

Analyse de données

Sous l'appellation "analyse des données" on entend en statistiques un ensemble de méthodes descriptives multidimensionnelles. Ces méthodes nécessitent en général une information organisée de la façon suivante: " n " individus statistiques (entités spatiales, ménages, firmes...) décrits par " p " variables. Ces méthodes permettent de résumer l'information contenue dans des tableaux de données aux dimensions importantes (tableau n lignes x p colonnes). On peut distinguer deux " familles " de méthodes:

Les analyses factorielles : elles consistent à transformer le tableau de données initial en un nouveau tableau contenant la même information, mais sous forme *hiérarchisée*. Il est composé *d'axes factoriels*. Le premier axe factoriel correspond à la combinaison linéaire des variables initiales qui différencie au maximum les individus entre eux. Il est de *variance maximum*. Les axes factoriels sont indépendants les uns des autres et classés en fonction de leur variance. En général il suffit d'un petit nombre d'axes factoriels (trois ou quatre) pour rendre compte de l'essentiel de l'information contenue dans le tableau initial. L'interprétation de ces axes factoriels permet de mettre en évidence la forme des interrelations entre les variables étudiées et les ressemblances et dissemblances entre les individus relativement à ces variables. Les deux méthodes les plus communément utilisées sont l'analyse en *composantes principales* (adaptée pour des données hétérogènes combinant des variables exprimées dans des échelles de mesure différentes, ou encore pour des variables exprimées en pourcentages) et l'analyse des *correspondances* (adaptée pour des tableaux de contingence ou de [variables qualitatives](#)).

Les classifications : elles permettent d'élaborer des typologies et de regrouper les individus par classes en fonction de leurs ressemblances par rapport à l'ensemble des variables. Un critère souvent utilisé du point de vue technique est de chercher la classification qui minimise la variance intraclasse (variabilité entre les individus d'une même classe) et maximise la variance interclasse (la variabilité entre les classes). Les méthodes les plus classiques sont la classification *ascendante hiérarchique* et la classification par *nuées dynamiques*.

Voir aussi :

Variables quantitatives

Bibliographie