1. Définitions importantes

A. Population

L'ensemble des personnes ou des objets sur lesquels porte une étude statistique.

Exemple : L'ensemble des élèves du Collège Regina Assumpta.

B. Échantillon

Sous-ensemble de la population.

Exemples : 1) *Les élèves de 5e secondaire.*
2) *Les élèves de sexe masculin.*

C. Taille d'une population ou d'un échantillon

Nombre d'éléments qui composent la population ou l'échantillon. On représente par N la taille d'une population et par n la taille d'un échantillon.

Exemple : $N = 2200$ *élèves* $n = 900$ *garçons*

D. Variable ou caractère d'une étude statistique et modalités

La variable ou le caractère d'une étude statistique est l'élément(s) sur lequel (lesquels) porte l'étude statistique.

Exemples : *nombre de frères ou soeurs (0, 1, 2, 3, ...)*
couleur des yeux (bleu, vert, brun, noir, ...)
nombre d'heures de sommeil par nuit (8,5; 6,25; 7,2; 9; 8,333...)

Les modalités d'une variable sont les valeurs possibles de la variable.

E. Les types de variables ou de caractères

On distingue 2 types de variables : **qualitatif** et **quantitatif.**

Variable *qualitative* : Les données recueillies sont des **mots** ou des **codes**.

Variable *quantitative* : Les données recueillies sont des **nombres** qui représentent une quantité (compter, dénombrer).

Il est possible de distinguer deux types de variables quantitatives : **discrètes** et **continues**.

Les données recueillies pour une **variable quantitative continue** sont susceptibles de prendre n'importe quelles valeurs réelles dans un intervalle.

Les données recueillies pour une **variable quantitative discrète** sont des valeurs isolées les unes des autres et qui sont généralement entières.

Exemple : Dans les situations suivantes, donne des valeurs possibles et identifie le type de caractère

Caractère	Valeurs possibles	Type de caractère
a) L'argent que l'on a en poche.		*quantitatif discret*
b) La taille d'un individu.		*quantitatif continu*
c) Le nombre de cheveux sur la tête d'une personne.		*quantitatif discret*
d) La couleur préférée des élèves.		*qualitatif*
e) La superficie d'un terrain.		*quantitatif continu*
f) Le nombre de parties gagnées par une équipe sportive durant la saison.		*quantitatif discret*

Bref, pour vérifier si un caractère est quantitatif discret ou continu, on doit se poser la question suivante : **« Est-ce que je peux fractionner les valeurs du caractère autant de fois que je le veux ? »**

F. Le sondage, l'enquête et le recensement

Le **recensement** est une étude statistique menée sur l'ensemble de la population concernée par l'étude.

Le **sondage** est une étude statistique menée sur un échantillon.

G. Distribution statistique

Une distribution statistique à une variable est présentée par un tableau, appelé tableau de distribution, qui donne la répartition des n données de l'échantillon selon les modalités de la variable.

Exemple : On a relevé le nombre de buts marqués au hockey lors des 20 (n) premières parties de hockey de la saison par les Canadiens de Montréal.

2	3	4	3	1	0	2	3	2	4
3	4	2	1	3	2	4	3	3	3

La variable « nombre de buts » prend les valeurs (modalités) 0, 1, 2, 3 et 4. Le tableau ci-contre indique la répartition des 20 parties de hockey selon la variable « nombre de buts ». Complétez le tableau.

Distribution des parties selon le nombre de buts

Nombre de buts (x_i)	Nombre de parties (n_i)
0	*1*
1	*2*
2	*5*
3	*8*
4	*4*
Total	*20*

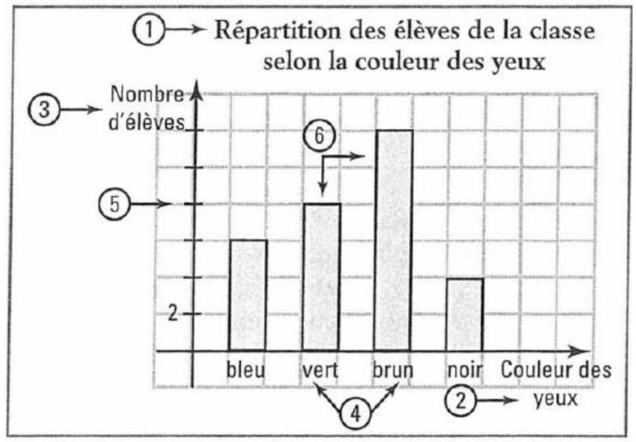
H. Diagramme à bandes

Un diagramme à bandes permet d'illustrer une variable **qualitative** ou **quantitative discrète**.

Exemple : Complétez le tableau de distribution représenté par le diagramme à bandes.

Répartition des élèves de la classe selon la couleur des yeux

Couleur des yeux	Nombre d'élèves
bleu	*6*
vert	*8*
brun	*12*
noir	*4*
Total	*30*



Les principaux éléments sont :

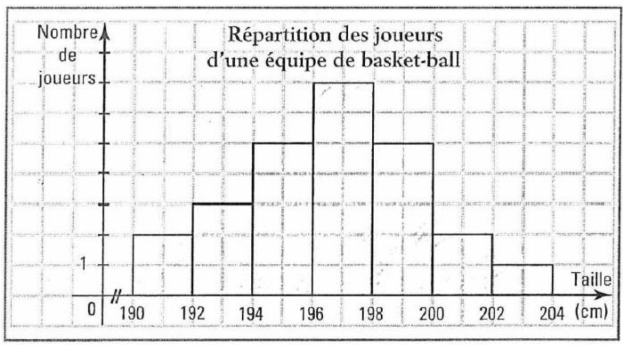
1- le titre ;
2- l'identification de l'axe horizontal : la variable « couleur des yeux » ;
3- l'identification de l'axe vertical : les fréquences « nombre d'élèves » ;
4- l'identification des bandes (les modalités de la variable) : bleu, vert, … ;
5- la graduation de l'axe vertical : l'échelle utilisée tient compte des fréquences ;
6- les bandes ont toutes la même largeur et sont également espacées. La hauteur de chaque bande est proportionnelle à la fréquence.

Dans un diagramme à bandes, les bandes peuvent être représentées **verticalement** ou **horizontalement**.

I. Histogramme

- L'histogramme représente une distribution où les données sont principalement quantitatives continues.
- Diagramme formé de bandes rectangulaires collées les unes sur les autres.
- La largeur des bandes est toujours identique dans l'histogramme.
- Les données sont regroupées par classe ou par intervalles de même longueur et placées sur l'axe horizontal.
- La hauteur des bandes, lue sur l'axe vertical, est en fonction de l'effectif, c'est-à-dire en fonction du nombre d'éléments de chaque classe.
- Un titre présente toujours l'histogramme.
- Les deux axes sont bien identifiés et gradués, c'est-à-dire que les bonds sont réguliers sur chacun des axes.
- Il est possible de faire une coupure d'axe seulement sur l'axe horizontal.

Exemple : L'histogramme ci-contre illustre la répartition des joueurs d'une équipe de basket-ball selon la taille (en cm).



On observe que :
- les joueurs ont une taille variant entre 190 cm et 204 cm.
- La classe [196, 198[renferme le plus grand nombre de joueurs.

2. Les mesures de tendance centrale

Lorsque vient le moment d'analyser un ensemble de données recueillies lors d'une étude statistique, il est naturel de s'intéresser aux données qui représentent le mieux le *centre* de la distribution. Par contre, ce centre peut être vu de différentes manières. Nous verrons trois mesures de tendance centrale : la moyenne, le mode et la médiane.

(A) *(B)* *(C)*

A. Moyenne arithmétique

Lorsqu'on désire représenter une série de données par un seul nombre, comme les notes d'un examen, la première mesure à laquelle on pense est la moyenne des données. La moyenne est la mesure de tendance centrale la plus connue et la plus utilisée comme représentante des données d'une série statistique. Avec la mise en situation suivante, nous apprendrons à représenter graphiquement une moyenne et à la calculer de trois façons différentes : avec les données brutes, avec les effectifs du tableau de distribution et avec les pourcentages du tableau de distribution.

Prenons la série statistique donnant le nombre de programmes préuniversitaires offerts dans chacun des 48 collèges publics du Québec en 2006 :

4	3	5	5	3	4	5	7	5	5	6	6
6	3	3	3	5	5	6	4	5	5	4	6
4	5	3	5	4	4	4	4	7	4	3	4
6	7	4	3	7	4	3	6	4	7	3	5

Source : Ministère de l'Éducation, Direction de la recherche, des statistiques et des indicateurs, Système prévisionnel SIPEEC, 2006.

i. Calcul de la moyenne avec des données brutes

Comme nous le savons, le calcul de la moyenne avec des données brutes consiste à additionner l'ensemble de ces données, puis à diviser la somme obtenue par le nombre total de données.

Moyenne avec des données brutes
$\text{Moyenne} = \dfrac{\text{somme des données}}{\text{nombre total de données}}$

Écriture symbolique :

Nous noterons la moyenne par le symbole μ, qui se lit « mu » (*m* dans l'alphabet grec), et le nombre de données par la lettre *N.*

En représentant chaque donnée de la série statistique par les symboles : x_1, x_2, x_3, x_4 et ainsi de suite, on obtient la formule suivante pour décrire le calcul d'une moyenne avec les données brutes :

$$\mu = \frac{x_1 + x_2 + x_3 + \cdots + x_N}{N}$$

On peut simplifier cette formule à l'aide de la notation sigma, symbolisée par « Σ » (S dans l'alphabet grec). Ce symbole indique que l'on doit faire la somme de tous les termes de forme x_i, l'indice i variant de 1 à *N*.

$$\mu = \frac{\Sigma x_i}{N}$$

Dans le cadre d'un sondage, on emploie le symbole $\overline{x}$ pour désigner la moyenne des données de l'échantillon et μ pour la moyenne des données de la population. De même, on utilise le symbole n pour désigner le nombre de données de l'échantillon et N pour le nombre de données de la population. La formule pour obtenir la moyenne des données de l'échantillon s'écrit donc :

$$\overline{x} = \frac{\Sigma x_i}{n}$$

Pour la moyenne des données de la mise en situation, on a :

$$\mu = \frac{\Sigma x_i}{N} = \frac{4+3+5+\cdots+3+5}{48} = \frac{223}{48} \approx 4{,}6 \text{ programmes préuniversitaires par collège}$$

Interprétation : En 2006, si tous les collèges publics du Québec avaient offert le même nombre de cours, il y aurait eu 4,6 programmes préuniversitaires dans chaque collège.

Attention!!: Le résultat du calcul d'une moyenne brute ne doit pas être arrondi à l'entier sous prétexte que les données brutes sont entières : la moyenne est un nombre théorique. Nous conviendrons de conserver une décimale après la virgule.

ii. Calcul de la moyenne avec les effectifs du tableau de distribution

Répartition des 48 collèges publics du Québec selon le nombre de programmes préuniversitaires offerts, 2006

Nombre de programmes préuniversitaires	Nombre de collèges	Pourcentage de collèges
3	10	20,8 %
4	14	29,2 %
5	12	25,0 %
6	7	14,6 %
7	5	10,4 %
Total	48	100,0 %

Avec les effectifs du tableau de distribution, on calcule ainsi la moyenne du nombre de programmes préuniversitaires offerts par les collèges publics du Québec :

Moyenne avec des effectifs
$\text{Moyenne} = \frac{\text{somme des produits de chaque valeur de la variable par son effectif}}{\text{nombre total de données}}$

Écriture symbolique :

En notant par la lettre n_i l'effectif correspondant à la valeur x_i, on obtient la formule suivante : $\mu = \frac{x_1 n_1 + x_2 n_2 + x_3 n_3 + \cdots + x_k n_k}{N}$, où k représente le nombre de valeurs de la variable.

En utilisant la notation sigma, on obtient :

$$\mu = \frac{\Sigma x_i n_i}{N}, \text{ pour } i \text{ variant de 1 à } k.$$

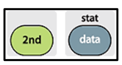
Dans l'exemple, on obtient donc :

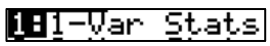
$$\mu = \frac{\Sigma x_i n_i}{N} = \frac{3 \times 10 + 4 \times 14 + 5 \times 12 + 6 \times 7 + 7 \times 5}{48} = \frac{223}{48} \approx 4{,}6$$

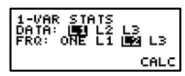
Entrer les données dans la calculatrice TI-30XS MultiView

1- 

: Entrer les modalités dans la première colonne et les effectifs dans la deuxième colonne

2- 

→ 

→ 

iii. Calcul de la moyenne avec les pourcentages du tableau de distribution

Pour trouver une formule qui permettrait de calculer une moyenne avec les pourcentages, reprenons le calcul de la moyenne avec les effectifs et apportons les modifications suivantes :

$$\mu = \frac{3\times10 + 4\times14 + 5\times12 + 6\times7 + 7\times5}{48} = 4{,}6$$

$$\mu = \frac{3\times10}{48} + \frac{4\times14}{48} + \frac{5\times12}{48} + \frac{6\times7}{48} + \frac{7\times5}{48} = 4{,}6$$

$$\mu = 3\times\frac{10}{48} + 4\times\frac{14}{48} + 5\times\frac{12}{48} + 6\times\frac{7}{48} + 7\times\frac{5}{48} = 4{,}6$$

$$\mu = 3\times20{,}8\,\% + 4\times29{,}2\,\% + 5\times25\,\% + 6\times14{,}6\,\% + 7\times10{,}4\,\% = 4{,}6$$

Moyenne avec des pourcentages
Moyenne = somme des produits de chaque valeur de la variable par son pourcentage

Attention!!: Il est à remarquer que nous n'avons pas à diviser, dans ce cas-ci, par le total des données, car cette division a déjà été faite dans le calcul du pourcentage.

Écriture symbolique

En notant f_i la fréquence relative correspondant à la valeur x_i, on obtient la formule suivante :

$$\mu = x_1 f_1 + x_2 f_2 + x_3 f_3 + \cdots + x_k f_k = \Sigma x_i f_i$$

pour i variant de 1 à k, où k est le nombre de valeurs de la variable.

Dans l'exemple, on obtient donc :

$$\mu = \Sigma x_i f_i = 3 \times 20{,}8\ \% + 4 \times 29{,}2\ \% + 5 \times 25\ \% + 6 \times 14{,}6\ \% + 7 \times 10{,}4\ \% \approx 4{,}6$$