



STATISTIQUES

Introduction au vocabulaire des statistiques

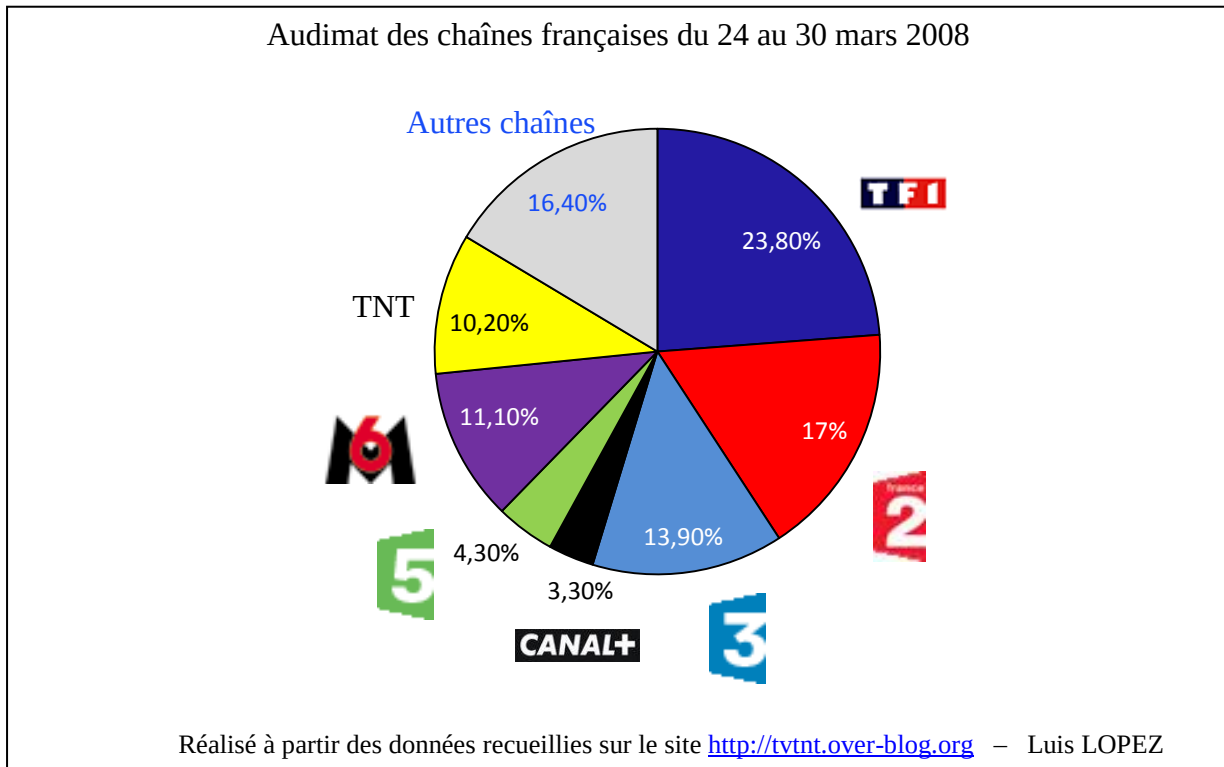
Les études statistiques s'appliquent à des ensembles d'éléments nommés **populations**.

Chaque élément de la population étudiée est **une unité statistique ou un individu**.

La partie de la population qui sert de support à l'étude est **l'échantillon**.

La propriété sur laquelle porte l'étude statistique est **le caractère ou variable statistique**.

Activité 1



Nom de ce graphique : **Ce graphique est un diagramme à secteurs circulaires**

Sujet d'étude : **L'étude porte sur les chaînes regardées par les français entre le 24 et 30 mars 2008.**

Population sur laquelle porte cette étude : **les téléspectateurs français du 24 au 30 mars 2008.**

Variable (ou caractère) de cette étude : **chaîne de télévision**

Propriété du caractère : **Le caractère étudié ici est non mesurable. On dit qu'il est qualitatif.**

Méthode pour construire ce graphique : **Les valeurs des angles sont calculées à partir des effectifs ou des fréquences**

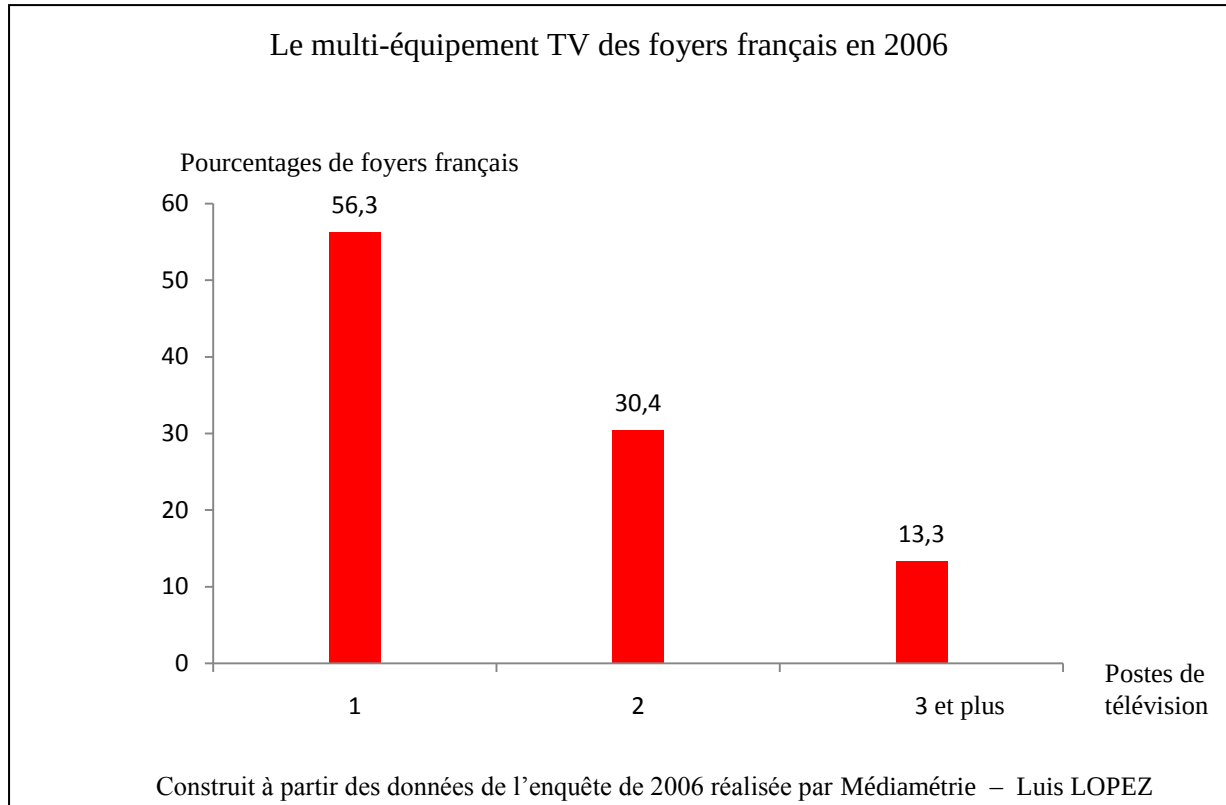
Calculer le pourcentage manquant. À quoi peut correspondre cette valeur ?

$100 - 23,80 - 17 - 13,90 - 3,30 - 4,30 - 11,10 - 10,20 = 16,40$ soit 16,40 %.

Ce pourcentage correspond à l'audimat de toutes les chaînes non mentionnées dans le graphique (chaînes à abonnement par exemple).



Activité 2



Nom de ce graphique : **Ce graphique est un diagramme à bâtons**

Additionner tous les pourcentages.

$$56,3 + 30,4 + 13,3 = 100 \text{ soit } 100 \%$$

Peut-on penser que tous les français disposent d'un poste de télévision ?

Comme 0 n'apparaît pas en abscisse, on peut supposer qu'on a ignoré ceux qui n'avaient pas de poste de télévision. On ne peut donc rien conclure.

Sujet d'étude : **L'étude porte sur le nombre de postes de télévision détenus par les foyers possédant au moins une TV en France en 2006**

Population sur laquelle porte cette étude : **les foyers français possesseurs de TV (en 2006)**

Variable (ou caractère) de cette étude : **nombre de poste de télévision**

Propriété du caractère : **Le caractère étudié ici est quantitatif (mesurable). Ce caractère ne peut prendre que des valeurs isolées (ici : entières et positives). On dit alors qu'il est discret.**

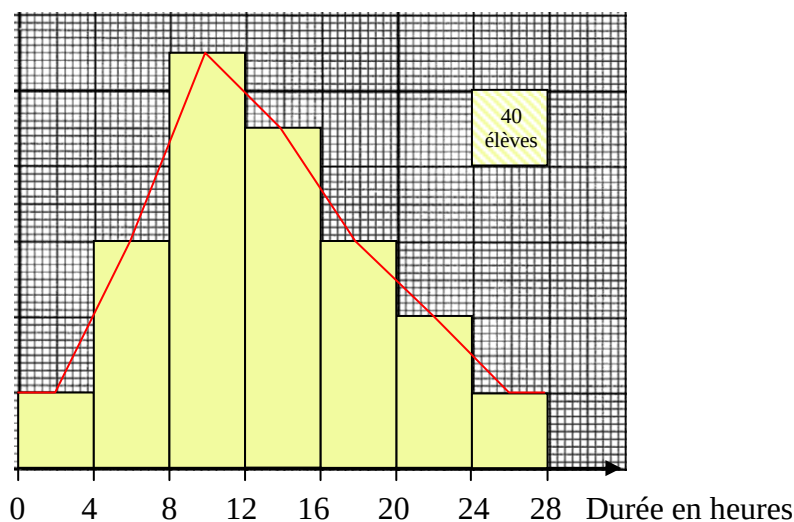
Méthode pour construire ce graphique : **Les hauteurs des bâtons sont proportionnelles aux effectifs ou aux fréquences.**

Mode (valeur du caractère correspondant au plus grand effectif) de cette étude statistique : **Le mode de cette étude statistique est 1.**



Activité 3

Temps consacré chaque semaine par les 800 élèves d'un lycée à regarder la télévision



Graphique réalisé par Luis LOPEZ

Nom de ce graphique : **Ce graphique est un histogramme.**

Sujet d'étude : **L'étude porte sur le temps consacré chaque semaine par les 800 élèves du lycée à regarder la télévision.**

Population sur laquelle porte cette étude : **les 800 élèves du lycée.**

Variable (ou caractère) de cette étude : **temps passé à regarder la télévision**

Propriété du caractère : **Le caractère étudié ici est quantitatif. Ce caractère peut prendre toutes les valeurs dans un intervalle donné appelé classe. On dit alors que le caractère est continu.**

Méthode pour construire ce graphique : **Les aires des rectangles sont proportionnelles aux effectifs.**

Mode de cette étude statistique : **La classe modale (classe correspondant au plus grand effectif) de cette étude statistique est [8 ; 12[. Le mode est le centre de cette classe : 10.**

On peut tracer sur le diagramme le polygone des effectifs.

Comparer la surface de l'histogramme et la surface incluse dans le polygone.

La surface incluse dans le polygone et celle de l'histogramme ont même aire.



Activité 4 : tableau statistique – calcul d’une moyenne

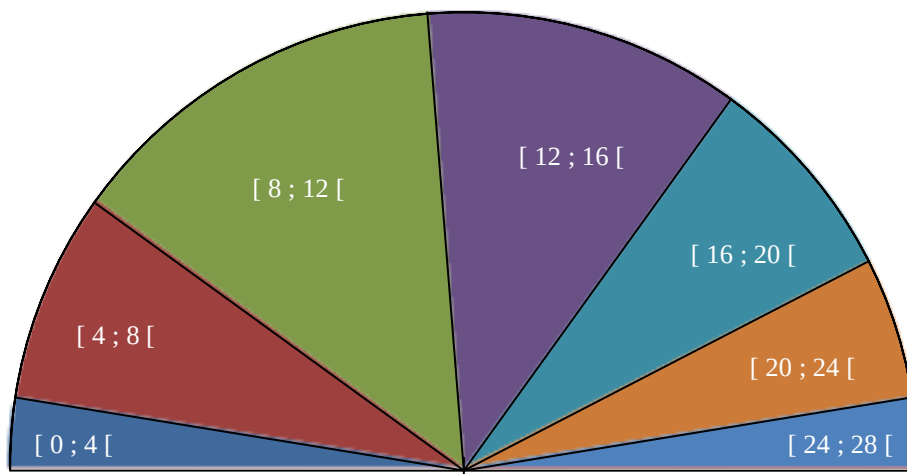
À partir de l’histogramme de l’activité 3, compléter le tableau suivant :

Classes	Effectifs n_i	Centres de classe x_i	Fréquence	Angle	$n_i \times x_i$
[0 ; 4 [	40	2	0,05	9	80
[4 ; 8 [	120	6	0,15	27	720
[8 ; 12 [	220	10	0,275	49,5	2 200
[12 ; 16 [	180	14	0,225	40,5	2 520
[16 ; 20 [	120	18	0,15	27	2 160
[20 ; 24 [	80	22	0,10	18	1 760
[24 ; 28 [	40	26	0,05	9	1 040
Total :	800		1	180	10 480

Détails des calculs de la première ligne :

Centre de classe	Fréquence	Angle
$\frac{0+4}{2} = 2$	$\frac{40}{800} = 0,05$	$0,05 \times 180 = 9$

Construire le diagramme à secteurs semi-circulaires correspondant à cette étude :



Calculer le temps moyen passé par les élèves devant le poste de télévision.

$$\bar{x} = \frac{10\,480}{800} = 13,1 \text{ soit } 13,1 \text{ heures ou } 13 \text{ heures et } 6 \text{ minutes.}$$

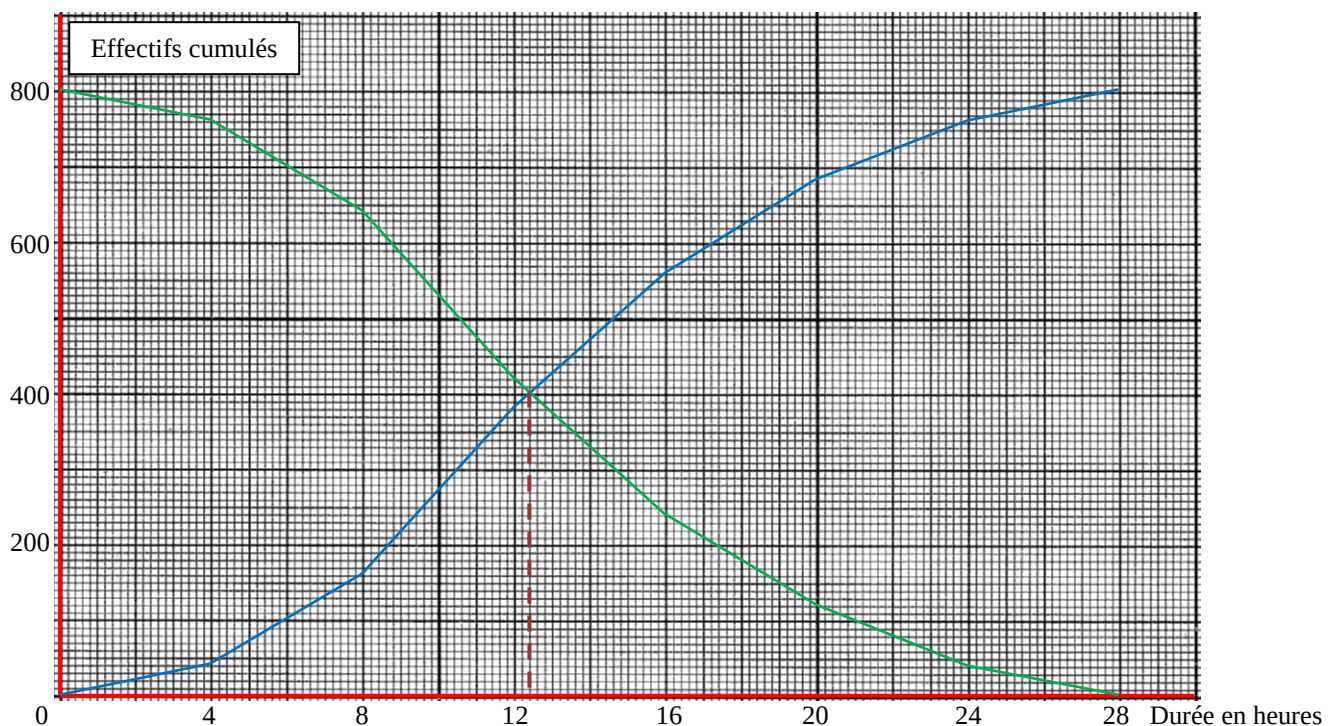


Activité 5 : calcul d'une médiane

En reprenant l'exemple utilisé dans les deux activités précédentes, compléter le tableau suivant :

Classes	Effectifs n_i	Effectifs cumulés croissants	Effectifs cumulés décroissants
[0 ; 4 [	40	40	800
[4 ; 8 [	120	160	760
[8 ; 12 [	220	380	640
[12 ; 16 [	180	560	420
[16 ; 20 [	120	680	240
[20 ; 24 [	80	760	120
[24 ; 28 [	40	800	40
Total :	800		

Construire le polygone des effectifs cumulés croissants et le polygone des effectifs cumulés décroissants dans le repère ci-dessous :



Que remarque-t-on ? L'ordonnée du point d'intersection correspond à 400 (moitié de 800).

Lire la valeur de la médiane (valeur pour laquelle il y a autant de valeurs inférieures que de valeurs supérieures). Vérifier par un calcul.

Graphiquement on lit : 12,4 heures.

Par le calcul, on résout $\frac{x-12}{16-12} = \frac{400-380}{560-380}$ soit $\frac{x-12}{4} = \frac{20}{180}$

ce qui nous donne $x-12 = \frac{80}{180}$ d'où $x = \frac{4}{9} + 12$. En arrondissant au dixième : $x \approx 12,4$



Activité 6 : tableau statistique – paramètres de dispersion

En reprenant l'exemple précédent, compléter le tableau suivant :

Classes	Effectifs n_i	Centres de classe x_i	$n_i \times x_i^2$
[0 ; 4 [	40	2	160
[4 ; 8 [	120	6	4 320
[8 ; 12 [	220	10	22 000
[12 ; 16 [	180	14	35 280
[16 ; 20 [	120	18	38 880
[20 ; 24 [	80	22	38 720
[24 ; 28 [	40	26	27 040
Total :	800		166 400

Étendue

Calculer l'étendue de cette série statistique (différence entre la valeur maximale et minimale) :

L'étendue est : $28 - 0 = 28$ soit 28 heures.

Variance

La variance permet de mesurer la dispersion des valeurs autour de la moyenne.

Elle est obtenue par la formule $V = \frac{1}{N} \sum_{i=1}^{i=p} n_i (x_i - \bar{x})^2 = \left(\frac{1}{N} \sum_{i=1}^{i=p} n_i x_i^2 \right) - \bar{x}^2$

Calculer la variance dans notre exemple.

$V = \left(\frac{1}{800} \times 166\,400 \right) - 13,1^2 = 36,39$. La valeur de la variance est 36,39.

Écart-type

L'écart-type σ est un autre indicateur pour évaluer la dispersion des valeurs. Il est obtenu en calculant la racine carrée de la variance. Calculer l'écart-type dans notre exemple.

$\sigma = \sqrt{36,39}$ soit environ 6,03. La valeur de l'écart-type est 6,03 heures.

Capabilité machine : application au secteur industriel

La capabilité d'un processus, ou d'une machine, est l'aptitude de ce processus, ou de cette machine, à fabriquer des pièces bonnes. En appelant T_s et T_i respectivement limite supérieure et limite inférieure de la tolérance :

$$C_{m_k} = \frac{T_s - \bar{x}}{3\sigma} \text{ et } C_{m_k} = \frac{\bar{x} - T_i}{3\sigma}$$

L'aptitude de capabilité est reconnue bonne si chacun de ces deux facteurs est égal ou supérieur à 1.