

## SERIE 1

### Statistique descriptive - Graphiques

#### Collecte de l'information, dépouillement de l'information et vocabulaire

La collecte de l'information peut être :

- **directe**: sondages
- **indirecte**: on utilise des données existantes (bilans, ...)

Cette collecte doit être objective avec suffisamment de données mais sans excès pour rester utilisable.

L'ensemble examiné est appelé la **population** (pas nécessairement des gens). Par exemple : des personnes, des objets, des lieux, des moments, ... **L'individu** est un l'élément de la population. On s'intéresse alors à son **caractère** qui est la propriété subjective qui nous intéresse (âge, sexe, nombre d'enfants, taille, ...)

On distingue différents types de **caractère** :

- qualitatif (par exemple catégorie professionnelle, couleur d'un objet, parti politique)
- quantitatif discret<sup>1</sup> (par exemple nombre d'enfants ou note trimestrielle)
- quantitatif continu<sup>2</sup> (par exemple longueurs, vitesses des particules d'un gaz, ...)

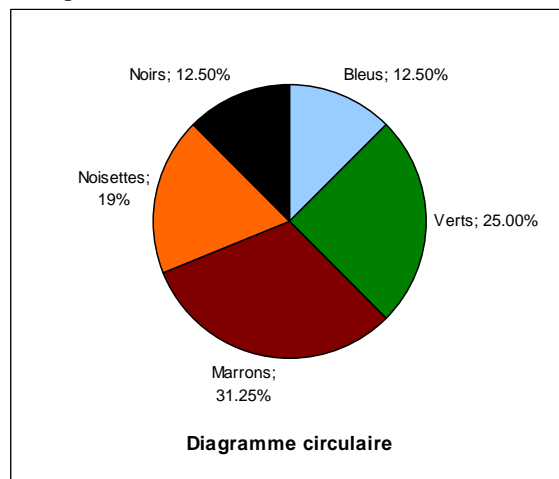
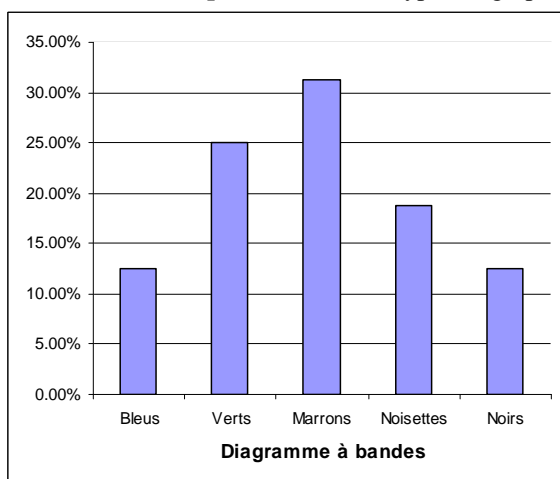
Dans le cas quantitatif, on parle aussi de **variables statistique**, plutôt que de caractère.

On reporte ensuite les caractères suivant une **partition** (partage d'un ensemble en sous-ensembles disjoints et exhaustifs). Il est bien sûr absolument indispensable que le caractère de chaque individu observé puisse être reporté de façon claire et univoque dans un et un seul sous-ensemble de la partition. Exemples de partitions: sexe, âge (en années), tranches de revenu etc.

La **population** est donc l'ensemble des **individus** sur lesquelles porte une étude. Ses individus sont classés selon un ou plusieurs **caractères** suivant des partitions qui sont des **classes**.

#### Visualisation des données, effectifs et fréquences

Il est possible de passer directement à la partie calculatoire, mais il est néanmoins appréciable de visualiser ces données. La représentation graphique fait également partie de la statistique descriptive. Les données brutes sont préalablement regroupées<sup>3</sup> et mises sous forme de tableau triée en **classes exhaustives** (partition), d'amplitudes à choix, dont on répertorie l'**effectif** ou la **fréquence**. Les deux types de graphiques les plus courants sont :



#### Définition :

La **fréquence** (relative) d'une classe est définie par :  $f_{classe} = \frac{\text{Effectif de la classe}}{\text{Effectif total}}$   
 exprimée généralement en %.

<sup>1</sup> variable numérique

<sup>2</sup> il peut prendre toute valeur dans un certain intervalle réel.

<sup>3</sup> Le fait qu'il y a un choix implique nécessairement une certaine subjectivité.

### Exemple 1 :

La **population** est l'ensemble des vélos produits par CIPEDVELO en 2001.

L'étude porte sur le type de vélos (VTT, ville, course, ...).

Le **caractère** est le type de vélos et les **classes** sont : VTT, course, ville, ...

Un **individu** est un vélo.

### Exemple 2 :

Une enquête portant sur la couleur des cheveux de 15 personnes a donné les résultats bruts suivants :

« Roux, Bruns, Bruns, Bruns, Bruns, Blonds, Blonds, Blonds, Blonds, Blonds, Noirs, Noirs, Noirs, Châtains, Châtains »

La **population** est : les 15 personnes

Un **individu** est : une personne

Le **caractère qualitatif** est : la couleur des cheveux

Les **classes** sont : Roux, Blonds, Bruns, Noirs et Châtains

La première étape consiste à ranger ces résultats bruts dans un tableau :

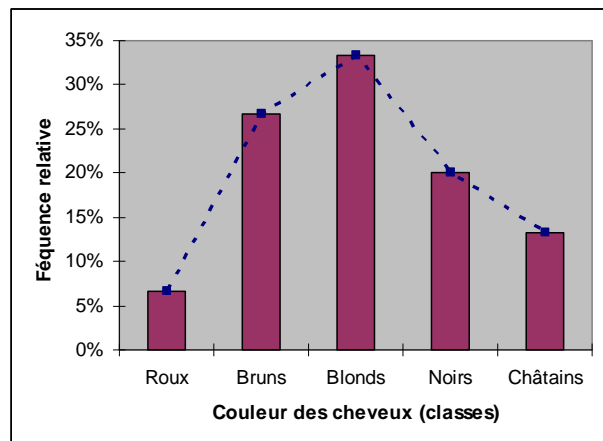
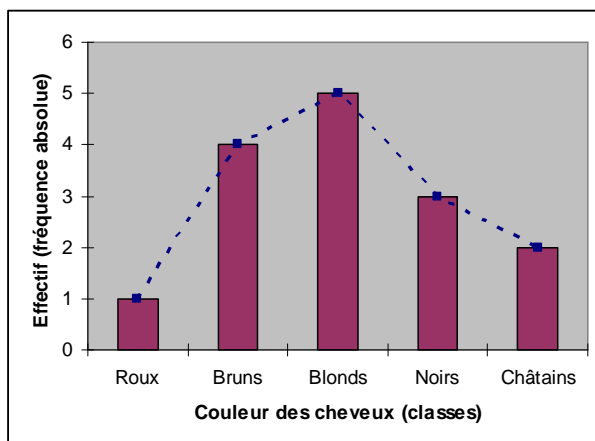
Couleur des cheveux	Effectif (fréquence absolue)	Fréquence relative	Fréquence relative en %
Roux	1	0,067	6,7 %
Bruns	4	0,267	26,7 %
Blonds	5	0,333	33,3 %
Noirs	3	0,200	20,0 %
Châtains	2	0,133	13,3 %
<b>TOTAL</b>	<b>15</b>	<b>1</b>	<b>100 %</b>

On remarque que 4 personnes sur 15 ont des cheveux bruns.

Pour la classe des blonds, la **fréquence** se calcule de la manière suivante :

$$f_{\text{blonds}} = \frac{5}{15} = 0,333 = 33,3\%$$

On peut alors donner la représentation graphique avec un **diagramme à bande (ou histogramme)** de deux manières :



Parfois on représente aussi le **polygone des fréquences** (ligne brisée qui joint les extrémités des bâtons).

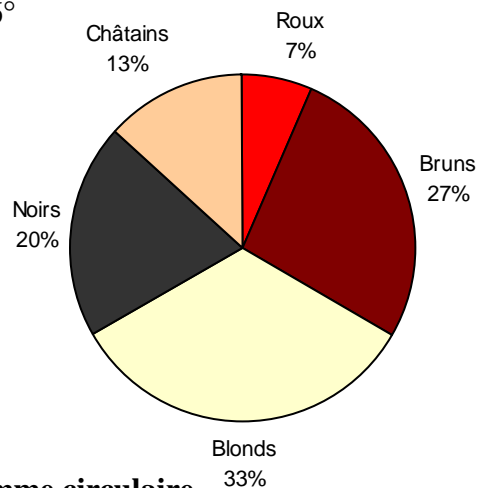
Pour le **diagramme circulaire**, il faut préalablement calculer l'angle de chaque "tranche" à l'aide d'une règle de trois.

Pour la classe des marrons :  $\frac{5}{16} = \frac{x}{360^\circ} \Rightarrow x = 112,5^\circ$

Remarque : Pour aller plus vite, on peut aussi utiliser la règle suivante :  
 « Si la fréquence de la classe vaut  $f$  en %, alors  $\boxed{\text{angle} = f \cdot 3,6}$  »

Pour la classe des cheveux bruns : l'angle =  $26,7 \cdot 3,6 = 96^\circ$

Couleur des cheveux	Effectif (fréquence absolue)	Fréquence relative en %	Angle
Roux	1	6,7 %	24°
Bruns	4	26,7 %	96°
Blonds	5	33,3 %	120°
Noirs	3	20,0 %	72°
Châtains	2	13,3 %	48°
<b>TOTAL</b>	<b>15</b>	<b>100 %</b>	<b>360°</b>



**Diagramme circulaire**

Exemple 3 :

25 élèves ont passé un examen. Voici, en vrac, leurs résultats :

5 3 5 5 4 4 4 5 4 5 1 5 3  
 2 3 3 4 4 6 5 6 4 3 5 2

La **population** est : les 20 élèves  
 Un **individu** est : un élève  
 Le **caractère quantitatif discret** est : la note de l'examen  
 Les **classes** sont : 1 ; 2 ; 3 ; 4 ; 5 ; 6 (variables statistiques)

On peut ranger ces résultats de la manière suivante :

Note $x_i$	Effectif $n_i$
1	1
2	2
3	5
4	7
5	8
6	2



Les **variables statistiques** sont  $x_1 = 1, x_2 = 2, \dots, x_5 = 5, x_6 = 6,$

Et dans chaque cas on a l'**effectif de chaque classe** :

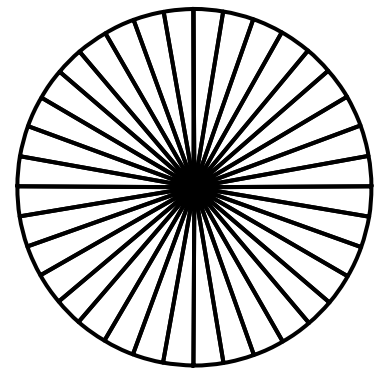
$n_1 = 1, n_2 = 2, \dots, n_5 = 8, n_6 = 2,$

Illustrer ces résultats à l'aide d'un diagramme à bande.

Exercice 1 :

- a) Compléter le tableau ci-dessous
- b) Illustrer ce tableau par deux diagrammes différents.

Type de vélo vendu	Effectif	Fréquence
VTT	53	
Ville	45	
Course	32	
Enfant	47	
Tricycle	13	

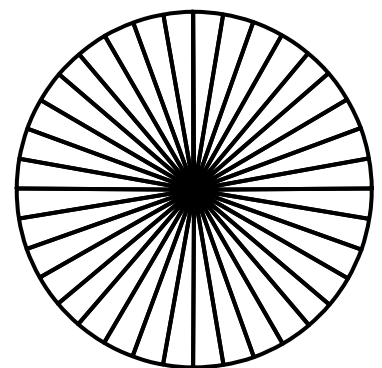


Exercice 2 :

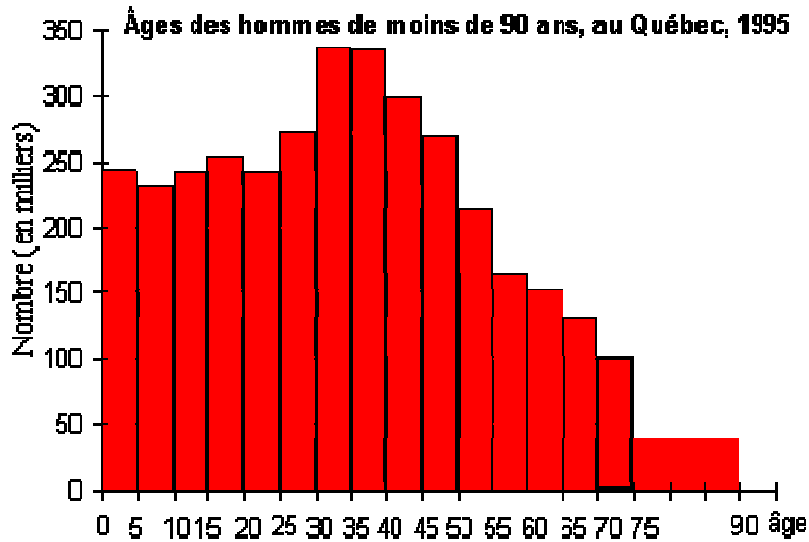
18 élèves ont passé un examen. Voici, en vrac, leurs résultats :

5    3    5    5    4    4    4    5    4  
2    3    3    4    4    6    5    6    4

Ranger ces résultats dans un tableau et illustrer-les à l'aide des différents diagrammes.



Exercice 3 :

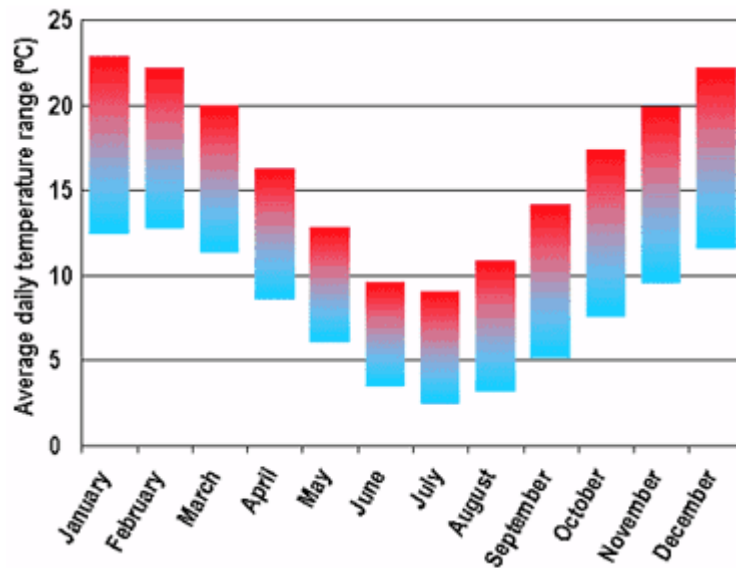


a) Quel pourcentage de la population des hommes a entre 30 et 45 ans ?

b) Quel pourcentage de la population des hommes a moins de 20 ans ?

Exercice 4 :

**Climat dans la zone de Katoomba (Australie)  
Étendue de la température moyenne pour chaque mois**



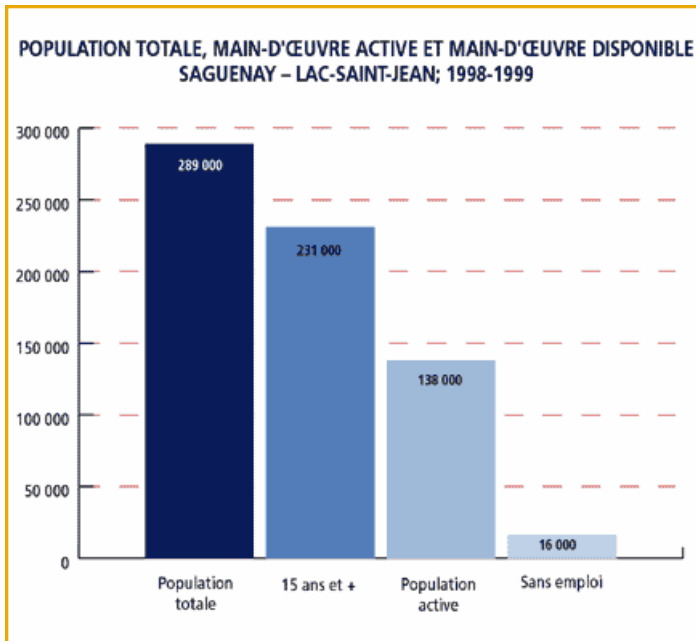
c) Quelle est la température la plus basse sur l'année et quand ? .....

d) Sur combien de mois par année la température atteint les 20° C ou plus ? .....

e) Quelle est la moyenne des températures maximales ? .....

Et minimales ? ..... Quelle est la moyenne annuelle ? .....

Exercice 5 :



- a) Quelle est la population non active ?
- b) Quelle est la population de moins de 15 ans ?
- c) Quelle est la population de 15 ans ou plus qui est probablement en formation scolaire ?