

SONDAGE EXHAUSTIF (B3, M3)

(23 / 03 / 2020, © Monfort, Dicostat2005, 2005-2020)

Le **sondage exhaustif** est un sondage élémentaire très utilisé, seul ou combiné avec d'autres types de sondage.

(i) Soit Ω une **population** finie et Π un **plan de sondage** de taille fixe N .

On dit que Π est un **plan de sondage exhaustif (équiprobable)** ssi Π est une probabilité uniforme sur l'ensemble $\mathcal{T}_N$ des échantillons sans répétition de taille N extraits de Ω , ie (cf aussi **loi uniforme discrète**) :

$$(1) \quad \Pi(A) = (A_M^N)^{-1}, \quad \forall A \in \mathcal{T}_N,$$

où A est un **échantillon sans remise** de taille N (ie $\text{card } A = N$) et où $\text{card } \Omega = M$.

Le sondage défini en (1) par Π est dit **sondage exhaustif (équiprobable)** (cf **tirage exhaustif**).

(ii) Le tirage exhaustif étant un tirage sans remise des unités, l'échantillon observé $y = (y_1, \dots, y_N)$ est un échantillon non indépendant, mais équidistribué selon la **loi de probabilité** (ou « distribution ») P_M , aussi notée P^n , qui est la distribution de la **variable d'intérêt** η définie sur la population Ω (cf **théorie des sondages**).

Le terme équiprobable signifie ici que la n -ième unité a_n de A est tirée au hasard (uniforme) dans l'ensemble $\Omega \setminus \{a_1, \dots, a_{n-1}\}$ (cf **événements équiprobables**). Sa probabilité d'inclusion est donc :

$$(2) \quad \Pi(a_n \in A) = \{M - (N - 1)\}^{-1}, \quad \forall n = 1, \dots, N.$$

(iii) La **loi hypergéométrique** intervient souvent dans l'étude des sondages exhaustifs, laquelle fait aussi usage de la notion de **variable de CORNFELD**.

(iv) Des **estimateurs** usuellement associés à ce type de sondage sont les suivants (on suppose que η prend ses valeurs dans $\mathcal{Y} = \mathbf{R}$) :

(a) $T_N' = \bar{y}_N$ (**moyenne empirique**), qui estime la moyenne théorique $\bar{Y}$;

(b) $T_N'' = M \cdot \bar{y}_N$, (total empirique) qui estime le total théorique $T = M \cdot \bar{Y}$.

On montre que les estimateurs précédents sont des **estimateurs sans biais** :

$$(3) \quad \begin{aligned} E T_N' &= \bar{Y}, \\ E T_N'' &= T. \end{aligned}$$

Leurs **dispersions** sont resp :

$$(4) \quad \begin{aligned} V T_N' &= (M - 1)^{-1} \cdot (M - N) \cdot (\sigma^2 / N), \\ V T_N'' &= M^2 \cdot V T_N', \end{aligned}$$

avec $\sigma^2 = M^{-1} \cdot \sum_{n=1}^M (Y_n - \bar{Y})^2$ (**variance** théorique).

Elles sont resp estimées sans biais par :

$$(5) \quad \begin{aligned} v T_N' &= M^{-1} \cdot (M - N) \cdot (S_N^2 / N), \\ v T_N'' &= M^2 \cdot v T_N', \end{aligned}$$

avec $S_N^2 = (N - 1)^{-1} \sum_{n=1}^N (y_n - \bar{y}_N)^2$ (variance empirique « corrigée », calculée avec $y = (y_1, \dots, y_N)$).

Dans la formule (4), le coefficient :

$$(6) \quad E = (M - N) / (M - 1)$$

est appelé **coefficient d'exhaustivité**, ou **facteur d'exhaustivité**. Comme $E \leq 1$, il en résulte que (à taille N donnée) le sondage exhaustif est plus précis que le **sondage bernoullien**.