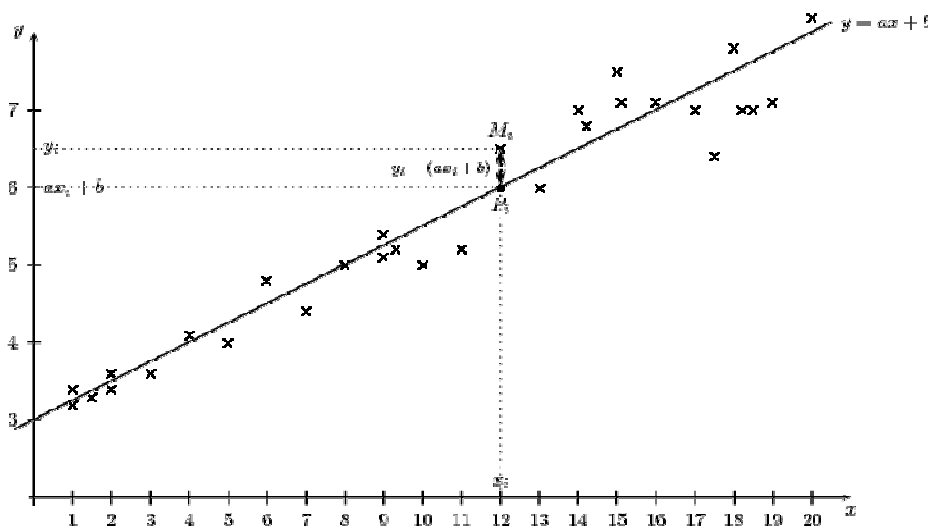


1 – PREAMBULE

Dans bien des situations il est courant d'avoir à sa disposition deux ensembles de données de taille n , $\{y_1, y_2, \dots, y_n\}$ et $\{x_1, x_2, \dots, x_n\}$, obtenus expérimentalement ou mesurés sur une population.

On peut alors se proposer de rechercher un modèle mathématique représentant ces ensembles, c'est-à-dire rechercher une relation de la forme $y = f(x)$ pouvant éventuellement exister entre les x_i et les y_i .

Lorsque la relation recherchée semble être affine, c'est-à-dire de la forme $y = a \cdot x + b$, on parle de régression linéaire.



2 – CRITERE POUR DEFINIR LA DROITE

Considérons une droite « quelconque » d'équation $y = a \cdot x + b$. Prenons maintenant un point $M_i(x_i, y_i)$ du nuage de point. Dans le cas général, le point M_i n'appartient pas à la droite, c'est-à-dire que ses coordonnées ne vérifient pas l'équation qu'on s'est donné : $y_i \neq a \cdot x_i + b$.

Il y a donc un écart entre les ordonnées des points M_i (du nuage) et P_i (de la droite) qui ont la même abscisse x_i . Et ceci est généralement vrai pour tous les points du nuage !

Soit e_i l'écart d'ordonnée entre les points M_i et P_i : on a : $e_i = y_{P_i} - y_{M_i}$.



On pose alors comme critère la droite qui, parmi toutes, minimise la somme des carrés des écarts.

3 – COEFFICIENTS a ET b DE LA DROITE

Partant du critère énoncé, on montre que les paramètres de l'équation de la droite se calculent ainsi :

$$a = \frac{N \sum_{n=1}^N (n \times D_n) - \sum_{n=1}^N n \cdot \sum_{n=1}^N D_n}{N \sum_{n=1}^N n^2 - (\sum_{n=1}^N n)^2}$$

$$b = \frac{\sum_{n=1}^N D_n}{N} - a \cdot \frac{\sum_{n=1}^N n}{N}$$