

Chap 4 : Statistiques à deux variables

I. Série statistique

1) Définition

Définition 1 : On appelle **série statistique à deux variables** x et y toute liste de couples $(x_1; y_1), (x_2; y_2), \dots, (x_n; y_n)$.
 n est le nombre de ces couples.

Exemple : Par exemple

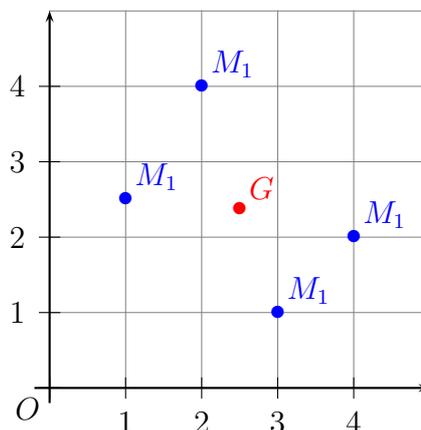
x_i	1	2	3	4
y_i	2,5	4	1	2

2) Nuages de points

Définition 2 : Etant donné une série statistique on appelle **nuage de points** associé l'ensemble des n points M_1, M_2, \dots, M_n du plan de coordonnées $(x_1; y_1), (x_2; y_2), \dots, (x_n; y_n)$.

Définition 3 : On appelle **point moyen** d'un nuage de points le point G de coordonnées $(\bar{x}; \bar{y})$ où \bar{x} est la moyenne de x_1, x_2, \dots, x_n et \bar{y} est la moyenne de y_1, y_2, \dots, y_n .

Exemple : Pour notre exemple on a le graphique suivant :



II. Ajustement affine

Lorsque les points du nuage semblent alignés dans le nuage de points on peut s'intéresser à une droite qui passe très près de tous ces points. Elle donne une idée du « lien » éventuel entre x et y .

Exemple : Un chef d'entreprise a fait un relevé sur cinq années de l'évolution du pourcentage d'emplois à temps partiel dans son entreprise, voici la série correspondante :

année	2002	2003	2004	2005	2006
rang x_i	1	2	3	4	5
% d'emplois partiels y_i	6,5	12,5	16,9	20,6	23,5

Faire un graphique représentant le nuage de points ainsi que le point moyen.

1) Méthode graphique

On peut trouver graphiquement une droite qui passe très près de tous les points et l'utiliser comme lien entre x et y .

Exemple : Avec notre exemple précédent on peut proposer comme droite d'ajustement la droite (M_2M_5) .

2) Méthode des moindres carrés

On peut trouver mathématiquement **la** droite qui passe le plus près possible de tous les points.

Définition 4 : On appelle **droite d'ajustement affine** la droite qui passe le plus près possible de tous les points du nuage.

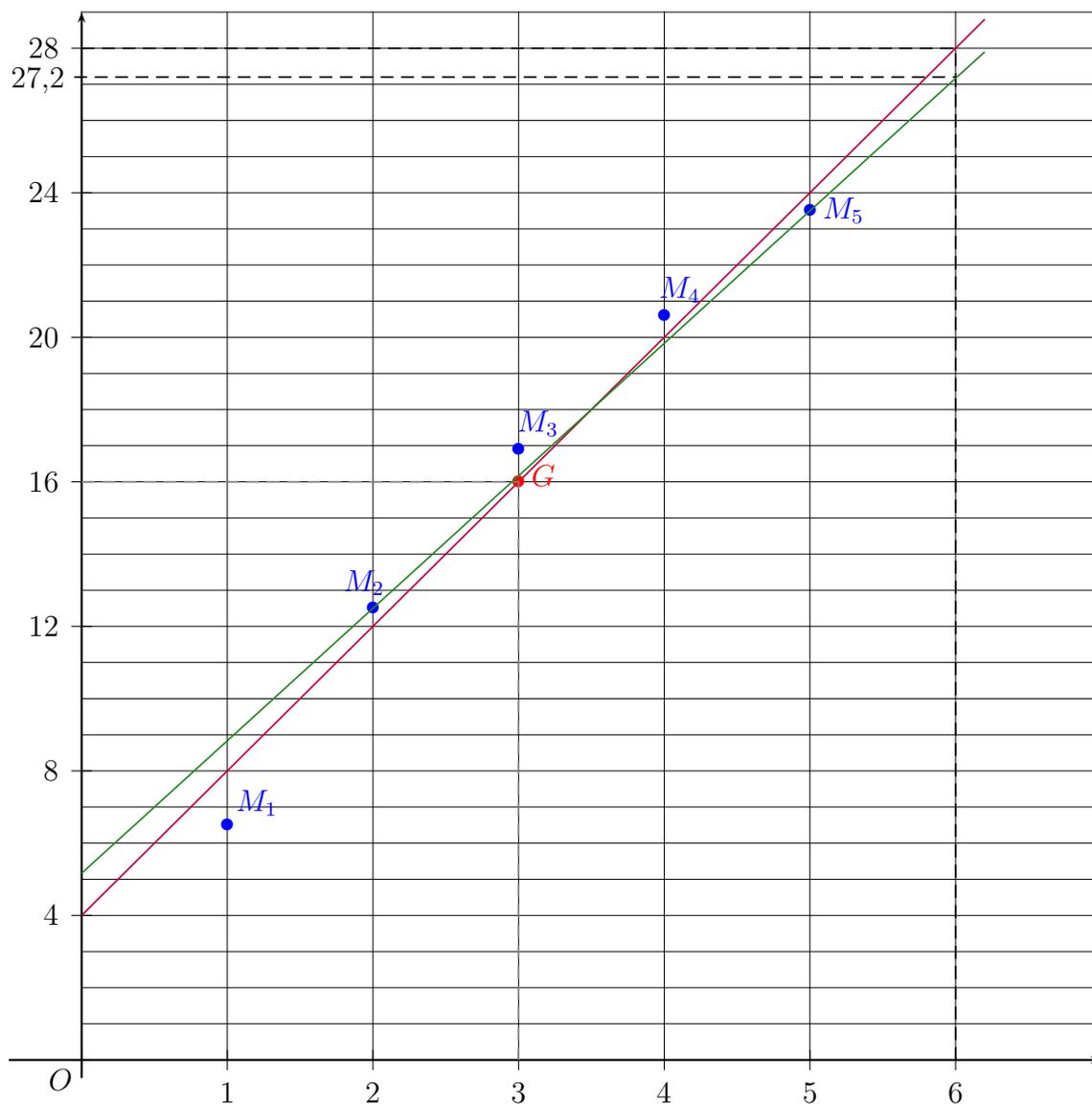
Son équation $y = ax + b$ se trouve à la calculatrice.

Remarque : Cette droite passe par le point moyen du nuage.

A quoi ça sert ?

Dans notre exemple les points semblent approximativement alignés on peut donc penser qu'il y a un lien entre le nombre d'emplois précaires et le rang de l'année.

Comment ? suivant à peu près l'équation de la droite d'ajustement affine, soit $y = 4x + 4$.



Si on suppose que le lien va perdurer on peut estimer le nombre d'emplois partiels pour 2007: $y \approx 27,22$ par lecture graphique sur la droite (M_2M_5) . On peut donc penser qu'il y aura environ 27,2% d'emplois partiels dans cette entreprise en 2007 par un ajustement affine graphique.

On peut penser qu'il y aura environ 28% ($4 \times 6 + 4 = 28$) d'emplois partiels dans cette entreprise en 2007 avec la méthode des moindres carrés.