$\sqrt{f(x)}$

MATHEMATIQUES APPLIQUEES

Coefficients de corrélation et de détermination

13

1 – PREAMBULE

Etudier la corrélation entre deux ou plusieurs variables, c'est étudier **l'intensité de la liaison qui peut exister entre ces variables**.

Une mesure de cette corrélation est obtenue par le calcul du *coefficient de corrélation linéaire* (dit de Bravais-Pearson).

2 – FORMULE

On considère deux séries de même longueur $X\{x_1, x_2, \ldots\ldots, x_n,\}$ et $Y\{y_1, y_2, \ldots\ldots, y_n,\}$. Pour connaître le coefficient de corrélation liant ces deux séries, on applique la formule suivante :

$$\boxed{r = \frac{\sum_{i=1}^{N}(x_i - \overline{x})\cdot(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2 \cdot \sum_{i=1}^{N}(y_i - \overline{y})^2}}}$$

Avec :

$$\begin{cases} \frac{1}{N}\cdot\sum_{i=1}^{N}(x_i - \overline{x})\cdot(y_i - \overline{y}) & \text{Covariance entre } X \text{ et } Y \\ \sqrt{\frac{1}{N}\cdot\sum_{i=1}^{N}(x_i - \overline{x})^2} & \text{Ecart-type de } X \\ \sqrt{\frac{1}{N}\cdot\sum_{i=1}^{N}(y_i - \overline{y})^2} & \text{Ecart-type de } Y \end{cases}$$

3 – INTERPRETATION

Le coefficient de corrélation r est compris entre -1 et $+1$. Plus le coefficient est proche des valeurs extrêmes -1 et $+1$, plus la corrélation entre les variables est forte ; on emploie simplement l'expression « fortement corrélées » pour qualifier les deux variables. Une corrélation égale à 0 signifie que les variables sont linéairement indépendantes.

Donc :

$r = +1$: corrélation positive parfaite

$r = -1$: corrélation négative parfaite

$r = 0$: absence totale de corrélation

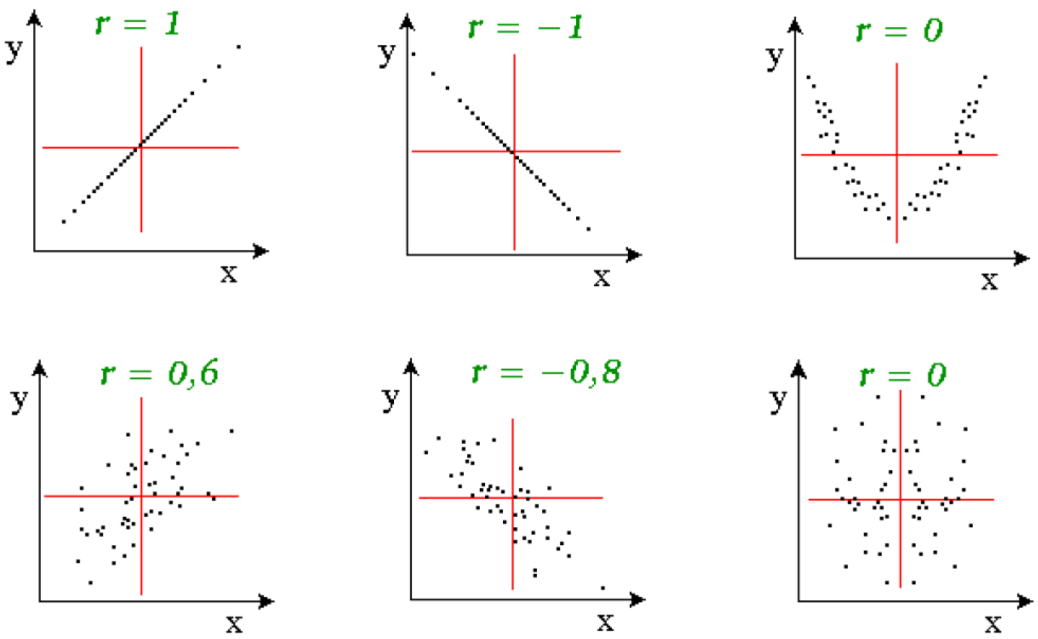
Le coefficient de corrélation nous donne des informations sur **l'existence d'une relation linéaire** (sous forme d'une droite) entre deux grandeurs considérées.

Un coefficient de corrélation nul **ne signifie pas l'absence de toute relation** entre les deux grandeurs. Il peut exister une relation non linéaire entre elles.

Le coefficient de corrélation n'est pas **sensible aux unités de chacune des variables**. Ainsi, par exemple, le coefficient de corrélation linéaire entre l'âge et le poids d'un individu sera identique que l'âge soit mesuré en semaine, en mois ou en année(s).

En revanche, ce coefficient de corrélation est **extrêmement sensible à la présence de valeurs aberrantes** ou extrêmes dans notre ensemble de données (valeurs très éloignées de la majorité des autres, pouvant être considérées comme des exceptions).

4 – REPRESENTATION GRAPHIQUE DU COEFFICIENT DE CORRELATION



5 – COEFFICIENT DE DETERMINATION

Quelle que soit la technique utilisée pour construire un droite de tendance (via les moindres carrés ou autre), on s'assure de la viabilité du modèle retenu en utilisant un critère comme le coefficient de détermination « r^2 ».

$$0 \leq R^2 \leq 1$$

$R^2 \to 1 \quad \Rightarrow$ bon modèle

$R^2 \to 0 \quad \Rightarrow$ mauvais modèle.

Le coefficient de détermination est donc égal au carré du coefficient de corrélation.