

— 5 —

Séries statistiques

I. Nuage de points

Définition 1

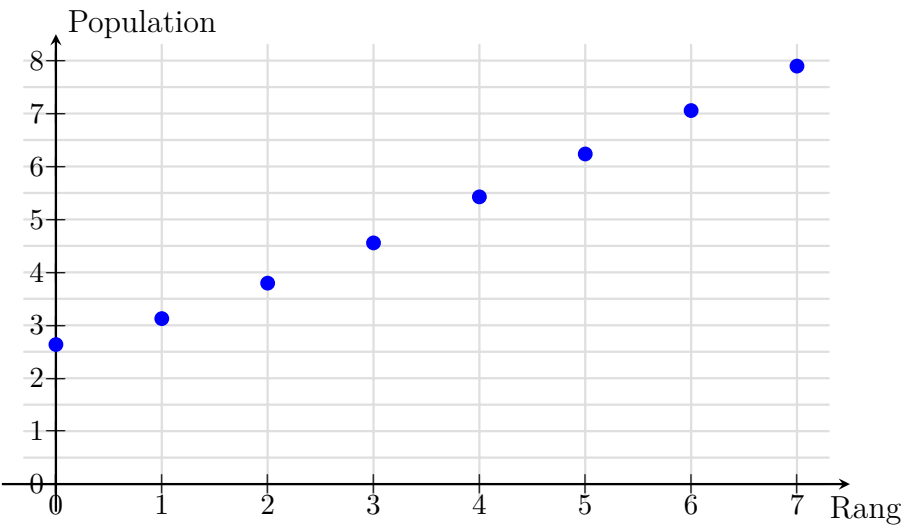
L'ensemble des points M_i de coordonnées (x_i, y_i) est appelé le **nuage de points** associé à la série statistiques à deux variables $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$.

Exemple :

On donne dans le tableau ci-dessous la population mondiale mesurée par année :

Année	1950	1960	1970	1980	1990	2000	2010	2020
Rang de l'année	0	1	2	3	4	5	6	7
Population mondiale (en milliards)	2,54	3,03	3,70	4,46	5,33	6,14	6,96	7,80

Le nuage de point suivant représente la population mondiale en fonction du rang de l'année :



II. Ajustement affine

1. Définition

L'objectif de cette partie est d'utiliser les données statistiques fournies afin d'estimer/approximer les valeurs inconnues.

Lorsque la valeur recherchée fait partie du domaine d'étude, on parle d'**interpolation**. Dans la situation précédente, il s'agirait d'estimer la population en 1965 par exemple.

Lorsque la valeur recherchée est hors du domaine d'étude, alors on parle d'**extrapolation**. Dans la situation précédente, il s'agirait de prévoir la population en 2030 par exemple.

Pour cette approximation, ce qui est généralement fait est d'observer le nuage de points obtenu afin de déterminer une courbe qui approche au mieux les points. Dans ce cours, on s'intéresse principalement au cas où l'on peut utiliser une droite comme approximation.

Définition 2

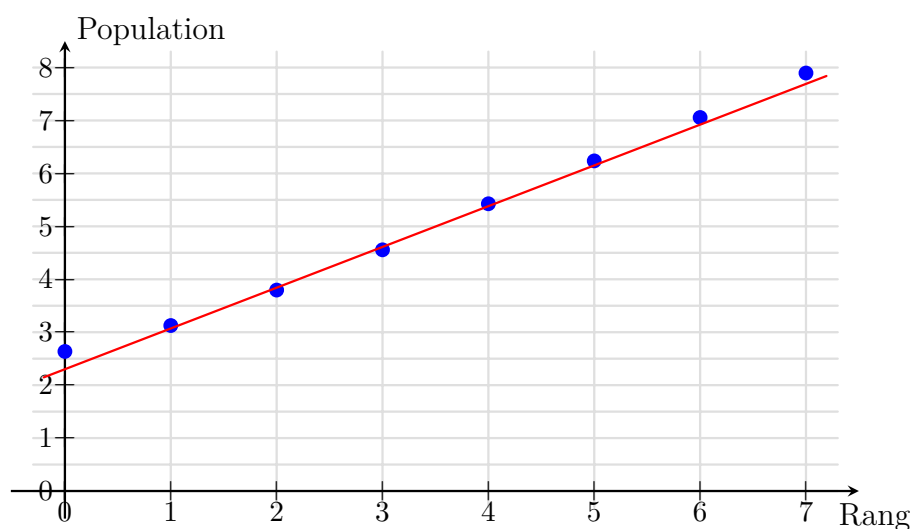
Lorsque les points sont sensiblement alignés, on peut construire une droite, appelée droite d'ajustement, passant « au plus près » de ces points.

2. Méthode « au jugé »

Sans technique particulière, on trace une droite qui semble passer près de chacun des points. Cette méthode est évidemment peu précise mais est très abordable.

Exemple :

On reprend l'exemple précédent et on trace la droite d'ajustement :



Celle-ci permet, par exemple, d'estimer qu'en 1965, la population mondiale était proche de 3,5 milliards d'habitants. (En réalité, celle-ci était d'environ 3,34 milliards)

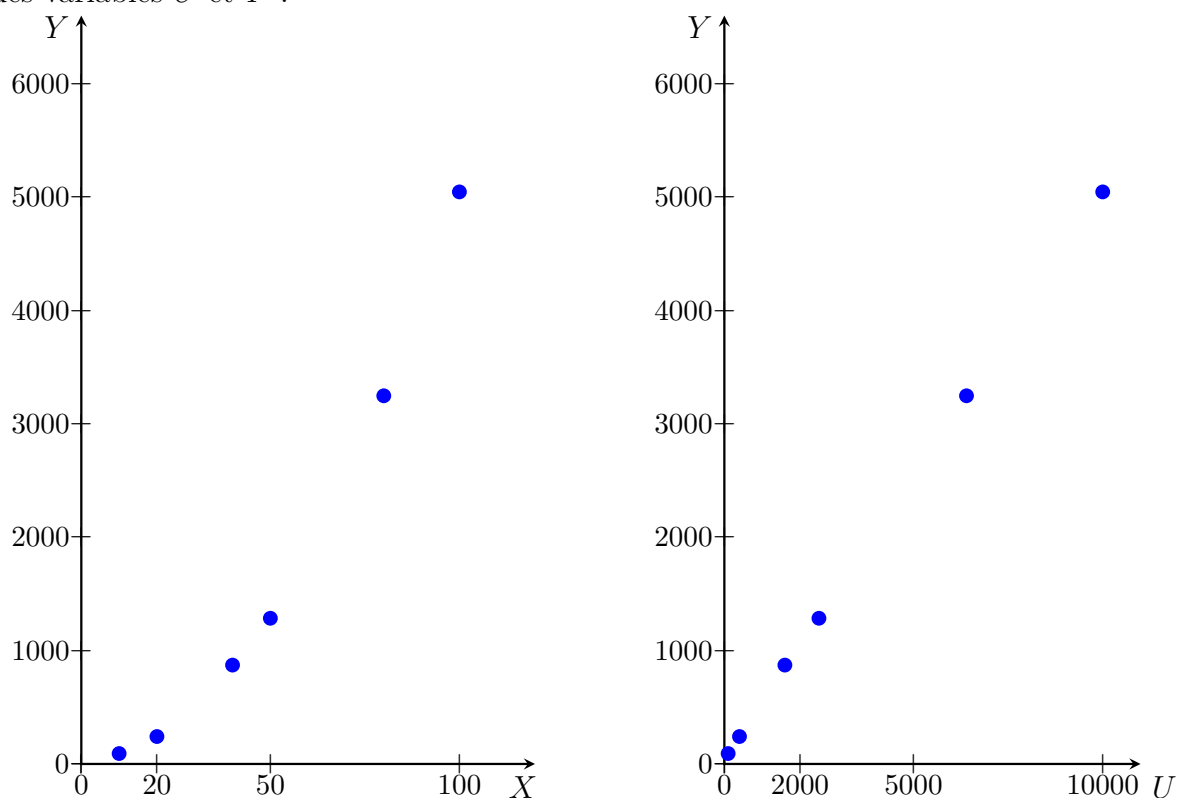
III. Changement de variable

Lorsqu'un nuage de point ne semble pas approximable par une droite, on peut tout de même réaliser un ajustement affine à l'aide d'un changement de variable.

On s'intéresse aux données du tableau suivant :

x_i	10	20	40	50	80	100
$u_i = x_i^2$	100	400	1600	2500	6400	10000
y_i	47	197	827	1240	3202	5000

On représente ci-dessous le nuages de points des variables X et Y , puis le nuage de points des variables U et Y :



On remarque que le changement de variable $U = X^2$ a permis d'obtenir un nuage de points sensiblement alignés (ce qui n'était pas le cas pour le premier graphique). On peut alors donner une droite d'ajustement.

On admet que cette droite admet pour équation : $Y = 0,5U + 2$.

Puisque $U = X^2$, alors on en déduit que $Y = 0,5X^2 + 2$.