

MATRICE CENTRÉE (A3, C1, F1)

(05 / 12 / 2019, © Monfort, Dicostat2005, 2005-2019)

L'expression **matrice centrée** est parfois donné à une **matrice** de **données** dont les marges sont nulles. De telles matrices sont utilisées eg en **analyse des données** ou dans l'étude de la **régression linéaire**.

(i) Soit $(\Omega, \mathcal{F}, P)$ un **espace probabilisé** et $X : \Omega \mapsto M_{NK}(\mathbf{R})$ une matrice aléatoire (cf **matrice d'observation**) associée à N **observations** et K **variables**. On pose :

$$(1) \quad X = (x_{nk})_{(n,k)} = (X_1 \text{ /// } X_N) = [x_1, \dots, x_K],$$

où /// désigne les N-2 lignes intermédiaires entre les lignes X_1 et X_N .

Le vecteur $\bar{X}_N$ des K moyennes empiriques des variables s'écrit (cf **moyenne empirique**) :

$$(2) \quad \bar{X}_N = e_N' X / e_N' e_N = [\bar{x}_1, \dots, \bar{x}_K],$$

avec $\bar{x}_k = N^{-1} \sum_{n=1}^N x_{nk}$ (moyenne empirique de la variable d'**indice** k).

Le vecteur $\bar{x}_K$ des N moyennes des observations s'écrit :

$$(3) \quad \bar{x}_K = X e_K / e_K' e_K = (\bar{X}_1 \text{ /// } \bar{X}_N)',$$

avec $\bar{x}_n = K^{-1} \sum_{k=1}^K x_{nk}$ (moyenne empirique de l'observation d'indice n).

On appelle **matrice centrée** la matrice Y associée à X et dont le terme général est :

$$(4) \quad y_{nk} = x_{nk} - \bar{x}_n - \bar{x}_k + \bar{X},$$

où $\bar{X} = e_N' X e_K / \{(e_N' e_N) (e_K' e_K)\}$ est la moyenne générale des termes x_{nk} .

(ii) On montre que :

(a) $e_N' Y = 0$ et $Y e_K = 0$ (d'où $e_N' Y e_K = 0$), propriété qui explique la terminologie ;

(b) la **matrice de dispersion** empirique S_N (des variables) de X s'écrit simplement sous la forme du produit $Y' Y$.