

STATISTIQUES

D) VOCABULAIRE :

- Une étude statistique s'effectue sur un ensemble appelé **population** dont les éléments sont appelés **individus** et consiste à observer et étudier un même aspect sur chaque individu, appelé **caractère**
- Chaque relevé est appelé une **donnée**.
- Les **valeurs** d'un caractère étudié sont les données différentes qu'il peut prendre.

STATISTIQUES

- Le caractère peut être **qualitatif** si les données ne sont pas des nombres : couleur, profession, ...
- Il est **quantitatif** si les valeurs sont des nombres : taille, masse, ...

Un caractère quantitatif peut être :

- **Discret** si les valeurs du caractère sont isolées (par exemple : nombre d'enfants). Les valeurs d'un caractère discret sont appelées les **modalités**.
- **Continu** si les valeurs du caractère sont regroupées en intervalles appelés **classes** (par exemple : taille $[1.60,1.70[$, $[1.70,1.80[$...).

STATISTIQUES

II) EFFECTIFS, FRÉQUENCES :

- Dans une série statistique, l'**effectif** d'une valeur est le nombre de fois où la valeur apparaît.
- Le **mode** de la série est la valeur du caractère ayant le plus grand effectif.
- L'**effectif total** est le nombre total de données
- On peut ranger les résultats dans un **tableau d'effectifs**.

STATISTIQUES

- La **fréquence** f d'une valeur se calcule en divisant l'effectif de cette valeur par l'effectif total.
- Les fréquences sont des nombres **compris entre 0 et 1**. On les exprime souvent en pourcentages en multipliant le quotient par 100.

	Ouvriers	Secrétaires	Cadres	Total
Femmes	13	4	3	20
Hommes	39	0	7	46
Total	52	4	10	66

Fréquence d'ouvriers : $\frac{52}{66} = \frac{26}{33} \approx 0,8$

Fréquence en pourcentages

De femmes dans l'entreprise : $\frac{20}{66} \times 100 \approx 30,3\%$

De cadres hommes $\frac{7}{66} \times 100 \approx 10,6\%$

Comparer le pourcentage de cadres parmi les hommes et

parmi les femmes : $H: \frac{7}{46} \times 100 \approx 15,2\%$
 $F: \frac{3}{20} \times 100 = 15\%$

Les pourcentages sont presque égaux.

STATISTIQUES

On peut ranger les valeurs d'une variable quantitative dans l'ordre croissant ou décroissant.

- l'effectif cumulé croissant (ECC) d'une valeur est la somme de l'effectif de cette valeur et des effectifs des valeurs précédentes.
- La fréquence cumulée croissante (FCC) d'une valeur est la somme de la fréquence de cette valeur et des fréquences de toutes les valeurs précédentes.
- On peut représenter une série statistique par une courbe des effectifs (ou des fréquences) cumulé(e)s croissant(e)s. Les valeurs sont sur l'axe des abscisses et les ECC ou FCC sont sur l'axe des ordonnées.

STATISTIQUES

Une société de services en informatique fait une analyse des temps d'utilisation devant un ordinateur ; elle réalise une enquête auprès d'un échantillon de 200 clients et obtient les résultats suivants :

Temps de connexion en h/an	[200;400[[400;600[[600;800[[800;1000[[1000;1200[[1200;1400[
Nombre d'utilisateurs	15	32	35	78	31	9
Fréquences (en%)						
FCC (en%)						

- Donner la classe modale : [800;1000[
- Compléter le tableau avec les fréquences en pourcentages

STATISTIQUES

Une société de services en informatique fait une analyse des temps d'utilisation devant un ordinateur ; elle réalise une enquête auprès d'un échantillon de 200 clients et obtient les résultats suivants :

Temps de connexion en h/an	[200;400[[400;600[[600;800[[800;1000[[1000;1200[[1200;1400[
Nombre d'utilisateurs	15	32	35	78	31	9
Fréquences (en%)	7,5	16	17,5	39	15,5	4,5
FCC (en%)						

- Compléter le tableau avec les fréquences cumulées croissantes en pourcentages

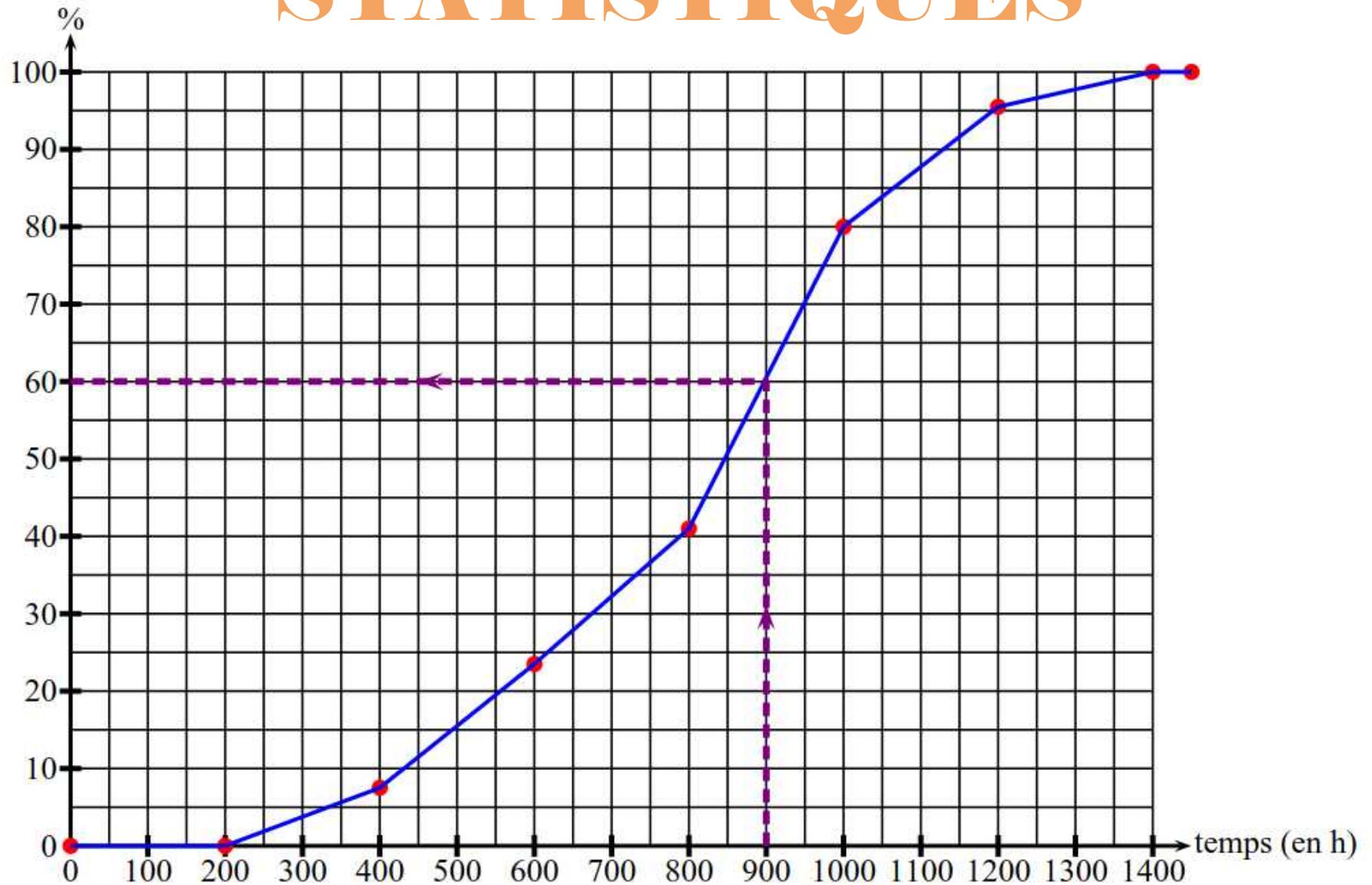
STATISTIQUES

Une société de services en informatique fait une analyse des temps d'utilisation devant un ordinateur ; elle réalise une enquête auprès d'un échantillon de 200 clients et obtient les résultats suivants :

Temps de connexion en h/an	[200;400[[400;600[[600;800[[800;1000[[1000;1200[[1200;1400[
Nombre d'utilisateurs	15	32	35	78	31	9
Fréquences (en%)	7,5	16	17,5	39	15,5	4,5
FCC (en%)	7,5	23,5	41	80	95,5	100

- Quel est le pourcentage d'utilisateurs qui se connectent au moins 1000 h/an ? $100 - 80 = 20 \%$

STATISTIQUES



- On détermine graphiquement que le nombre d'utilisateurs qui se connectent moins de 900 h/an est de 60 %

STATISTIQUES

- Caractéristiques d'une série statistique :
 - De position :
 - Moyenne pondérée :

Pour calculer une moyenne pondérée \bar{x} , on ajoute tous les produits des valeurs par leurs effectifs et on divise par l'effectif total.

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + \dots + x_p n_p}{n_1 + n_2 + \dots + n_p}$$

- Remarque : la fréquence de la valeur x_i est calculée par :

$$f_i = \frac{n_i}{N} \text{ alors } \bar{x} = f_1 x_1 + f_2 x_2 + \dots + f_p x_p$$

- Si les valeurs sont regroupées en classes, on prend le centre de chaque classe pour calculer la moyenne.

STATISTIQUES

- Médiane

La médiane notée Me d'une série statistique de N données, est le nombre qui partage cette série rangée dans l'ordre croissant en deux parties de même effectif.

- Si la série comporte un nombre impaire de données, la médiane est la donnée « du milieu »
- si la série comporte un nombre paire de données, la médiane peut être tout nombre compris entre les deux données « du milieu ». En pratique, on prend la demi-somme de ces deux données.

c'est la valeur qui est à la $\left(\frac{n+1}{2}\right)^{ème}$ position

STATISTIQUES

- Quartiles :
- Le premier quartile Q_1 est la plus petite valeur de la série telle que au moins 25% des données de la série soient inférieures ou égales à Q_1 .
- Le troisième quartile Q_3 est la plus petite valeur de la série telle que au moins 75% des données de la série soient inférieures ou égales à Q_3 .

STATISTIQUES

- Méthode :

On commence par classer les données de la série dans l'ordre croissant, on compte son effectif total N et on calcule $\frac{N}{4}$

- Si N est divisible par 4, le nombre trouvé donne le rang de la donnée qui est le premier quartile.
- Si N n'est pas divisible par 4, on prend la valeur approchée par excès du résultat comme rang du premier quartile.
- Pour le troisième quartile, on fait la même chose en calculant $\frac{3 \times N}{4}$ pour déterminer le rang.

STATISTIQUES

- Remarque :
- La médiane n'est pas obligatoirement une donnée de la série.
- Les quartiles sont toujours des données de la série.

STATISTIQUES

- De dispersion
 - L'étendue d'une série statistique est la différence entre la plus grande valeur et la plus petite valeur de cette série.
 - L'intervalle interquartile est l'intervalle $[Q1 ; Q3]$. La longueur de cet intervalle s'appelle l'écart inter-quartile et vaut $Q3 - Q1$.
- Étude d'une série statistique avec la calculatrice :
exercice n°2 pages 166-167