

STATISTIQUES

1. Statistiques à deux variables

1. Série statistique double

Une série statistique double est l'étude simultanée de deux caractères des individus d'une population

Exemple : Le revenu mensuel x_i et le nombre de jours de vacances aux sports d'hivers y_i durant les 3 dernières années d'un groupe de personnes.

x_i (euros)	954	1247	1295	1578	1982	2045	2489	2784	3578	4225
y_i (jours)	0	7	3	5	35	21	19	0	79	21

2. Nuage de points

Chaque individu d'une série statistique double peut être représenté dans un repère par le point de coordonnées $(x_i; y_i)$. L'ensemble des points représentant les individus de la série est appelé nuage.

3. Le point de coordonnées $(\bar{x}; \bar{y})$ est appelé point moyen de la série.

2. Ajustement linéaire par la méthode des moindres carrés.

1. Droite de régression de y en x

Il s'agit de la droite d telle que la somme des carrés des écarts entre les points du nuage et leurs projections parallèlement à (Oy) sur d soit minimale.

Cette droite passe par le point moyen de la série.

Cette droite a pour coefficient directeur le nombre a défini par :

$$a = \frac{1}{V(x)} \times \frac{1}{n} \times \sum (x_i - \bar{x})(y_i - \bar{y})$$

Son équation est donc $y = a(x_i - \bar{x}) + \bar{y}$

2. Somme des carrés des résidus

On appelle résidu la distance entre un point du nuage et sa projection verticale sur la droite de régression. On note S la somme des carrés des résidus. C'est cette valeur que minimise la droite de régression de y en x.

3. Compléments

Il arrive qu'une série double présente une autre allure que celle d'une droite. On peut alors essayer de faire un ajustement linéaire sur une série déduite. Par exemple, si le nuage de points évoque une fonction exponentielle, on pourra envisager de faire un ajustement linéaire sur la série $(x_i; \ln(y_i))$.