Chapitre 2 : Les Méthodes Statistiques Relatives à la Corrélation

1. Principes fondamentaux

Les conditions d'application de ces méthodes :

* Les deux (2) variables sont indépendantes et ont une distribution normale à 2 démentions.

* Les échantillons sont aléatoires simple et indépendant les uns des autres dans les comparaisons de 2 ou plusieurs coefficient de corrélation.

2. Objectifs de la corrélation linéaire (Positionnement):

* Variations respectives de plusieurs grandeurs dans une même population. *exp: relation entre poids et taille*

* Courbe associée à la fonction ***y = f (x)***.

* Si la loi est définie, la connaissance de « *x* » suffit à déterminer « *y* » : *Relation fonctionnelle (sciences exactes)*.

* Les fluctuations statistiques à une valeur d'une des variables correspondent une distribution de valeur de l'autre variable.

3. Représentation graphique :

Si on représente les couples de valeurs (*x*, *y*) sur le même repère orthonormé on obtient un nuage de point **(Fig. 1)**.

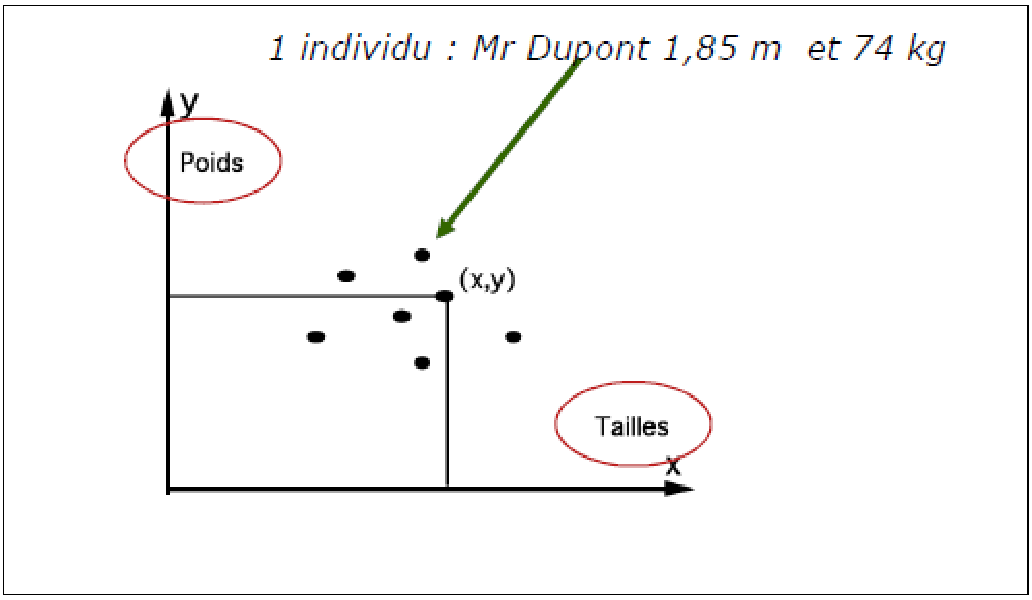


Figure 1. Le nuage de point

4. Corrélation et régression :

La régression permet d'étudier l'association entre deux variables quantitatives, en étudiant les variations de l'une en fonction des valeurs de l'autre.

Le coefficient de corrélation « *r* » est une mesure d'association entre deux variables quantitatives faisant jouer des rôles symétriques aux valeurs. Les deux variables peuvent être placées indifféremment en abscisse ou en ordonnées. On cherche à savoir simplement s'il existe une liste entre ces deux variables et à quantifier l'intensité de la liaison.

$$\boxed{\mathbf{r}_{xy} = \mathbf{cov}_{xy} \;/\; \mathbf{s}_x\, \mathbf{s}_y}$$

$\mathbf{cov}_{xy}$ **:** *Covariance* : $\mathbf{cov}_{xy} = (1/n \sum x_i\, y_i) - \bar{X}\,\bar{y}$

$\mathbf{s}_x$ **:** Ecart type de la variable « $\boldsymbol{x}$ »

$\mathbf{s}_y$**:** Ecart type de la variable « $\boldsymbol{y}$ »

Quels que soit l'unité et les ordres de grandeur de « *x* » et « *y* », le coefficient de corrélation est un nombre sans unité, compris entre (−1) et (+1). Il traduit la plus ou moins grande dépendance linéaire de « *x* » et « *y* » ou, géométriquement, le plus ou moins grand aplatissement du nuage de points.

4.1. La corrélation :

« *x* » et « *y* » sont des variables quantitatives. Pour dire que « *x* » et « *y* » sont corrélées, c'est affirmer qu'il existe une liaison entre ces deux variables.

Plus « *x* » varie dans un sens, plus « *y* » varie. Si « *y* » varie dans le ***même sens*** que « *x* »: ***la corrélation est positive***.

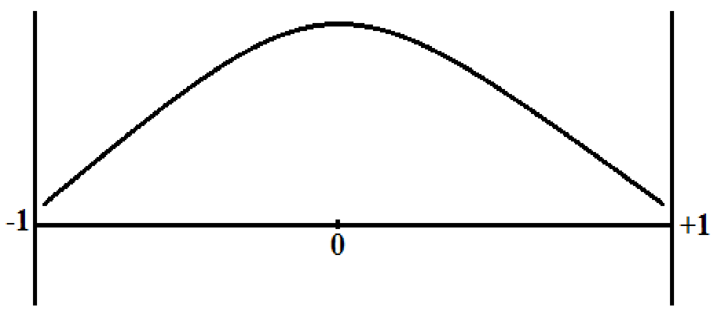
Si « *y* » varie dans le ***sens opposé*** de « *x* » : ***la corrélation est négative***.

Si « *x* » et « *y* » varient ***indépendamment*** de l'un de l'autre : ***les variables ne sont pas corrélées***.

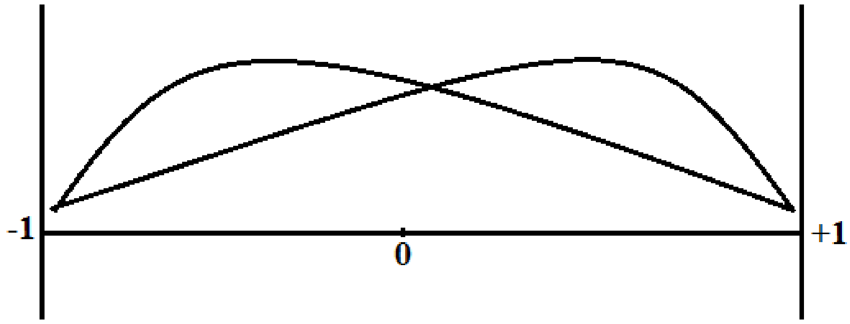
Quand le coefficient de corrélation est proche de **1** ou **−1**, les caractères sont dits "***fortement corrélés***".

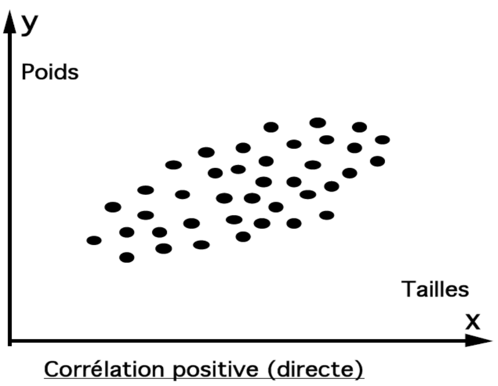
4.2. Distribution :

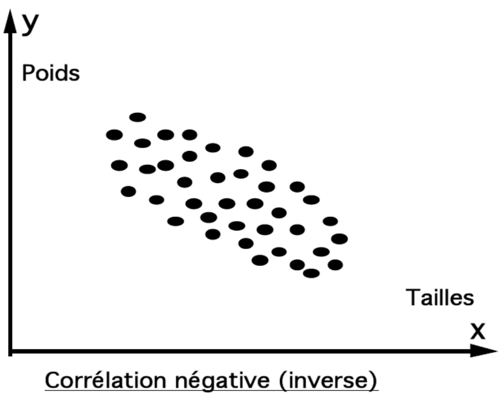
Lorsque la corrélation « r » est nulle (r= 0) la distribution d'échantillonnage est *symétrique* et elle est représentée comme suit :

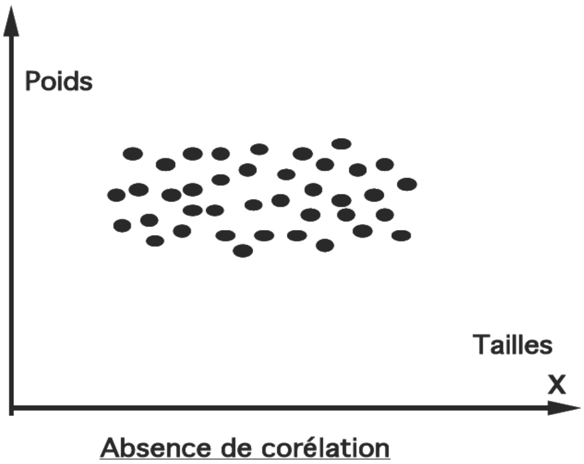


Lorsque le coefficient de corrélation « r » est différent du « 0 », la distribution est ***dissymétrique*** et elle est représentée comme suit :









Nuage de points diffus = Absence de liaison entre les caractères étudiés