

## STRUCTURE LATENTE (K, K12)

(12 / 06 / 2020, © Monfort, Dicostat2005, 2005-2020)

Les méthodes statistiques se réfèrent généralement à des « données » **observables**, ie à l'observation ou à la « mesure » de grandeurs attachées à des **unités statistiques** au sein d'une **population** d'unités. Les seules informations inobservables sont généralement les **paramètres** associés à un **modèle**, ou encore certaines variables « résiduelles » (cf eg **perturbation aléatoire**).

Cependant, dans certaines **situations statistiques**, des données peuvent se trouver être, par essence, implicites, non observables avec les moyens d'un **système statistique**.

(i) Au sens large, une **structure latente**, ou **structure sous-jacente**, ou encore **structure cachée**, est constituée (cf **inobservabilité**, **structure**) :

(a) d'un **ensemble** de **variables** (**caractères statistiques**, **attributs**, etc) non directement observables sur certaines unités statistiques. Les observations de ces variables sont indisponibles (impossibilités, erreurs ou altérations diverses) ;

(b) d'un ensemble de relations inobservables entre ces variables, ou entre ces variables et d'autres variables observables.

Le **problème statistique** est le suivant : déterminer des hypothèses « probabilistes » tq ces **variables cachées**, aussi bien que les relations dans lesquelles elles interviennent, puissent être estimées à partir d'un **échantillon** de données observables sur ces unités (cf **observabilité**).

Le procédé de **passage de l'inobservable à l'observable** est un procédé fondamental en **Statistique**, donc dans chaque **domaine de connaissance** :

(a) **relation fonctionnelle** ou **modèle de régression**, pour lesquels sont inobservables le **paramètre d'intérêt**  $b$  et la **perturbation**  $u$  ;

(b) estimation en présence d'une censure ou d'une loi tronquée (cf **censure**, **troncature**).

(ii) Cependant, on réserve plus spécifiquement l'expression « **analyse des structures latentes** » à deux grandes catégories de méthodes statistiques :

(a) l'**analyse des données**. Les méthodes, algébriques ou géométriques, qui en relèvent font jouer un rôle de **structure latente** aux divers « **facteurs** » de l'analyse : **analyse factorielle** (ou **analyse en facteurs communs et spécifiques**), **analyse en composantes principales**, **analyse factorielle des correspondances**, **analyse sphérique**, etc ;

(b) le **modèle de structures latentes**. Dans ce cadre, les observations sont supposées être des **va** multinomiales multidimensionnelles (eg des variables dichotomiques, à valeurs dans  $\{0, 1\}$ ) (cf **loi multinômiale**, **variable binaire**). Ce type de modèles se fonde sur l'hypothèse que la **loi de probabilité** des observations est un **mélange de lois** multidimensionnelles, mélange (inobservable) dont la « *loi mélangeante* » (qui définit les « poids ») n'est autre que la **loi des variables latentes**.