

## Chapitre 2 : Analyse de l'information chiffrée

### I. Analyse croisée de deux caractères

#### 1. Tableau croisé d'effectifs

Lorsqu'on étudie une population, on souhaite souvent déterminer si des liens existent entre les différents caractères des individus qui composent cette population

Exemple : on peut étudier le lien entre le salaire et le genre des employés d'une entreprise, ou bien entre le rhésus et le groupe sanguin d'un individu comme dans le tableau ci-dessous.

Numéro individu	Groupe sanguin	Rhésus
1	O	négatif
2	O	positif
3	A	positif
4	O	négatif
5	O	positif
6	A	négatif
7	O	négatif
8	AB	positif
9	A	positif
10	B	positif

**Définition :** Un tableau croisé d'effectifs (ou tableau à double entrée) est un tableau dans lequel les valeurs d'un caractère sont présentées en ligne et celles de l'autre caractère sont présentées en colonne.

A l'intersection d'une ligne et d'une colonne, on indique le nombre d'individus présentant simultanément la valeur du premier caractère correspondant à la ligne, et la valeur du deuxième caractère correspondant à la colonne.

L'**effectif marginal** d'une valeur est le nombre d'individus de la population présentant cette valeur du caractère.

Exemple : Si l'on reprend le tableau précédent, on peut obtenir le tableau croisé d'effectifs ci-dessous.

Groupe Rhésus	A	B	AB	O	Total
Positif	2	1	1	2	6
Négatif	1	0	0	3	4
<b>Total</b>	3	1	1	5	10

L'effectif marginal de la valeur A est 3, ce qui signifie que, dans la population étudiée, 3 individus sont de groupe sanguin A.

## 2. Utilisation du tableur

Les données concernant une population peuvent être regroupées dans un fichier texte au format CSV.

Ce format très répandu permet de représenter des données sous forme de valeurs généralement séparées par des virgules ou des points virgules.

La première ligne du fichier CSV contient les descripteurs (le nom des caractères) des données contenues dans ce fichier comme dans l'exemple ci-contre.

```
N° individu;Groupe;Rhésus
1;O;négatif
2;O;positif
3;A;positif
4;O;négatif
5;O;positif
6;A;négatif
7;O;négatif
8;AB;positif
9;A;positif
10;B;positif
```

Pour traiter directement les données d'un fichier CSV, on peut faire appel à un tableur en double-cliquant sur le fichier ou :

- avec Excel, on utilise le menu « Données », puis « Convertir » : dans l'assistant conversion, on coche ensuite « Délimité, puis « Point-virgule » (si c'est le cas) ;
- avec OpenOffice, quand on charge un fichier CSV, la boîte de dialogue « Import de texte » s'ouvre : on coche « Point-virgule » (si c'est le cas).

On obtient alors le tableau ci-contre :

La recherche d'une information dans un fichier CSV contenant plusieurs centaines voire plusieurs milliers de lignes peut être automatisée à l'aide d'outils logiciels comme le tableur.

Le tableur permet de construire un **tableau croisé dynamique** : après avoir sélectionné la plage entière des données en prenant les en-têtes de colonnes, il suffit d'aller dans le menu « Données » et de sélectionner tableau croisé. (Voir les tutoriels mis sur Pronote).

	A	B	C
1	N° d'individu	Groupe	Rhésus
2	1	O	négatif
3	2	O	positif
4	3	A	positif
5	4	O	négatif
6	5	O	positif
7	6	A	négatif
8	7	O	négatif
9	8	AB	positif
10	9	A	positif
11	10	B	positif

On obtient alors le tableau ci-dessous :

	A	B	C	D	E	F
1	Filtre					
2						
3	NB - Groupe	Groupe				
4	Rhésus	A	AB	B	O	Total Résultat
5	positif	2	1	1	2	6
6	négatif	1			3	4
7	Total Résultat	3	1	1	5	10

Le tableur possède également des fonctions logiques préprogrammées (OU, ET, NON, SI et NB.SI) qui permettent de sélectionner ou de dénombrer des données selon certains critères :

Fonction	Syntaxe	Commentaires
<b>OU</b>	=OU(C1;C2)	La fonction <b>OU</b> renvoie VRAI si au moins une des deux conditions C1 ou C2 est vérifiée, et renvoie FAUX dans le cas contraire.
<b>ET</b>	=ET(C1;C2)	La fonction <b>ET</b> renvoie VRAI si les deux conditions C1 et C2 sont vérifiées, et renvoie FAUX dans le cas contraire.
<b>NON</b>	=NON(C)	La fonction <b>NON</b> renvoie VRAI si la condition C est fausse, et renvoie FAUX lorsque C est vraie.
<b>SI</b>	=SI(C;valeur1;valeur2)	La fonction <b>SI</b> renvoie la valeur 1 lorsque la condition C est vraie, et renvoie la valeur 2 lorsque C est fausse.
<b>NB.SI</b>	=NB.SI(zone;critères)	La fonction <b>NB.SI</b> renvoie le nombre de cellules de la plage de cellules zone qui satisfont les critères.

Exemple

- Un donneur universel est de groupe sanguin O et de rhésus négatif. Pour signaler à chaque ligne de la colonne D si l'individu est « donneur universel », on entre en D2 la formule =ET(B2="O";C2="négatif"), puis on la recopie vers le bas.
- En saisissant la formule =NB.SI(D2:D11;VRAI) dans la cellule D12, on obtiendra le nombre de donneurs universels en explorant la plage de cellules D2:D11.

	A	B	C	D
1	N° d'individu	Groupe	Rhésus	Donneur universel
2	1	O	négatif	VRAI
3	2	O	positif	FAUX
4	3	A	positif	FAUX
5	4	O	négatif	VRAI
6	5	O	positif	FAUX
7	6	A	négatif	FAUX
8	7	O	négatif	VRAI
9	8	AB	positif	FAUX
10	9	A	positif	FAUX
11	10	B	positif	FAUX
12	Nombre de donneurs universels :			3

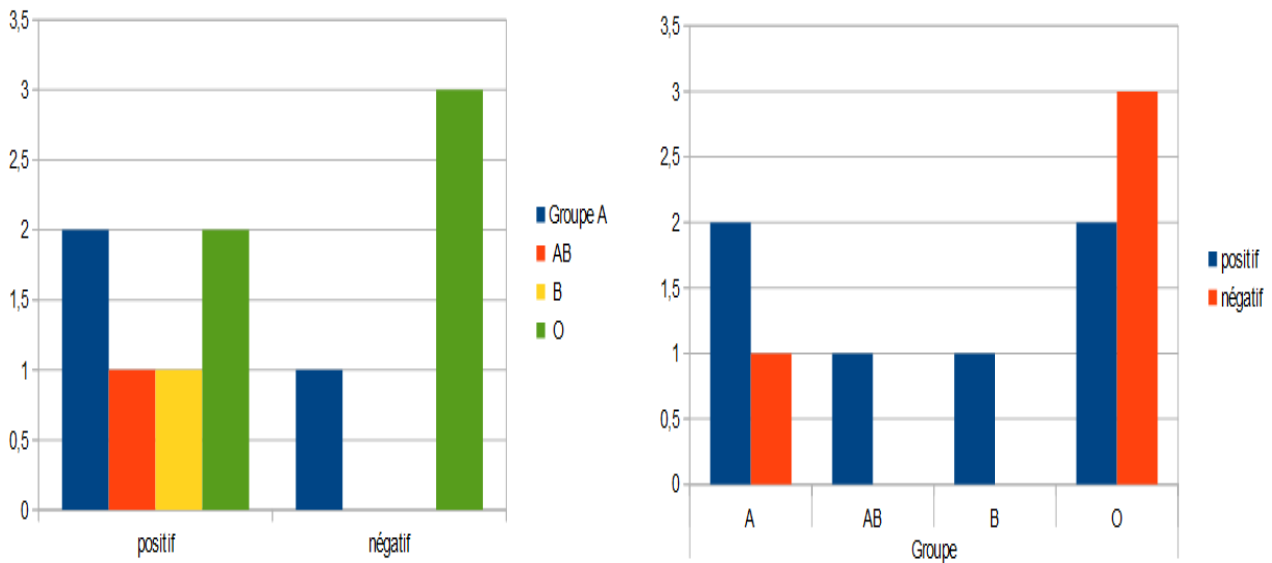
Remarque : Le tableur permet aussi de trier et/ou de filtrer les différentes valeurs d'un caractère. Pour cela, on sélectionne les colonnes correspondantes puis, dans le menu « Données », on choisit soit « Trier » soit « Filtre » puis « Autofiltre » (dans OpenOffice) ou « Filtrer » (avec Excel)

## II. Représentations graphiques

Pour interpréter les données d'une population, il peut être utile de faire des diagrammes. Il y a plusieurs possibilités :

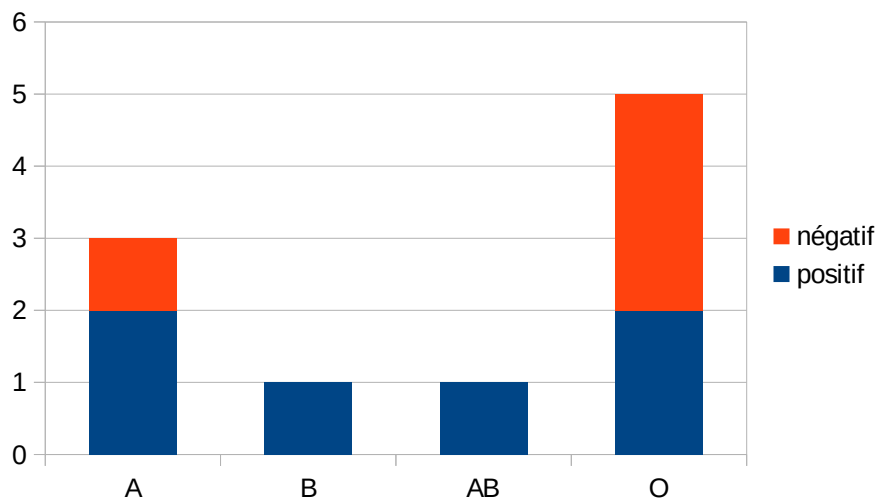
(a) **Le diagramme en barre (ou en bâtons)** : à chaque valeur on associe un segment vertical, ou un rectangle vertical de largeur constante et réduite, dont la longueur est proportionnelle à l'effectif.

Exemple : Si l'on reprend l'exemple des groupes sanguins, on a suivant les caractères étudiés :



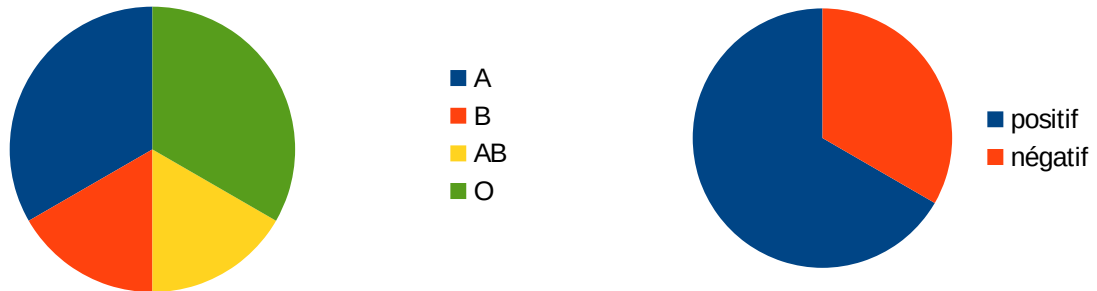
Remarques :

- Dans les deux exemples ci-dessus, on parle de diagramme **multi-barre** car il y a plusieurs barre pour un même caractère étudié d'où la nécessité d'utiliser des couleurs différentes.
- On peut également représenter les rectangles avec différentes couleurs suivant les caractères étudiés et les cumuler comme dans l'exemple ci-dessous



(b) **Le diagramme circulaire (ou « camembert »)** : c'est un disque partagé en secteurs angulaires dont la mesure de chaque angle est proportionnelle à l'effectif.

Exemples : Si l'on reprend l'exemple des groupes sanguins, on a suivant le caractère étudié :



Remarque : Pour construire ces diagrammes, il faut s'appuyer sur le tableau croisé d'effectifs et utiliser la proportionnalité en mettant en correspondance l'effectif total avec l'angle de 360°.

(c) **Le nuage de points** : il permet de représenter une population lorsque celle-ci est caractérisée par deux caractères quantitatifs (caractère caractérisé par des nombres). Chaque point du nuage a pour abscisses une valeur d'un caractère et pour ordonnées la valeur correspondante de l'autre caractère.

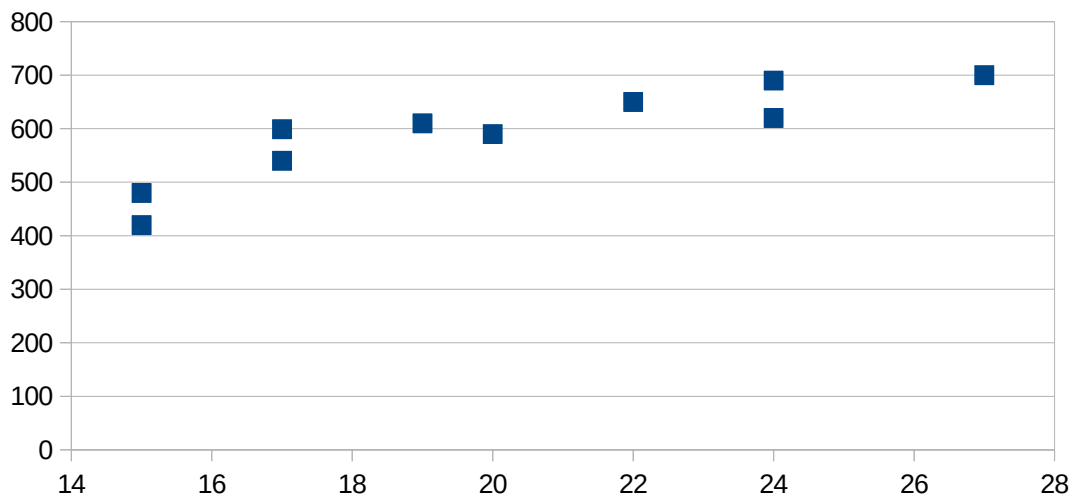
Remarque : On ne peut pas réaliser un nuage de points avec l'exemple des groupes sanguins, en effet, les groupes et le rhésus ne sont pas désignés par des nombres (« A », « B », « AB » et « O » pour les groupes, « positif » et « négatifs » pour les rhésus.)

Exemple : Une entreprise vend des ordinateurs, elle prélève dans son stock dix ordinateurs pour relever la taille de l'écran (en pouces) et le prix de vente (en euros). On obtient les résultats suivants.

Numéro	1	2	3	4	5	6	7	8	9	10
Taille	15	15	17	17	19	20	22	24	24	27
Prix	420	480	540	599	610	590	650	620	690	700

Ici les deux caractères étudiés (taille et prix) sont quantitatifs car représentés par des nombres. On peut donc construire un nuage de points en prenant pour abscisses la taille et pour ordonnées le prix.

Sur le nuage de points qui suit, chaque point représente un ordinateur.



Remarques :

- Avant de se lancer dans la construction d'un diagramme, il faut bien observer l'ordre de grandeur des valeurs pour anticiper le choix des échelles.
- Lorsque que l'on réalise un nuage de points, il n'y a pas besoin d'un tableau croisé d'effectifs.
- Sur le nuage de points ci-dessus, on peut constater que les points sont « presque alignés » ce qui peut révéler une situation de progression régulière.
- Pour représenter graphiquement des données, on peut utiliser un tableur. Il suffit de sélectionner les plages de données correspondantes puis, dans le menu « Insertion », on choisit « Diagramme » (avec OpenOffice) ou « Graphique » (avec Excel) puis le type de diagramme que l'on souhaite réalisé (des réglages plus fin sont souvent nécessaire).