

CLASSIFICATION STATISTIQUE (I7)

(19 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

Une **classification statistique** est une opération par laquelle une **population** d'**unités statistiques** est partitionnée (classification stricte) ou approximativement partitionnée (classification floue) entre classes réputées homogènes (cf **homogénéité**, **partition**, **théorie des parties floues**).

Un **problème de classification** est donc un **problème de décision** statistique basé, à la fois, sur :

(a) une **population** (resp un **échantillon**) ;

(b) une liste pré-établie de **variables** observées sur les éléments (unités statistiques) composant la population (resp l'échantillon). Ces variables sont supposées pertinentes relativement au **problème statistique** considéré ;

(c) une notion de « **proximité** » (**distance**, **indice de similarité**) entre éléments, calculable à partir des variables précédentes.