

ANALYSE DES PROXIMITÉS (K10)

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

(i) L'**analyse des proximités** est une méthode d'**analyse des données** consistant à construire un **ensemble** :

(a) soit constitué de classes d'**unités statistiques**,

(b) soit constitué de classes d'**observations** de **variables** relatives à ces unités,

à partir des **proximités** existant entre elles (cf eg **classification automatique**).

(ii) Les proximités en questions (cf **indice de dissimilarité**, **indice de similarité**) sont basées sur les **valeurs des variables** observées sur les unités précédentes. Ces valeurs étant le plus souvent numériques (additives), on peut alors associer à chaque classe une valeur agrégée, ou « **magnitude de classe** » (cf **agrégation, statistiques, statistique**).

(iii) Les classes obtenues par ces méthodes sont, en principe, disjointes entre elles (cf **partition**).

Cependant, si l'appartenance des unités à des classes données est incertaine, et peut être considérée comme floue (cf **théorie des parties floues**), on en déduit des classes elles-mêmes floues (cf **partie floue**).

Ainsi (sociologie : économie), les activités économiques françaises ont été structurées en termes de « filières de production » à partir des comptes nationaux détaillés :

(a) d'une part, le tableau des échanges interindustriels (TEI) : celui-ci décrit la répartition des consommations intermédiaires (CIHTD) $(c_{ij})_{(i,i)}$ des branches j en produits i ;

(b) d'autre part, le tableau des investissements des branches en produits (IBP) : celui-ci décrit la répartition de la formation brute de capital fixe (FBCF) $(f_{ij})_{(i,i)}$ des branches j en produits i .

Certaines branches j appartenant à certaines filières, se sont trouvées en liens forts avec d'autres appartenant à des filières différentes, les proximités correspondantes impliquaient de les affecter à des filières distinctes (mais non disjointes).