

CHAPITRE N° : SERIES STATISTIQUES À DEUX VARIABLES

I. Série statistique à deux variables

1°) Définition : Pour une population donnée, on peut être amené à étudier conjointement deux variables (ou caractères quantitatifs) notés généralement X et Y.

On relève pour chacun des individus la valeur x_i de X et y_i de Y. On associe ainsi à chaque individu un couple de nombres (x_i, y_i) .

L'ensemble de ces couples constitue une série statistique à deux variables X et Y.

2°) Exemple : On peut relever la taille en cm et le poids en kg de 5 garçons de 18 ans.

La série statistique à deux variables X (taille en cm) et Y (poids en kg) est constituée de 5 couples, par exemple : (174 ; 72) (178; 82) (175; 70) (169 ; 72) et (172 ; 68)

3°) Présentation des données : les résultats sont généralement présentés sous forme de tableaux.

Pour l'exemple ci-dessus :

Garçon	A	B	C	D	E
Taille (en cm) x_i	174	178	175	169	172
Poids (en kg) y_i	72	82	70	72	68

4°) Série chronologique :

Une **série** est dite **chronologique** lorsque la variable X est fonction du temps. On remplace souvent la valeur de l'année par son rang.

Exemple : Un chef d'entreprise a fait un relevé sur cinq années de l'évolution du pourcentage d'emplois à temps partiels dans son entreprise :

Année	2001	2002	2003	2004	2005
Rang x_i	1	2	3	4	5
Pourcentages d'emplois partiels y_i	6,5	12,5	16,9	20,6	23,5

Autre exemple : la série des chiffres d'affaires mensuels d'un hypermarché sur une année.

II. Nuage de points

1°) Définition : Le plan étant muni d'un repère orthogonal, chaque couple (x_i, y_i) d'une série statistique peut être considéré comme les coordonnées d'un point M_i .

L'ensemble des points M_i est appelé nuage de points associé à la série statistique.

2°) Représenter graphiquement le nuage de points d'une série (x_i, y_i) à partir d'un tableau de données et placer le point moyen :

a) Point moyen d'un nuage

Définition : On appelle point moyen du nuage de points associé à une série statistique à deux variables X et Y le point G de coordonnées (\bar{X}, \bar{Y}) où \bar{X} est la moyenne arithmétique des valeurs prises par X et où \bar{Y} est la moyenne arithmétique des valeurs prises par Y.

Remarque : Calculer \bar{X} et \bar{Y} revient à déterminer les moyennes arithmétiques des séries statistiques simples suivantes :

Valeur prise par X	x_1	x_2	...	x_n	$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n}$
Effectif	1	1	...	1	
Valeur prise par Y	y_1	y_2	...	y_n	$\bar{Y} = \frac{y_1 + y_2 + \dots + y_n}{n}$
Effectif	1	1	...	1	

b) *Exercice n°1* : L'évolution depuis 1998 du nombre d'adhérents à un club de tennis est représentée dans ce tableau :

Année x_i	1998	1999	2000	2001	2002	2003	2004	2005
Nombre d'adhérents y_i	129	134	118	145	142	104	150	136

Représenter le nuage de points correspondant à cette série. On pourra prendre 1 cm pour une année en abscisse en commençant par l'année 1998, puis 1 cm pour 5 personnes en ordonnée en commençant par 100.

Calculer les coordonnées du point moyen G de cette série et le placer sur le graphique.

Méthode :

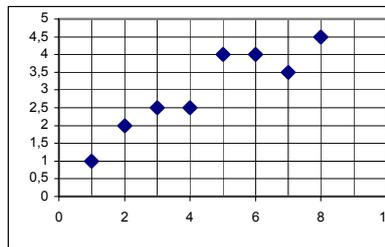
- Le plan étant rapporté à un repère orthogonal, on choisit une unité appropriée pour chaque axe :

CHAPITRE N° : SERIES STATISTIQUES À DEUX VARIABLES

- en repérant dans le tableau de données la plus petite et la plus grandes des valeurs prise par chacune des variables X et Y.
- en précisant si nécessaire le point d'intersection des axes (les graduations ne commençant pas obligatoirement à 0).
- On place sur le graphique tous les points de coordonnées (x_i, y_i) issus du tableau sans les relier.

3°) Dresser un tableau de données d'une série (x_i, y_i) à partir de la représentation graphique du nuage de points.

Exercice n°2 : dresser le tableau de données de la série statistique représentée par le nuage de points :



Méthode :

- Sur la figure, l'ensemble des points de coordonnées (x_i, y_i) correspondent aux valeurs prises respectivement par chacune des variables x et y.
- On dresse un tableau de données comprenant dans la première ligne (ou colonne) les valeurs prises par la variable x rangées dans l'ordre croissant et dans la deuxième ligne (ou colonne) les valeurs correspondantes prises par la variable y.

III. Interprétation du nuage de points

1°) Il n'existe pas de lien entre les deux variables.

Lorsque les points du nuage se répartissent de façon dispersée, il n'y a pas d'ajustement possible et il n'y a pas a priori de lien entre les deux variables de la série statistique : voir exercice n°1.

2°) Il existe un lien entre les deux variables :

Lorsqu'un ajustement s'impose au vu de la forme du nuage, on est amené à penser que l'un des caractères dépend de l'autre.

Pour des raisons de simplicité on cherche deux types de liaison :

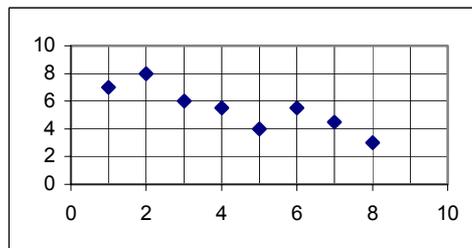
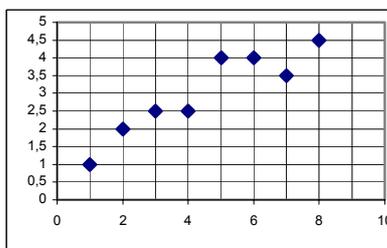
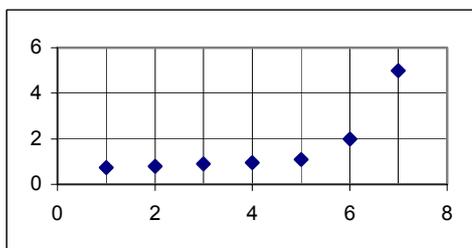
Liaison linéaire ou liaison exponentielle : voir exercice n°3.

IV. Ajustement affine

1°) Définition : lorsque le nuage a une allure globalement rectiligne, on peut effectuer un **ajustement affine**, c'est à dire trouver une droite qui passe au plus près des points de ce nuage.

2°) Faire un ajustement au jugé.

Exercice n°3 : Soit trois nuages de points 1,2 et 3.



a) Pour l'un des nuages, un ajustement affine ne semble pas approprié, lequel ? Pourquoi ?

b) Pour les deux autres droites, indiquer, parmi les équations données, l'équation de leur droite d'ajustement :

- $y = 0,5x + 3$;
- $y = 0,45x + 1$;
- $y = -0,6x + 8$;
- $y = -0,57x + 2$.

3°) Ajustement affine au jugé

On admettra que pour que l'ajustement soit le meilleur possible, il faut que la droite d'ajustement affine passe par le point moyen G de ce nuage.

Exercice n°4 (bac STT CG Pondichery 2006) : Le tableau suivant donne l'évolution du SMIC mensuel brut pour 169 heures de travail.

Les montants sont donnés en euros pour les années 2001, 2002, 2003 et 2004, en francs pour les autres années.

Année	1997	1998	1999	2000	2001	2002	2003	2004
Rang de l'année : x	0	1	2	3	4	5	6	7
Montant du SMIC en francs	6663,67	6797,18	6881,68	7101,38				
Montant du SMIC en euros : y	1127	1154	1215	1286

(Données INSEE)

Rappel : 1 euro = 6,55957 francs.

CHAPITRE N° : SERIES STATISTIQUES À DEUX VARIABLES

- a. A l'aide de la calculatrice, déterminer les montants du SMIC mensuel en euros arrondis à l'unité pour les années 1997, 1998, 1999 et 2000 et compléter le tableau.
- b. Sur la feuille de papier millimétré fournie, représenter dans un repère orthogonal (O, \vec{i}, \vec{j}) du plan, le nuage des points de coordonnées $(x; y)$. On prendra comme unités graphiques :
- ◆ 2 cm pour une année en abscisse en commençant au rang 0 ;
 - ◆ 1 cm pour 20 euros en ordonnées en commençant la graduation à 900.
- c. Déterminer les coordonnées du point moyen G de ce nuage et le placer sur le graphique précédent.
- d. On considère la droite D passant par G et de coefficient directeur 37,5 et réalisant un ajustement affine du nuage précédent.
- Déterminer l'équation réduite de la droite D.
 - Tracer cette droite sur le graphique.
 - Déterminer graphiquement une estimation du SMIC mensuel pour l'année 2006. On fera apparaître sur le graphique les tracés nécessaires à la lecture.
 - Retrouver, par le calcul, le résultat précédent.

4°) Remarques :

- a) Estimation : En supposant que la tendance observée se maintienne, la droite d'ajustement permet de faire des estimations. Elles s'effectuent soit par lecture graphique, soit en utilisant l'équation de la droite d'ajustement.
- b) La droite que l'on a tracé au jugé n'est pas unique. On aurait pu tracer de nombreuses droites approchant le nuage et passant par le point moyen.
 Cette méthode manque de précision.
 On a cherché des méthodes plus "calculatoires" pour avoir une plus grande précision.

5°) Ajustement affine par la méthode de Mayer :

On découpe le nuage en deux sous-nuages de quantités identiques (à une unité près pour le cas où le nombre de points est impair), puis on calcule les deux points moyens G_1 et G_2 des deux sous-nuages : la droite (G_1G_2) est un ajustement affine de la série.

Remarque : la droite (G_1G_2) passe par le point moyen du nuage.

Exercice n°5 : appliquer la méthode de Mayer.

Voici une série statistique double :

x_i	1	2	3	4	5	6	7	8
y_i	110	90	85	77	60	54	55	43

- a) Déterminer les coordonnées du point moyen G de cette série.
- b) G_1 désigne le point moyen des quatre premiers termes de la série et G_2 le point moyen des quatre derniers; Calculer les coordonnées de G_1 et G_2 .
- c) Donner une équation de la droite passant par G_1 et G_2 .
- d) Vérifier que la droite (G_1G_2) passe par G.

6°) Ajustement affine par la méthode des moindres carrés :

L'ajustement affine effectué par la calculatrice est obtenu par une méthode appelée méthode des moindres carrés. Ma calculatrice donne le coefficient directeur a et l'ordonnée à l'origine b de la droite.

Cette droite est appelée droite de régression de y en x.

Utilisation de la calculatrice : voir photocopie.

Exercice n°6 : Le tableau ci-dessous donne la population d'un pays au cours de six années consécutives.

Rang de l'année x_i	0	1	2	3	4	5
Population (en milliers) y_i	60 150	60 481	60 853	61 235	61 614	61 984

- a) Représenter le nuage de points de la série statistique (x_i, y_i) .
 Un ajustement affine apparaît-il envisageable ?
- b) Déterminer une équation de la droite d'ajustement de y en x par la méthode des moindres carrés.
- c) En déduire un effectif prévisionnel de la population pour l'année de rang 7, puis pour l'année de rang 9.

Ce que je dois savoir faire :

Etablir un tableau de données à partir du graphique d'un nuage de points.
 Représenter une série statistique à deux variables par un nuage de points.
 Calculer les coordonnées du point moyen d'un nuage de points
 Réaliser graphiquement un ajustement affine
 Réaliser graphiquement ou par le calcul des estimations à l'aide d'une droite d'ajustement.