

## VARIABLE D'INTÉRÊT (C1, E, G11, H7, J9)

(18 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) Une **variable d'intérêt** est une **variable statistique** ou une **variable aléatoire** dont on étudie les propriétés à titre principal, par distinction avec une **variable « secondaire »**, qui peut en faciliter l'étude, ou avec une **variable instrumentale**, etc.

(ii) Au sein d'un **domaine de connaissance** donné, l'étude d'un **phénomène** comporte souvent un choix parmi les **variables** qui le décrivent. Les raisons en sont multiples : **adéquation** au problème, **observabilité**, absence d'**erreurs**, de **censure**, etc. La notion de variable d'intérêt est donc relative à la **représentation statistique** considérée.

Ainsi (cf **relation fonctionnelle**) :

(a) dans un **modèle de régression linéaire** standard écrit dans l'**espace des observations** sous la forme  $y = Xb + u$ , où  $E u / X = 0$  et  $V u / X = \sigma^2 I_N$ , le **couple aléatoire** observé  $(X, y)$  joue un rôle principal, tandis que la **perturbation aléatoire**  $u$ , qui résume l'effet des variables omises, joue un rôle conceptuel accessoire, malgré son utilité ou son caractère nécessaire.

Dans ce modèle, la variable d'intérêt est, par essence, la variable endogène (ou variable de gauche), que l'on cherche à « expliquer », mais la **spécification** du modèle attribue aussi à la variable exogène (variable de droite) un rôle important ;

(b) il en va de même pour un **modèle d'interdépendance**, dans lequel le couple  $(X, Y)$  des données observées est d'intérêt, la matrice  $U$  des perturbations étant secondaire.

Dans ce modèle, la variable d'intérêt est généralement le couple  $(X, Y)$  constitué des observations de la **variable exogène** (vectorielle à valeurs dans  $\mathbf{R}^K$ ) et de la **variable endogène** (vectorielle à valeurs dans  $\mathbf{R}^G$ ).