

## CHAPITRE 5 – Statistiques à deux variables

### A) Tableaux croisés d'effectifs

#### 1) Rappels sur la fréquence

##### a) Rappel

Dans une série statistique, la fréquence associée à une valeur est le quotient de l'effectif lié à cette valeur sur l'effectif total.

##### b) Exemples :

i) Soit une entreprise employant 150 femmes et 50 hommes, la fréquence des femmes sera le quotient de 150 sur 200, soit trois quarts, 0,75 ou encore 75% (ces trois expressions se rapportent à trois écritures différentes du même nombre !).

De même la fréquence des hommes sera alors le quotient de 50 sur 200, soit un quart, ou 0,25, ou encore 25% (même remarque que ci-dessus).

ii) Soit la série statistique suivante, représentant les achats de médicaments dans une pharmacie : compléter la ligne fréquence et la colonne total !

Classe de montant des achats	]0 ; 2000[	[2000 ; 4000[	[4000 ; 6000[	[6000 ; 8000[	Total
Effectif	15	30	20	15	
Fréquence					

### 2) Fréquences dans un tableau croisé d'effectifs

#### a) Tableau croisé d'effectifs

Supposons que l'on veuille préciser ce tableau en tenant compte de l'heure d'achat des médicaments, par intervalles de 8h à 11h, de 11h à 15h et de 15h à 18h : on obtient alors un tableau croisé d'effectifs.

##### Exemple :

Classes de montant Classes d'heures	]0 ; 2000[	[2000 ; 4000[	[4000 ; 6000[	[6000 ; 8000[	Total
[7 ; 11[	10	15	10	10	(45)
[11 ; 15[	2	5	5	3	(15)
[15 ; 18[	3	10	5	2	(20)
Total	15	30	20	15	(80)

i) Compléter la dernière colonne.

ii) Refaire ce tableau en remplaçant les effectifs par des fréquences (on exprimera toutes les fréquences en pourcentages et on les arrondira à deux décimales).

Classes de montant Classes d'heures	]0 ; 2000]	[2000 ; 4000[	[4000 ; 6000[	[6000 ; 8000[	Total
[7 ; 11[	(12,5%)	(18,75%)	(12,5%)	(12,5%)	(56,25%)
[11 ; 15[	(2,5%)	(6,25%)	(6,25%)	(3,75%)	(18,75%)
[15 ; 18[	(3,75%)	(12,5%)	(6,25%)	(2,5%)	(25%)
Total	(18,75%)	(37,5%)	(25%)	(18,75%)	100%

**b) Fréquence conjointe :**

Les fréquences définies par une classe de montants et une classe d'heures (qui dépendent donc des deux variables choisies) s'appellent fréquences conjointes : par exemple, à la 3ème colonne, ligne 4, on trouve 12,5%, la fréquence conjointe de [2000 ; 4000[ et de [15 ; 18[.

C'est la fréquence de "l'achat avec montant entre 2000 et 4000F et effectué entre 15h et 18h".

En termes probabiliste d'événements, on dirait que c'est la fréquence de  $A \cap B$ , si l'on appelle A l'événement "montant entre 2000 et 4000F" et B l'événement "achat effectué entre 15h et 18h".

On notera  $f_{A \cap B}$  la fréquence conjointe de A et B.

**Exemples :**

- . Trouver la fréquence des achats effectués entre 15h et 18h d'un montant compris entre 6000 et 8000F.
- . Trouver la fréquence des achats effectués entre 7h et 11h d'un montant compris entre 0 et 2000F.

**c) Fréquence conditionnelle :**

C'est la fréquence des achats "sachant que ...", autrement dit on ne s'intéresse qu'aux achats ayant une certaine particularité, et on fait des statistiques dans ce cadre limité.

On notera  $f_B(A)$  la fréquence de A sachant B. Pour la calculer, on s'intéresse uniquement la ligne (ou la colonne) B et on calcule la fréquence de A dans cette ligne.

**Exemples :**

- . Trouver la fréquence des achats effectués de 7h à 11h **sachant que** leur montant est entre 0 et 2000F.
- . Trouver la fréquence des achats de montant entre 2000 et 4000F **sachant que** ils ont "été effectués entre 11h et 15h".
- . Trouver la fréquence des achats de montant entre 4000 et 6000F **sachant que** ils ont "été effectués entre 15h et 18h".
- . Trouver la fréquence des achats effectués de 11h à 15h **sachant que** leur montant est entre 6000 et 8000F.

**d) Propriété de la fréquence conditionnelle :**

Pour tous A et B, on a : 
$$f_B(A) = \frac{f_{A \cap B}}{f_B}$$

Si l'on appelle A l'événement "montant entre 2000 et 4000F" et B l'événement "achat effectué entre 15h et 18h", on aura :

$$f_B(A) = \frac{f_{A \cap B}}{f_B} = 12,5\% / 25\% = 50\% .$$

**e) La fréquence marginale :**

On appelle ainsi tout simplement la fréquence d'une colonne ou d'une ligne dans l'ensemble des colonnes

ou des lignes.

On les trouve dans la dernière ligne (ou colonne) du tableau. Ci-dessus,  $f_B$  (25%) est une fréquence marginale.

## **B) Corrélation entre deux variables**

### **a) Présentation du problème**

On remarque parfois que deux variables varient de façon similaire sur un échantillon statistique, ce qui peut amener à se poser la question d'une relation de cause à effet.

Ceci amène à deux interrogations à examiner : d'abord, voir si ces deux variables semblent vraiment être corrélées, et ensuite voir si cette corrélation peut être expliquée logiquement.

Par les statistiques, on peut tenter de répondre à la première question. Ceci se fait grâce à des tableaux de données judicieusement construits, tels des nuages de points qui peuvent ou non être significatifs d'une corrélation.

On peut préciser la corrélation en trouvant une droite ou une courbe qui puisse matérialiser cette corrélation.

### **Exemples sur des tableaux de données**

#### **Exemple 1 :**

Prenons un tableau donnant la tension artérielle maximale des femmes selon leur âge :

Âge	36	42	48	54	60	66
Tension max	11,8	13,2	14	14,4	15,5	15,1

#### **Exemple 2 :**

La consommation d'essence au cent kilomètres d'une voiture en vitesse stabilisée sur le plat semble dépendre de sa vitesse :

Vitesse (km/h)	10	20	30	40	50	60	70	80	90
Conso (l/100)	16,5	11,5	9	7,5	6,8	6,6	7	7,5	9

### **b) Nuages de points**

On représente ensuite ces deux tableaux dans un repère adapté, comme un ensemble de points qui semblera plus ou moins organisé, une variable servant d'abscisse et l'autre d'ordonnée.

Les tableaux permettent de créer cette représentation facilement.

### **c) Point moyen**

Le point moyen dans ce nuage de points est le point dont les coordonnées sont la moyenne des coordonnées de tous les points (moyenne des abscisses pour l'abscisse, moyenne des ordonnées pour l'ordonnée).

### **d) Droite de régression**

Lorsque le nuage de points ressemble plus ou moins à une droite, on peut tracer la "droite de régression", droite qui se rapproche le plus de l'ensemble des points (cette droite passe par le point moyen).

Parfois, le nuage ressemble plutôt à une courbe dont peut alors chercher l'équation.

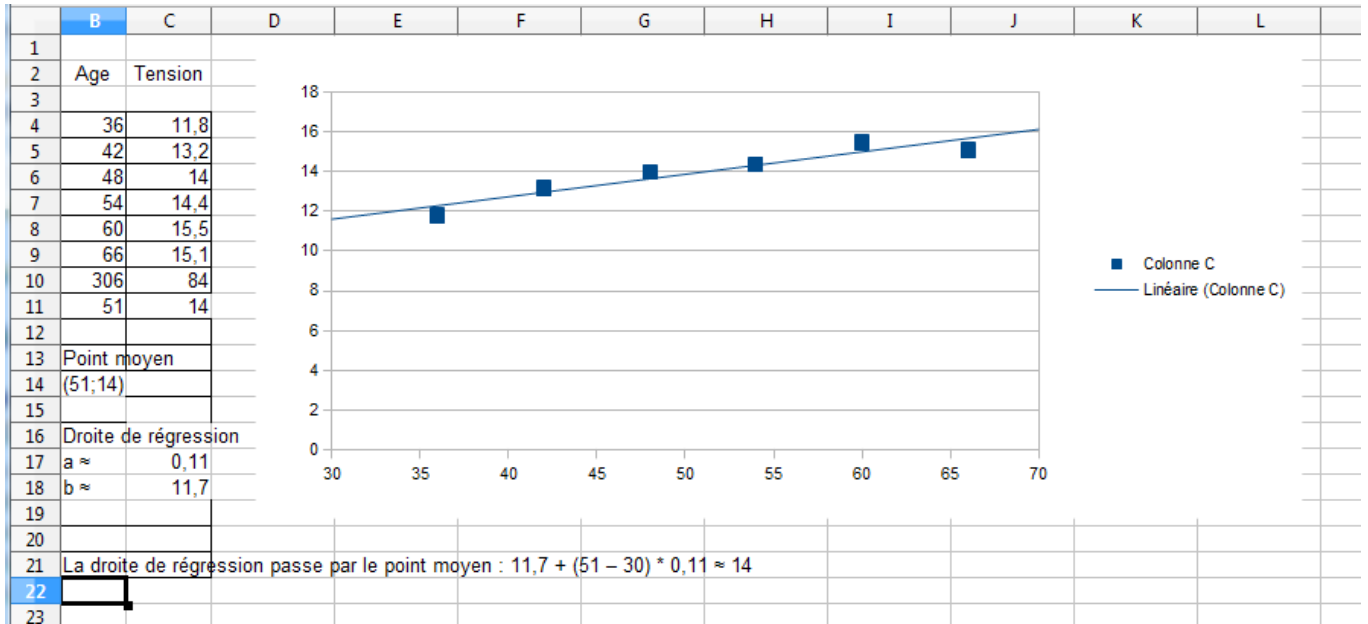
**e) Base théorique**

Attention, chercher des corrélations partout n'est pas forcément une bonne idée, deux données qui semblent corrélées ne sont pas forcément en relation de cause à effet, et même si elles le sont, cela ne dit pas laquelle est la cause et laquelle l'effet !

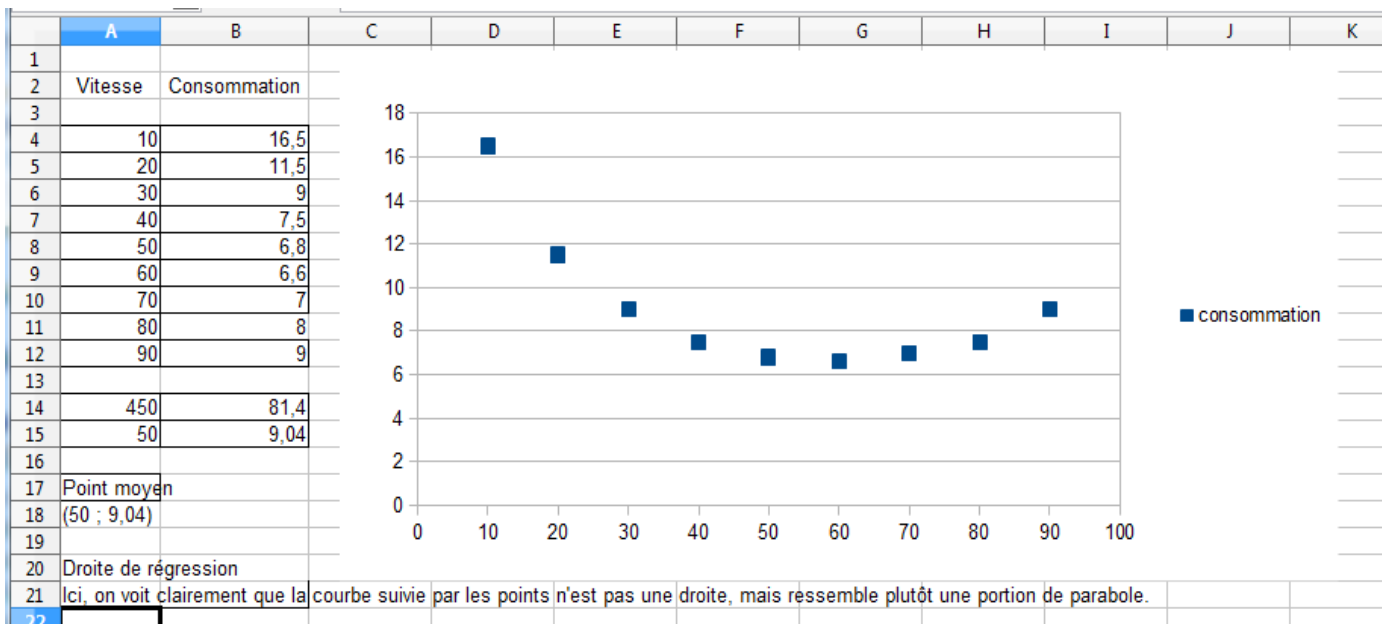
Il est donc nécessaire avant d'affirmer cette relation d'avoir derrière une explication théorique du phénomène.

**f) Exemples**

Application à l'exemple 1 :



Application à l'exemple 2 :



## Statistiques à deux variables – Fiche de révision

### Tableau croisé d'effectifs

C'est un tableau à deux dimensions (lignes et colonnes) dans lequel on note les fréquences d'apparition simultanée de deux valeurs (ou classes de valeurs) des deux variables étudiées.

### Fréquence conjointe

C'est la fréquence qui apparaît dans les cases centrales (hors totaux) d'un tableau croisé d'effectifs.

### Fréquence conditionnelle :

C'est la fréquence d'une valeur (ou classe de valeurs) d'une variable "sachant que" l'autre variable a une valeur (ou classe de valeurs) donnée.

### Propriété de la fréquence conditionnelle :

Pour toutes conditions A et B, on a :  $f_B(A) = \frac{f_{A \cap B}}{f_B}$ .

### Corrélation entre deux variables

On remarque parfois que deux variables varient de façon similaire sur un échantillon statistique, ce qui peut amener à se poser la question d'une relation de cause à effet.

### Nuages de points

On représente le tableau croisé d'effectifs dans un repère adapté, comme un ensemble de points qui semblera plus ou moins organisé, une variable servant d'abscisse et l'autre d'ordonnée.

### Point moyen

Le point moyen dans ce nuage de points est le point dont les coordonnées sont la moyenne des coordonnées de tous les points (moyenne des abscisses pour l'abscisse, moyenne des ordonnées pour l'ordonnée).

### Droite de régression

Lorsque le nuage de points ressemble plus ou moins à une droite, on peut tracer la "droite de régression", droite qui se rapproche le plus de l'ensemble des points (cette droite passe par le point moyen).

### Base théorique

Attention, chercher des corrélations partout n'est pas forcément une bonne idée, deux données qui semblent corrélées ne sont pas forcément en relation de cause à effet, et même si elles le sont, cela ne dit pas laquelle est la cause et laquelle l'effet !

Il est donc nécessaire avant d'affirmer cette relation d'avoir trouvé une explication théorique sérieuse du phénomène.