

STATISTIQUE (G1, I1, O)

(12 / 09 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) Le terme de « **statistique** » possède trois sens distincts :

(a) la « **Statistique** » (ici exprimée avec une majuscule) se réfère à l'**ensemble des concepts et méthodes statistiques** mis en oeuvre par le **statisticien** (science statistique ou savoir statistique) (cf aussi **production statistique, consommation statistique**). On peut alors distinguer entre :

(a)₁ **Statistique descriptive** ;

(a)₂ Statistique « observationnelle », ou « opérationnelle » : **expérience aléatoire, sondage** ;

(a)₃ **Statistique inférentielle** ;

(b) le **concept « technique » de « statistique »** est une **application mesurable** « fonction » des **observations**. Ce concept permet de définir diverses grandeurs d'intérêt (eg une **caractéristique** légale empirique : **moyenne, écart-type**, etc) à partir d'un **échantillon d'observations**. Il sert souvent pour l'**estimation** du **paramètre** d'une **loi** ou pour divers **tests** le concernant (cf **test d'hypothèses**) ;

(c) les « **statistiques** », exprimées au pluriel, désignent l'**ensemble des données observées**, ou **observations**, soit sous forme de données détaillées (ie individuellement attachées à des **unités statistiques**), soit présentées (ou classées) de façon agrégée en fonction de divers critères d'intérêt (**tableau statistique**). On peut notamment distinguer entre **statistiques produites, statistiques demandées** et **statistiques utilisées**.

C'est ce dernier sens qui est le plus communément compris à l'extérieur de la sphère scientifique : statistiques relatives à la météorologie, à la démographie, aux prix, etc.

(ii) La Statistique, comme science ou comme « savoir », intervient dans la plupart des étapes du travail scientifique. Ainsi :

(a) chaque **domaine de connaissance** comporte diverses « branches », dont chacune s'intéresse à des **phénomènes** particuliers :

(a)₁ physique, eg : astronomie (mouvement, masse ou énergie des corps célestes), physique des particules ou ondulatoire (mouvement, masse ou énergie corpusculaire) ;

(a)₂ biologie, eg : physiologie (échanges intercellulaires, multiplication cellulaire), épidémiologie (contaminations) ;

(a)₃ écologie (prédation, compétition ou parasitage entre espèces écologiques) ;

(a)₄ psychologie (comportement de l'enfant, attitudes vis-à-vis du risque) ;

(a)₅ sociologie, eg : économie (théorie du consommateur, théorie des organisations) ;

(b) la connaissance ou la **compréhension des phénomènes** nécessite un préalable : leur **observation**. Cette dernière peut, selon le domaine et le phénomène considérés, s'effectuer de deux façons :

(b)₁ soit par « **relevés spontanés** », ie effectués sans méthode particulière, ou d'une façon dénuée d' « **objectif statistique** » : sous-produits d'activités diverses : eg fichiers administratifs, fichiers privés, bibliothèques diverses, documents anciens (eg chroniques, archives, etc) ;

(b)₂ soit par l'une des deux méthodes d'observation statistique de base :

(b)₂₁ **expérimentation** : physique (résistance des matériaux), biologie (stimuli appliqués à des **unités expérimentales** et étude de leurs réponses), écologie (agronomie : rendement des surfaces cultivées, implantation de faunes ou de flores hors de leur milieu antérieur, etc) ;

(b)₂₂ **sondage** : comparaison des effets d'un **traitement** - cf **stimulus**, etc - sur un **échantillon test** avec un **échantillon témoin**, marquage d'unités composant diverses faunes, enquêtes auprès de personnes physiques ou morales.

La Statistique intervient donc, en premier lieu, aux deux niveaux précédents, qui concernent l'observation des phénomènes (cf **production statistique**) ;

(c) le travail scientifique débute généralement par la **description d'un phénomène** (cf **Statistique descriptive**). Cette description concerne des « **unités statistiques** » sur lesquelles sont observées diverses **variables**, parfois appelées « **descripteurs** » ou **attributs**. On distingue trois grands types de telles variables :

(c)₁ les **variables quantitatives** (ou variables numériques) ;

(c)₂ les **variables qualitatives** (ou **variables quantales**) ;

(c)₃ les **variables morphologiques**.

Ces descriptions accumulées permettent de constituer des « stocks » d'**observations**, ou de « **données observées** », qui sont parfois très importants (**grandes bases de données**) ;

(d) la description d'un phénomène particulier revient notamment à sélectionner une liste de variables pertinentes (cf **variable d'intérêt**). L'un des fondements du travail scientifique consiste souvent à assimiler ces dernières à des **variables aléatoires**, donc à des variables supposées engendrées par une **loi multivariée** (eg une **loi de probabilité multidimensionnelle**). La connaissance de cette loi équivaut donc, par principe, à une complète connaissance du phénomène.

Le plus souvent, pour des raisons intellectuelles ou « budgétaires » au sens large (**parcimonie**, limites d'appréhension des phénomènes, moyens d'observation limités, contraintes « techniques » diverses, etc), l'**homme de l'art** cherche seulement à

mettre en relation certaines variables entre elles. Ainsi, les notions suivantes sont destinées à servir ces objectifs :

(d₁) les notions de **corrélation**, d'**association** ou d'**interaction** sont relativement neutres du point de vue causal : elles mesurent la « force » des relations entre variables, mais non pas la force d'éventuelles relations de cause à effet ;

(d₂) la notion de **relation fonctionnelle** (eg fonctions de **régression** ou d'**interdépendance**) cherche, au contraire, à établir un lien d'influence, voire de **causalité**, entre les variables prises en compte.

Deux notions de « **loi** », relatives à un même phénomène, peuvent être qualifiées de « **loi scientifique** » : soit la **loi de probabilité** de l'**ensemble** des variables considérées (qui est généralement une **loi multivariée**), soit la **relation fonctionnelle** qui peut être exhibée au sein de ces variables (eg **modèle à équations simultanées**) ;

(e) les notions équivalentes de **modèle statistique**, de **structure statistique**, ou de **représentation statistique**, constituent une résultante importante de l'étape précédente. Ces notions permettent de construire un « vrai » **problème de décision**, ie un problème comportant généralement une phase d'**optimisation** spécifique (cf aussi **décision statistique**). On peut de la sorte résoudre diverses questions statistiques élémentaires : **classement** ou **classification** des données, évaluation des **paramètres** des modèles (**estimation**) (cf aussi **paramètre d'intérêt**, **paramètre principal**), vérification de certaines hypothèses (**test d'hypothèses**), vérification de la stabilité des modèles (cf **robustesse**), études « internes » au problème ou aux données (**interpolation**), études « externes » au problème ou aux données (**prévision**, **projection**, **extrapolation** ou rétropolation), etc ;

(f) les activités précédentes conduisent à mettre à la disposition de l'**homme de l'art** des « maquettes », ou « modèles réduits », plus ou moins sophistiqué(e)s, qui résument de façon intelligible, la **structure**, le **fonctionnement**, ou encore l'**évolution** du phénomène (cf aussi **niveau**, **répartition**, **évolution**). Ces outils peuvent alors servir à satisfaire divers besoins humains. Les propriétés des maquettes peuvent, en effet, être utilisées pour prévoir ou anticiper des évolutions ou des « ruptures » (eg inondations, avalanches, incendies, instabilité des sols ou séismes, vulcanologie), pour agir le cas échéant sur certains phénomènes (prévention ou précaution relatives à des phénomènes naturels ou épidémiologiques ; politique économique ; politique sociale).

(iii) On peut considérer que la Statistique est un **savoir-trépied** basé sur les entités conceptuelles suivantes :

(a) l'**état du « milieu » (ou « monde »)** étudié (ou à étudier). Dans l'approche statistique globale, cet état est supposé déterminé (« décidé ? ») par une entité indéfinie, appelée par commodité **Nature**. Cet état est donc la « **situation véritable** » de ce monde : il est caractérisé par sa structure, son fonctionnement et son évolution. Le **statisticien** considère que cet état est « gouverné » par une **loi** imposée (ou « décidée ») par la Nature. Cette loi, a priori complexe (cf **complexité**), joue le rôle de **paramètre** dont la connaissance est considérée comme

fondamentale. Mais cette loi est **inobservable** en général : seules ses « manifestations » **observables** peuvent être analysées, qu'elles soient « spontanées » (relevés divers, non statistiques, relatifs à des phénomènes naturels) ou qu'elles soient « provoquées » (**plans d'expérience, plans de sondage**) (cf supra). Cette démarche, basée sur la conception d'un univers **aléatoire** ou **stochastique**, s'est, à l'usage, révélée très fructueuse pour la science ;

(b) l'**observation** et la **mesure**, qui constituent une représentation du monde étudié, tq l'**Homme** (ou son mandataire) peut le percevoir à travers ses sens (perception sensorielle) et à l'aide de divers moyens d'investigation ou dispositifs d'observation (cf **dispositif expérimental**), ainsi qu'à travers sa propre réflexion mentale (formation du cerveau et de l'entendement) ;

(c) l'**opinion (personnelle)** du statisticien ou de l'homme de l'art, relativement aux phénomènes étudiés ; on peut lui associer une **utilité** (cf aussi **fonction d'utilité**). Cette troisième composante du trépied statistique a engendré une école de conceptualisation, appelée **école bayésienne** (cf **probabilité a priori, probabilité personnelle, probabilité subjective**). L'opinion en question est relative au « paramètre » précédent : cette opinion attribuée à ce paramètre (qu'il s'agisse de la loi gouvernant le phénomène ou du paramètre de cette loi) divers degrés de vraisemblance, eux-mêmes synthétisés dans la notion de **loi a priori**.

Les trois composantes précédentes (paramètre, observation, opinion) constituent ainsi les bases de l'**inférence statistique**.