

OBSERVATION (C1, F)

(31 / 01 / 2020, © Monfort, Dicostat2005, 2005-2020)

« L'adaptation de l'observation à une théorie relève de la falsification, celle de la théorie à des faits relève de la science » (cf **étude scientifique**)

« Les faits sont têtus » (tant qu'ils ne sont pas altérés)

Le terme « **observation** » fait référence :

(a) soit à l'**action** consistant à observer un **phénomène** (cf **système statistique, production statistique**) ;

(b) soit au **résultat** de cette action, ou **donnée**, qui est généralement une valeur observable prise par une **variable** associée à une **unité statistique** (cf **espace d'observation, observabilité, observable**).

(i) Dans ce dernier sens, Ω représente l'**ensemble fondamental** considéré et $\xi : \Omega \mapsto \mathcal{X}$ une **variable** (descripteur, etc) considérée comme **application** de Ω dans un ensemble de valeurs \mathcal{X} . On suppose généralement que toute valeur $\xi(\omega) \in \mathcal{X}$ est **observable** par l'intermédiaire d'un **système d'observation**, ou **dispositif d'observation** (cf **production statistique, dispositif expérimental**).

Cette propriété d'**observabilité** est fondamentale pour l'activité statistique. En effet, c'est à partir des valeurs observées $\xi(\omega)$ qu'il est possible de calculer des « fonctions des observations » (cf **statistique**) permettant l'**inférence statistique** (estimation, tests, classifications, etc).

La notion d'observation est donc liée à la nature des variables observées (**variable qualitative** ou **variable quantitative**), ainsi qu'au dispositif d'observation, aussi appelé **système d'observation (dispositif expérimental, etc)**.

(ii) La **spécification** d'un **modèle statistique** comporte parfois des variables **inobservables** (cf **variable cachée, variable latente, perturbation aléatoire, etc**).

En effet, la théorie émanant de l'**homme de l'art** peut, dans certains cas, être amenée à définir des variables non observables mais dont l'effet, ou l'« action », sur les autres variables (observables) semble avérée. Ces variables peuvent aussi correspondre à des « **concepts abstraits** » satisfaisants pour expliquer le phénomène, mais non utilisables.

Trois situations peuvent être traitées :

(a) la **procédure statistique** mise en oeuvre peut, d'elle-même, résoudre ce problème. Ainsi, la **perturbation aléatoire** u d'un **modèle de régression** tq $y = F(b) + u$ est un moyen commode pour analyser statistiquement ce modèle. Elle n'intervient pas dans la nature du lien F reliant variables expliquées (causées) et variables explicatives (causales) ;

(b) dans d'autres cas (**simulation, expérimentation** factice), il est possible de définir des fonctions reliant, de façon connue, les observables aux inobservables ;

(c) enfin, lorsque les fonctions précédentes ne sont pas connues, la résolution du modèle statistique peut, elle-même, conduire à une solution (liste des variables observables) globalement indépendante des inobservables (cf aussi **lacune, observation manquante**). Ainsi, lorsque le modèle théorique admet que η (observable) dépend de ζ (inobservable) selon $\eta = \psi(\zeta)$, et que ζ dépend à son tour de ξ (observable) selon $\zeta = \varphi(\xi)$, une substitution évidente permet de relier les observables selon une relation tq $\eta = \psi \circ \varphi(\xi)$, dont peut se déduire une relation fonctionnelle (modèle) de la forme $\eta = F(\xi) + \varepsilon$.

(iii) Ce qui est communément (médias, grand public) appelé, sous forme plurielle, « **statistiques** » n'est autre que l'ensemble des valeurs observées sur des unités participant à un phénomène, ie l'ensemble des observations disponibles relatives à ce phénomène (cf aussi **diffusion de l'information**).

Ainsi en est-il eg (physique) des observations météorologiques, (biologie) des cas de maladies épidémiques ou périodiques, (écologie) du dénombrement d'un cheptel, (sociologie : économie) des statistiques démographiques (dénombrement de la population, des naissances ou des décès, flux migratoires), données sur les prix, le chômage, le coût de la construction, (sociologie : jeu démocratique) résultats d'une élection.

C'est à partir de données de ce type (eg échantillon) que se construit la notion « technique » de **statistique** (concept statistique).