^{\frac{1}{2}}.$$

Cette formule est dite formule développée qui se prête plus aisément au calcul que la formule de définition.

3° Déterminer, en utilisant la formule développée, l'écart-type de la distribution

6, 18, — 24, 8, 24, — 14, 4, 12, — 26, 6, 20.

38. Écart-type : formule dans le cas d'observations groupées en classes. — On considère une série d'observations groupées en classes de valeurs centrales $x_1, x_2, \ldots, x_p$ ayant pour effectifs correspondants $n_1, n_2, \ldots, n_p$ tel que $\sum_{i=1}^{p} n_i = n$. La valeur moyenne est $\bar{x}$.

1° Démontrer que : $$\sum_{i=1}^{p} n_i(x_i - \bar{x})^2 = \sum_{i=1}^{p} n_i x_i^2 - n\bar{x}^2.$$

2° Déduire que l'écart-type est $$\sigma = \left[\frac{1}{n}\sum_{i=1}^{p} n_i x_i^2 - \bar{x}^2\right]^{\frac{1}{2}}.$$

3° Déterminer, en utilisant la formule de définition et la formule développée, l'écart-type de la série :

Classes	0 à 10	10 à 20	20 à 30	30 à 40
Effectifs	4	8	12	16

39. Écart-type : changement de variable. — On considère une série de n observations, de valeurs $x_1, x_2, \ldots x_n$ de moyenne arithmétique $\bar{x}$. On pose :

$x_i = x_0 + k X_i$ où x_0 et k sont des constantes.

1° Démontrer la formule $\bar{x} = x_0 + k\bar{X}$, où $\bar{X}$ est la moyenne arithmétique des X_i.

2° Démontrer que : $$\frac{1}{n}\sum_{i=1}^{n}(x_i - x_0)^2 = \frac{k^2}{n}\sum_{i=1} X_i^2.$$

3° Si on désigne par σ_x la fluctuation de la série des x_i et σ_X la fluctuation de la série des X_i, démontrer que $\sigma_X = k\sigma_x$.

4° *Application* : Répartition des salariés d'une entreprise suivant leur âge.

Age	20 à 25	25 à 30	30 à 35	35 à 40	40 à 45	45 à 50	50 à 55	55 à 60
Effectif	16	53	80	85	35	16	7	8

On pourra poser $x_0 = 37{,}5$ et $k = 5$.

40. Le classement de certaines catégories des personnels de deux entreprises A et B, effectué d'après le montant des salaires horaires, est indiqué dans le tableau ci-dessous.

1° Calculer, pour chacun de ces groupes :

a) la médiane et donner sa signification;
b) la moyenne et donner sa signification;
c) l'écart quadratique moyen (écart-type).

2° Quelles réflexions vous inspire la comparaison des résultats obtenus :

a) pour les valeurs centrales (valeurs typiques);
b) pour la dispersion ?

3° Expliquer en quoi le développement du machinisme peut amener une modification de la répartition des salaires dans une entreprise.

Salaires F	3 à 3,30	3,30 à 3,60	3,60 à 3,90	3,90 à 4,20	4,20 à 4,50	4,50 à 4,80	4,80 à 5,10	5,10 à 5,40	5,40 à 5,70	5,70 à 6
Entreprise A	95	184	265	235	182	166	84	73	47	18
Entreprise B	54	126	214	387	476	624	581	419	154	66

41. Une entreprise qui exploite un parc de taxis a relevé, pour 100 d'entre eux, les distances qu'ils avaient parcourues au moment de leur mise à la réforme.

Distances parcourues (*en milliers de km*)	Nombre de taxis
80 — 85	5
85 — 90	9
90 — 95	14
95 — 100	18
100 — 105	25
105 — 110	16
110 — 115	7
115 — 120	6

1° Indiquer la distance médiane.
2° Calculer la distance moyenne.
3° Calculer l'écart-type.
4° Indiquer la signification de ces trois résultats.

42. On a pesé individuellement 1 000 cigarettes consécutives à la sortie d'une machine à cigarettes (environ 30 secondes de fabrication). Les poids individuels ont été répartis dans les classes d'étendue 2 centigrammes.

Ex. : classe 8 : $1,18 \leqslant x_8 < 1,20$.

Numéros des classes	Limites des classes	Nombre de cigarettes
	grammes	
1	De 1,04 à 1,06	17
2	De 1,06 à 1,08	15
3	De 1,08 à 1,10	27
4	De 1,10 à 1,12	47
5	De 1,12 à 1,14	63
6	De 1,14 à 1,16	85
7	De 1,16 à 1,18	117
8	De 1,18 à 1,20	129
9	De 1,20 à 1,22	126
10	De 1,22 à 1,24	112
11	De 1,24 à 1,26	88
12	De 1,26 à 1,28	69
13	De 1,28 à 1,30	42
14	De 1,30 à 1,32	31
15	De 1,32 à 1,34	16
16	De 1,34 à 1,36	16
	Total	1 000

1° *a*) Établir la relation qui donne la moyenne arithmétique $\bar{x}$ des poids des cigarettes en fonction d'une moyenne provisoire x_0 (choisie parmi les centres de classes), de l'intervalle de classe k constant et de la moyenne $\overline{U}$ des déviations des centres des classes de la série par rapport à x_0, l'intervalle de classe k étant pris comme unité $\left(\text{on posera } U_i = \dfrac{x_i - x_0}{k}\right)$.

b) Appliquer cette relation au calcul du poids moyen des cigarettes.

2° *a*) Établir de même la formule qui donne l'expression de la variance de la série.

b) Appliquer cette formule au calcul de l'écart-type.

c) Donner la valeur du coefficient de variation.

43. Étant donné la série statistique suivante :

Répartition de la population salariée suivant la distance du lieu de travail

Distance en km	A domicile	De 0 à 1	De 1 à 2	De 2 à 5	De 5 à 10	De 10 à 20	De 20 à 50	Total
Nombre d'ouvriers	29	324	159	255	147	59	27	1 000

établir la relation $$\sum (x - \bar{x})^2 = \sum (x - a)^2 - n(a - \bar{x})^2$$

x = terme quelconque de la série; a = nombre quelconque; n = nombre de termes.

Calculer l'écart-type σ de cette série en appliquant cette formule. On donnera à a la valeur pouvant réduire au maximum les calculs.

(*Baccalauréat*)