

ERREUR (A à O)

(09 / 06 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) Outre son sens ordinaire, le mot **erreur** peut comporter, en science comme en statistique, diverses acceptions. Une erreur est, en général, une cause ou une **action** qui altère :

- (a) soit la réalisation d'une épreuve aléatoire (eg **expérience aléatoire**) ;
- (b) soit l'**observation** d'un **phénomène** ;
- (c) soit la conception même de la **loi** qui est censée régir ce phénomène.

Une erreur est souvent liée aux limites d'un dispositif d'observation (**système statistique**, **dispositif expérimental**, etc) (cas (b)).

(ii) Parmi les nombreux contextes, statistiques ou non, dans lesquels intervient ce mot, on peut relever les suivants :

(a) **erreur conceptuelle**, ou **erreur de concept** : dans un **domaine de connaissance** donné, une notion est utilisée à la place d'une autre. Ceci peut arriver lorsqu'on substitue un concept **observable** à un autre, qui est **inobservable** (cf **observabilité**) ou mal observable (cf eg **modèle à erreurs sur les variables**) ;

(b) **erreur de mesure**, ou **erreur d'observation** : réponse erronée à un **questionnaire** (mémoire défaillante, fausse réponse ou réponse approximative intentionnelles), **non réponse**, **observation manquante**, etc. Cette erreur peut souvent être détectée (traitement des observations manquantes, ou **lacunes**) et corrigée (cf eg **méthode de WARNER**) ;

(c) **erreur de calcul** : cette erreur peut provenir d'un calcul analytique ou d'un calcul numérique (arrondissement, etc). Le premier type d'erreur peut généralement être corrigé (critique académique, vérification ou révision des calculs). Le second est, généralement, d'ampleur négligeable ;

(d) **erreur sur une variable** statistique, ou **erreur sur des observations** de celle-ci. On peut traiter ce problème eg avec un **modèle à erreurs sur les variables** ;

(e) **erreur sur une équation**, ou **perturbation aléatoire**, dans une **représentation statistique**. Il s'agit d'un procédé destiné à « regrouper » dans un « résidu » un ensemble de variables ou d'observations portant sur celles-ci. La théorie considère alors, en effet, que ce résidu peut s'exprimer en termes de perturbation aléatoire dont l'effet est négligeable « en moyenne » (cf eg **régression**), ie en termes de **valeur centrale** ;

(f) **erreur d'estimation**, inhérente à la **variabilité** d'un **échantillon aléatoire**, lorsqu'un tel **jeu de données** est utilisé. Cette erreur peut généralement se concevoir soit au sens du **biais**, soit au sens de la **dispersion** (cf aussi **écart quadratique moyen**). On l'appelle parfois **erreur aléatoire**, ou **erreur d'échantillonnage**, alors que ni le biais, ni la dispersion, n'est une **variable aléatoire**.

Lorsque les données ne constituent pas un échantillon aléatoire, ie n'ont pas été obtenues selon un « processus aléatoire » (ou pouvant être considéré comme tel), ce type d'erreur est parfois difficile à appréhender : on parle d'**erreur non aléatoire**, ou d'**erreur systématique** (cf eg **sondage par choix raisonné**, **sondage par unités-types**, **sondage par quotas**). Cette **situation statistique** survient aussi, parfois, lorsque les données n'ont pas été « construites » par le **statisticien**, mais « imposées » de façon « exogène » (cf **production statistique**) : eg observations anciennes (ie à caractère « historique »), observations « uniques », etc ;

(g) **erreur de spécification d'un modèle statistique**. Ce type d'erreur peut consister en une erreur :

(g)₁ sur la « vraie » **population** étudiée ;

(g)₂ sur la **forme analytique** des **lois de probabilité** mises en jeu ;

(g)₃ sur la forme analytique des relations entre variables (eg endogènes et exogènes) (cf **relation fonctionnelle**, **fonction de régression**, **fonction d'interdépendance**) ;

(g)₄ sur la liste des variables (eg endogènes et exogènes) ;

(g)₅ ou, plus généralement sur la **famille** même des **lp** retenues ;

(h) **erreurs d'espèce** associées à un (problème de) **test d'hypothèses**. Ces concepts se réfèrent plutôt à une notion de **risque**, ou de « pari », accepté ou engagé par le statisticien, et relatif à une **hypothèse statistique** ;

(i) **erreur de prévision** portant sur un **processus** ou une **trajectoire** (cf **série temporelle**). Cette erreur peut souvent être assimilée (a) ex ante, à une erreur d'**estimation**, et (b) ex post, à un « **écart** » entre prévision et réalisation (cf **écart de prévision**) ;

(j) **erreur d'estimation**, d'**observation** ou altérant une **procédure statistique** lorsque celle-ci utilise des **observations aberrantes** ou des **variables aberrantes** (eg variables inadéquates ou non pertinentes) (cf **aberration**) ;

On peut encore parler d'erreur d'estimation lorsqu'on utilise volontairement un estimateur biaisé (cf **estimateur sans biais**), notamment parce qu'il est plus simple à calculer. Diverses techniques, eg la **méthode de QUENOUILLE-TUKEY**, permettent de réduire un biais de cette nature ;

(k) **erreur d'hypothèse**, soit lorsqu'une méthode statistique est appliquée à tort, ie sans que les hypothèses qui la sous-tendent n'aient fait l'objet d'une vérification préalable (**test d'hypothèses**), soit lorsque l'une des hypothèses faites est de nature mathématique (continuité, concavité, etc) et n'est pas vérifiée.

Ceci est le cas eg lorsqu'on utilise les propriétés de la **loi normale** sans avoir auparavant testé l'hypothèse de **normalité** (pour de petits échantillons) ou l'hypothèse de **normalité asymptotique** (pour de grands échantillons).

L'objectif d'une méthode robuste (cf théorie de la **robustesse**), parfois appelée **méthode affranchie** (libérée d'hypothèses trop contraignantes), est d'employer une **procédure statistique**, et notamment une **règle de décision** (estimation, tests, prévision), qui ne soit pas « trop » altérée lorsque les hypothèses de base retenues par le statisticien s'écartent des conditions réelles qui génèrent les données : ces conditions sont en général complexes, et parfois inconnues ;

(l) **erreur totale (théorie des sondages)** qui représente la somme de l'erreur d'**échantillonnage** et de l'erreur d'**observation** ;

(m) **erreur d'expérimentation (théorie des plans d'expérience)** qui, au cours d'une expérience planifiée, perturbe le **dispositif expérimental** (dispositif peu fiable dans certaines circonstances). Ainsi, certains outils (machines, instruments de mesure, etc) peuvent se dérégler au fur et à mesure de leur utilisation ;

(n) **loi des erreurs** : autre nom parfois donné à la **loi normale**.