

## QUESTIONNAIRE (L, M, O)

(08 / 01 / 2020, © Monfort, Dicostat2005, 2005-2020)

Utilité lors d'une investigation scientifique, un **questionnaire** (« statistique ») peut être considéré

(a) soit comme un ensemble d'**informations** ou d'**interrogations** formant un thème d'étude (eg analyse d'un **phénomène** donné). Ces informations ou interrogations sont habituellement représentées par diverses variables : ces variables peuvent être des **variables quantitatives** ou des **variables qualitatives** ;

(b) soit, une fois « complété », comme un ensemble d'**observations** relatives à ces variables et « mesurées » sur des **unités statistiques** (**unités expérimentales**, **unités de sondage**). Il représente donc alors « batterie » de données informatives concernant ces unités.

Ces données sont ainsi des « réponses » (directes ou indirectes) à des « questions » (posées par l'**homme de l'art** ou le **statisticien**), lesquelles font l'objet de l'investigation.

(i) Un questionnaire est donc très généralement un résultat de base de la **production statistique**. Il peut :

(a) résulter de diverses activités anthropiques non « statistiques ». Ainsi, un « formulaire administratif » (eg déclaration fiscale, demande de service, etc) est élaboré à des fins de gestion administrative, mais peut aussi être utilisé à des fins statistiques. De même, des relevés de **mesures** résultant d'observations diverses (météorologie, marquage d'animaux, etc) ne sont pas nécessairement organisés ni conçus par des statisticiens, mais peuvent cependant, très souvent, être traités statistiquement. Enfin, les **grandes bases de données** (notamment électroniques) sont généralement constituées à partir d'informations disponibles (serveurs informatiques, réseaux sociaux, etc) et doivent faire l'objet de traitements appropriés préalablement à leur utilisation ;

(b) être associé à l'une des deux grandes activités de production statistique :

(1) **sondage** (ou **recensement**) : cette activité conduit notamment à élaborer un questionnaire d'« **enquête** » ;

(2) **plan d'expérience** (par extension) : cette activité conduit notamment à élaborer un « protocole d'observation » qui, à la fois, spécifie ou décrit l'**expérience** à réaliser (tests de laboratoire, rendements de parcelles de terrain, stratigraphie archéologique, etc), et « consigne » les résultats de l'**observation** (observation des diverses variables) sur des documents ad hoc, généralement formalisés (« **relevés d'information** »).

(ii) Les **questions** faisant l'objet d'un questionnaire ont pour objectif l'observation d'un certain nombre de **variables**, généralement communes à l'ensemble des unités interrogées (comparabilité).

Les **réponses aux questions** constituent autant d'informations élémentaires, appelées « observations » des variables précédentes. Une **réponse** peut être :

(a) **directe** ou **indirecte**. Dans la réponse directe, l'unité répond elle-même et sur elle-même. Dans une réponse indirecte, l'unité répond pour une autre unité qui peut, éventuellement, être de « nature » différente ; ou encore, l'expérimentateur lui-même répond aux questions - qu'il s'est posées - au cours du relevé des observations effectuées sur des unités expérimentales. L'« unité répondante » peut ainsi fournir des informations la concernant directement ; elle peut aussi parfois fournir des informations indirectes, ie concernant d'autres unités (eg un « chef de ménage » renseigne sur l'ensemble des personnes du ménage, un chef de service sur l'ensemble du personnel à sa disposition, etc) ;

(b) **conditionnelle** ou **non conditionnelle**. Une réponse conditionnelle est subordonnée à la réalisation d'une modalité appartenant à une question précédente (questionnaire à structure hiérarchisée, ou arborescente). Une réponse non conditionnelle concerne dans le cas contraire (questionnaire à structure plate, ou structure linéaire) ;

(c) **sincère** ou **non sincère**. Si la réponse à la question ne soulève pas de problèmes de confidentialité (perçus psychologiques ou sociologiques), la réponse peut être sincère, donc exacte, au erreurs d'appréciation près (défauts de mémoire, d'intelligence de la question, etc). Diverses possibilités correspondent à une réponse non sincère : non-réponse pure et simple (ie questionnaire « vide »), non réponse partielle (questionnaire « rempli » en partie), mensonge (réponse fausse) ou triche (réponse fausse mais valorisante), évasion (réponse peu précise), etc ;

(iii) La **structure interne** d'un questionnaire est définie par la **disposition** des questions. En pratique, cette disposition varie souvent : mode linéaire (questions successives indépendantes), mode emboîté (questions hiérarchisées, ou « en cascade ») ou même mode croisé (questions dont les réponses constituent un tableau qui croise des critères d'intérêt). Dans certains cas (modalités réalisées, ou réponses sans objet), des « sauts » entre questions peuvent être précisés.

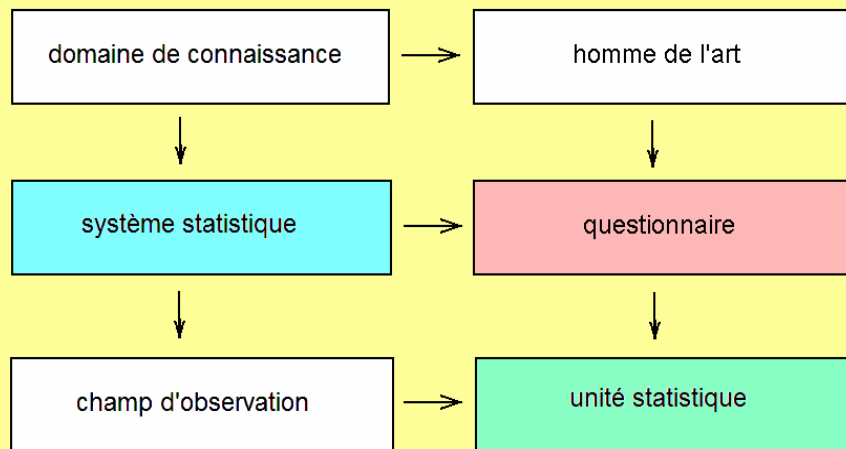
(iv) Les observations disponibles in fine (questionnaires de sondage, relevés d'expérience) peuvent être différentes des observations primaires décrites ci-dessus.

Les différences proviennent souvent d'**erreurs matérielles** : eg lors de la transcription ou de la codification des données, de la saisie informatique, de la reconnaissance de leurs formes, des contrôles de validité assortis de corrections, etc.

Par ailleurs, ces observations peuvent ne pas être toutes celles attendues (eg **non réponse** de certaines unités, le plus souvent), ou être affectées d'erreurs diverses (**erreur** d'observation).

(v) Pour un **système statistique** donné, un questionnaire représente un **moyen de collecte** privilégié d'informations élémentaires relevant de son champ. Il constitue donc un « pont » entre le domaine de connaissance concerné et l'observation de phénomènes divers relevant de ce domaine (cf schéma ci-après).

place du questionnaire dans l'investigation scientifique



(vi) Si, après observation, on note les questionnaires (objets matériels) à l'aide d'un indice  $n = 1, \dots, N$ , l'ensemble  $Z_n$  des données consignées dans chacun d'eux est assimilable à un « **échantillon** »  $Z = (Z_1, \dots, Z_N)$ . L'« observation »  $Z_n$  est constituée d'« observations » élémentaires  $(X_n, K_n)$  relatives à des variables définies, le plus souvent, dans le questionnaire (**variables quantitatives** ou « quantités »  $\xi$ , **variables qualitatives** ou « codes »  $\kappa$ ). S'il existe  $K$  variables quantitatives et  $H$  variables qualitatives, les  $X_n$  sont à valeurs dans  $\mathbf{R}^K$  et les  $K_n$  dans  $\mathbf{R}^H$  (ou, en pratique, dans les ensembles « décimaux » finis  $\mathbf{D}^K$  et  $\mathbf{D}^H$ ).

(vii) La place du questionnaire dans l'**investigation statistique** peut donc se représenter par à trois niveaux : domaine de connaissance (homme de l'art), système statistique (**statisticien**, questionnaire) et champ d'observation (unités statistiques).

En effet, le domaine de connaissance nécessite un système statistique afin d'analyser un champ d'observation donné, notamment dans le but d'agir sur celui-ci (lorsque cela est possible) ou sur des champs comparables.

Le questionnaire constitue donc l'information de base du travail statistique :

(a) il résulte, en amont, d'une activité de **production statistique** : observation (sondages ou plans d'expériences, autres modes d'observation), traitement des observations (mise en forme, **contrôle de qualité**, etc) ;

(b) il permet en aval, et après diverses transformations, d'effectuer des descriptions ou des analyses, de formuler des théories, de conseiller, etc.

(viii) La notion de questionnaire se relie ainsi directement aux divers concepts d'**information** et au problème de la **pérennité de l'information** accumulée : archivage, mobilisation, etc.

Le **support matériel du questionnaire** peut être phonique (eg il est d'abord oral lors d'une enquête « sur le trottoir ») ou davantage « matériel ». Dans ce dernier cas, il était souvent, traditionnellement, de type cellulosique (papier) ; il a ensuite évolué

vers des supports magnétiques (bandes magnétiques, disques), optiques (CD ROM, etc) ou à circuits intégrés (cartes, « disques » USB, etc) ainsi que sous forme de fichiers d'échange en réseaux (messageries, questionnaires « en ligne »), etc. Le support télématique (eg enquêtes par téléphone, FAX, etc) est aussi couramment employé.

(ix) La **saisie** d'un questionnaire est une activité qui consiste à opérer un **changement de support** (eg papier vers disque ou cassette) de ce questionnaire.

Elle vise, par principe, à en faciliter l'exploitation, notamment en permettant l'accès à la puissance de calcul des ordinateurs (« grosses bases de données »).

Diverses techniques, notamment celles basées sur la **reconnaissance des formes**, permettent diverses « conversions » entre types de supports : phonique vers phonique (traductions linguistique), phonique vers texte (reconnaissance vocale), image vers texte (reconnaissance de caractères), etc.