

Statistique descriptive à deux variables

Lorsqu'on étudie une population, il est fréquent que l'on recueille plusieurs types d'informations en même temps. Pour chaque individu, on dispose donc de plusieurs valeurs de caractères et on souhaite analyser les données qui nous sont fournies de cette manière. On peut par exemple chercher à savoir si l'un des caractères est lié à un autre et dans ce cas de quelle manière ou si les deux caractères sont totalement indépendants. Pour faire cela, il faut commencer par représenter les données qui nous sont fournies.

1 Méthodes de représentation

1.1 Tableau de données

Déterminer une distribution statistique à deux dimensions relative au couple $(X; Y)$, c'est connaître :

- les valeurs possibles pour le caractère X , ces valeurs seront notées x_1, \dots, x_n .
- les valeurs possibles pour le caractère Y , ces valeurs seront notées y_1, \dots, y_n .
- les valeurs correspondant à chaque observation ($X = x_i$ et $Y = y_j$). Cet effectif sera noté $n_{i,j}$. La fréquence correspondante est $f_{i,j} = \frac{n_{i,j}}{n}$. Il est commode de présenter ces renseignements à l'aide d'un tableau à double entrée.

Exemple 1.

age/poids	[50,80[[80,110[[110,140[
[20,30[10	5	1
[30,40[9	8	3
[40,50[5	7	2
[50,60[15	12	3

1.2 Nuage de points

Le plan étant muni d'un repère orthogonal, on peut associer à chaque point (x_i, y_i) le point M_i de coordonnées x_i et y_i . L'ensemble des points obtenus constitue le nuage de points représentant la série statistique.

1.3 Point moyen

On appelle point moyen d'un nuage de n points le point M de coordonnées

$$x_M = \bar{x} \text{ et } y_M = \bar{y}.$$