

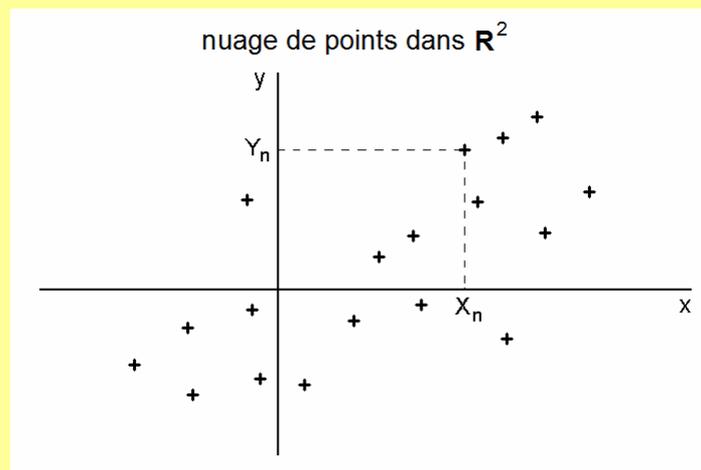
NUAGE DE POINTS (F, G11, H7, I9, J7, J9, K1, K15)

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

L'expression imagée **nuage de points**, à évocation géométrique et descriptive, intervient notamment en **Statistique** multivariée (cf **analyse multidimensionnelle**) : **analyse des données**, **modèle multi-indicé** (eg **modèle de régression**, **modèle d'interdépendance**). En effet, elle implique l'utilisation d'au moins deux **variables**.

(i) Soit (Ω, \mathcal{F}, P) un **espace probabilisé**, (ξ, η) un **couple aléatoire** réel (ie à valeurs dans \mathbf{R}^2) et (X, Y) un **N-échantillon** associé à (ξ, η) , avec $X = (X_1, \dots, X_N)$ et $Y = (Y_1, \dots, Y_N)$. Autrement dit, à tout élément (**unité statistique**) $\omega_n \in \Omega$ (**ensemble fondamental**) correspond un couple de valeurs (X_n, Y_n) , avec $X_n = \xi(\omega_n)$ et $Y_n = \eta(\omega_n)$, $\forall n \in \mathbf{N}_N^*$.

On appelle **nuage de points** la **représentation graphique** (dans \mathbf{R}^2) des N coordonnées (X_n, Y_n) (cf graphique ci-après).



(ii) Les observations (X_n, Y_n, Z_n) , $\forall n \in \mathbf{N}_N^*$, d'un triplet aléatoire (ξ, η, ζ) à valeurs dans \mathbf{R}^3 peuvent, de la même façon, être visualisées dans \mathbf{R}^3 .

(iii) Un (N,K) -**tableau statistique** T , à valeurs réelles et à deux dimensions, peut être considéré comme un couple constitué de deux « descripteurs » (eg N unités statistiques ou observations, d'une part, et K variables, d'autre part) et d'une **matrice** élément de $M_{NK}(\mathbf{R})$. Par suite :

(a) les lignes $(T_n)_{n=1, \dots, N}$ de T , peuvent être considérées comme des vecteurs de l'espace \mathbf{R}^K , formés des **observations** relatives à K variables $k \in \mathbf{N}_K^*$;

(b) de même, les colonnes $(t_k)_{k=1, \dots, K}$ de T , peuvent être considérées comme K vecteurs de l'espace \mathbf{R}^N , formés des K **variables** « observées » dans l'espace \mathbf{R}^N .

On peut alors parler du nuage des points T_n de \mathbf{R}^K et du nuage des points t_k de \mathbf{R}^N . Ces nuages appartiennent donc à des espaces de dimensions K ou N supérieures à 1.

(iii) Lorsque le tableau T est à $H > 2$ dimensions (H descripteurs), il est parfois possible de le « réduire » au cas précédent (tableau à deux dimensions), ce qui entraîne en général une **perte d'information** (eg sur la **structure** du tableau). Selon les questions traitées, on peut, alternativement :

(a) calculer des « **faces** » (ie des « **marges** » de dimension 2) de façon à se ramener à deux dimensions ;

(b) réordonner les k critères de croisement en privilégiant l'un d'entre eux (eg le « **temps** »), et en faisant le produit (cartésien) des $H - 1$ critères restants (cf aussi **vectorialisation d'une matrice**), ce qui ramène à un tableau à deux dimensions ;

(c) considérer directement T comme un « *tenseur* » ad hoc et d'utiliser les propriétés de l'algèbre tensorielle. En effet, T est alors identifiable à l'ensemble des projections d'une **forme multilinéaire** (forme H -linéaire), qui peut lui être associée.