ASSOCIATION (D2)
(24 / 03 / 2020, © Monfort, Dicostat2005, 2005-2020)

De façon générale, la notion d'**association** traduit l'existence d'une « liaison » entre des **événements** ou des **variables aléatoires** : cette liaison est souvent représentée par une **caractéristique légale** traduisant une notion de **dépendance** (cf **dépendance stochastique**).

(i) Ainsi, une **dépendance probabiliste** entre deux événements peut s'exprimer par une mise en défaut du **théorème des probabilités composées**. Deux évènements A et B seront associés ssi $P(A)\,P(B) \neq P(A \cap B)$. Par suite, plus la différence entre les deux membres sera grande, plus les événements A et B seront liés, ce qui peut se représenter de diverses façons.

De même, une **dépendance probabiliste**, ou parfois **dépendance statistique**, entre des va peut se représenter (de façon extensive) par (cf aussi **dépendance stochastique**) :

(a) une **corrélation**, par une **relation fonctionnelle** ;

(b) une (fonction de) **régression** ;

(c) ou encore une (fonction d') **interdépendance**.

(ii) Association et indépendance peuvent être considérés comme des ***antonymes statistiques***. Cependant, si l'indépendance peut se définir directement (définition « positive ») et de façon « unique », l'association peut revêtir des formes et des représentations variées. De plus, l'existence d'une association ne préjuge en rien de l'existence simultanée d'une **causalité** entre événements (resp variables).

(iii) On distingue souvent, en pratique, entre **« corrélation »** décrivant une liaison entre variables numériques, et **« association »** décrivant une liaison entre **variables qualitatives** (ou **attributs**) (cf **loi qualitative**).

Par ailleurs, il existe aussi des **liaisons mixtes** entre variables quantitatives et variables qualitatives.

On définit donc, selon le cas, des **coefficients de corrélation** (liant des variables numériques), des **coefficients d'association** (liant des variables qualitatives) ou des **coefficients mixtes** (cf eg **coefficient de corrélation ponctuel**).

De même, selon que les variables endogènes et exogènes sont de l'un ou l'autre type, on définit des **modèles quantitatifs** (eg modèles de régression ou d'interdépendance), des **modèles qualitatifs** (eg modèles à variables dépendantes qualitatives) ou des **modèles mixtes** (eg modèles d'analyse de la covariance ou de la variance).

(iv) Enfin, on peut, sous certaines conditions, adapter la formule définissant un coefficient de corrélation pour définir un coefficient d'association : mais cette transposition n'est pas toujours directe et dépend du **codage** choisi pour les variables qualitatives considérées.

Soit $(\Omega, \mathcal{T}, P)$ un **espace probabilisé**, $(\mathcal{X}, \mathcal{B})$ un espace numérique, $(\mathcal{K}, \mathcal{E})$ un espace non numérique et $(\xi, \kappa) : \Omega \mapsto \mathcal{X} \times \mathcal{K}$ un **couple aléatoire** composé d'une **variable quantitative** ξ et d'une **variable qualitative** κ. Le codage de κ en une variable $\zeta = c \circ \kappa$, où $c : \mathcal{K} \mapsto \mathcal{Z}$ est un **codage numérique** de κ et $(\mathcal{Z}, \mathcal{D})$ un espace numérique auxiliaire, conduit à calculer directement un **coefficient de corrélation linéaire** tq le suivant :

(1) $\rho_{\xi\zeta} = (V\,\xi)^{-1/2}\,(V\,\zeta)^{-1/2}\,C\,(\xi\,,\zeta) = (V\,\xi)^{-1/2}\,(V\,c \circ \kappa)^{-1/2}\,C\,(\xi\,, c \circ \kappa)$,

coefficient qui dépend ainsi, en général, de c.