

ÉCHANTILLON ALÉATOIRE (F, K, L, M, N)

(16 / 01 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) Un **échantillon aléatoire** (cf **échantillon**, **plan de sondage**) est un **échantillon** A tiré selon un **schéma probabiliste** Π au sein d'un **ensemble** $\Omega = \{\omega_1, \dots, \omega_M\}$, généralement appelé **population**, composé d'un nombre $M \geq 1$ d'**unités statistiques** ω_m (ie $\text{Card } \Omega = M$). Chacune de ces unités « porte » diverses « grandeurs », ou **variables**, (en nombre G) notées synthétiquement $\eta : \Omega \mapsto \prod_{g=1}^G \mathcal{Y}_g$ (où les \mathcal{Y}_g , $g = 1, \dots, G$, désignent des ensembles d'intérêt) : par suite, $Y_m = \eta(\omega_m) \in \mathcal{Y}_g$ ($\forall m \in N_M^*$). Par suite, Y prend ses valeurs dans l'ensemble puissance \mathcal{Y}^M , et $Y = (Y_1, \dots, Y_M)$ est un (G, M) -« tableau statistique ».

Le tirage conduit à « atteindre » un certain nombre N d'unités $\{a_1, \dots, a_N\}$ appelées **unités de sondage**, avec $a_n \in \Omega$ ($\forall n \in N_N^*$). Sur ces unités on observe donc les mêmes grandeurs η que précédemment : on note alors $y_n = \eta(a_n)$ ($\forall n \in N_N^*$) et $y = (y_1, \dots, y_N)$. Par suite, y prend ses valeurs dans l'ensemble puissance \mathcal{Y}^N .

(ii) Le schéma de tirage des unités est tq :

(a) soit les **probabilités** de tirage sont données a priori ;

(b) soit ces probabilités sont à définir (expliciter) ou à maîtriser.

(iii) La grandeur η peut comporter des variables de divers types : **variables quantitatives** (eg discrètes ou continues), **variables qualitatives** ou même **variables morphologiques**.

Dans le cas général, η est donc multivariée (eg **variable mixte**, **variable multidimensionnelle**).

(iv) On doit aussi distinguer entre diverses **structures** des **données** :

(a) structure vectorielle : dans le cas numérique, le **vecteur aléatoire** $\eta : \Omega \mapsto \mathbf{R}^G$ suit une **loi de probabilité** notée P^η , qui est une **loi multidimensionnelle** ;

(b) structure algébrique : l'**échantillon aléatoire** $Y : \Omega \mapsto \mathcal{Y}^M$ (avec, éventuellement, $\mathcal{Y} = \mathbf{R}^G$), dont la **loi** est notée P^Y ;

(c) structure ensembliste : Y peut aussi s'écrire $Y = \{Y_1, \dots, Y_M\}$, **suite** de données Y_m , chacune constituée d'une suite de G observations $Y_m = \{Y_{1m}, \dots, Y_{Gm}\}$.

Dans le cas général, les lois concernées sont des **lois multivariées**.