MÉTHODE PAR CAPTURE ET LIBÉRATION (B3, M6)
(17 / 06 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) Une **méthode par capture et libération** est une méthode de **sondage** visant l'estimation du cardinal d'un **ensemble** fini d'éléments (**population**) : ces éléments sont, en général, des **unités statistiques** (**unité de sondage**, **unité expérimentale**) que l'on ne peut (ou ne veut) recenser dans leur ensemble (cf **recensement**).

Cette méthode procède en quatre étapes :

(a) **prélèvement** (capture) d'un **échantillon** d'unités ;

(b) **« marquage »** des unités prélevées (ie attribution d'une **« identité »** ou d'un **mode de repérage** : baguage, émetteur d'ondes, etc) ;

(c) **relâche** (libération) des unités marquées ;

(d) enfin, après une certaine **période d'attente**, **nouveau prélèvement** d'un échantillon.

La méthode de base suppose notamment :

(a) que le dernier échantillon prélevé contient des unités des deux types (donc certaines des unités précédemment marquées), ce qui suppose que la taille d'échantillon est suffisante pr à celle de la population ;

(b) que le délai d'attente n'est ni trop court (mélange par brassage insuffisant entre unités hétérogènes), ni trop long dans le cas d'une population biologique ou écologique (variations de population dues à des naissances, à des décès ou à des migrations).

On peut ainsi évaluer eg l'**effectif** total d'une **population** dont le recensement peut poser problème : ainsi (écologie), les animaux aquatiques, terrestres, arboricoles ou aériens se déplacent indépendamment des frontières institutionnelles dans l'**espace** (zones d'étude, frontières étatiques) et le suivi de leur population en est facilité par marquage (bagues, émetteurs, etc).

(ii) Soit Ω (avec Card Ω = M) l'**ensemble** considéré, A le premier échantillon (avec Card A = N_A), qui est marqué, et B le second échantillon (avec Card B = N_B). On note N_B' le nombre d'éléments (unités) marqué(e)s de B, avec $0 \leq N_B' \leq \min (N_A , N_B)$. On suppose que chaque échantillon est un **échantillon aléatoire** et l'on veut estimer M.

(iii) Un **estimateur** de M est donné par :

(1) $$M^* = (N_B' + 1)^{-1} . (N_B + 1) . N_A .$$

Par suite, si A est prélevé de façon exhaustive, ie si A est un **échantillon sans remise** des unités tirées (cf **tirage exhaustif**), on montre que :

(a) si B est prélevé avec remise (**échantillon avec remise**) :

(2) $E\ M^* = \{1 - (1 - (N_A / M))^{(N(B) + 1)}\} . M$ (estimateur biaisé),

où N(B) désigne l'entier N_B . Le biais est faible si $N_A >> 0$;

(b) si B est prélevé sans remise :

(3)
$$E\ M^* = \begin{cases} (N_A+1)^{-1} . N_A . (M+1), & \text{si } N_A + N_B \geq M, \\ (N_A+1)^{-1} . N_A . \{1 - (C_1 / C_2)\} . (M+1), & \text{si } N_A + N_B \leq M-1, \end{cases}$$

où $C_1 = C_{M-N(A)}^{N(B)+1}$, $C_2 = C_{M+1}^{N(B)+1}$, et où, par commodité, N(A) désigne l'entier N_A et N(B) l'entier N_B .

(iv) Une variante de la méthode consiste à fixer a priori le nombre N_B' et à prélever l'échantillon de taille N_B suffisante. Une estimateur de M est alors donné par :

(4) $M^{**} = (N_B')^{-1} . N_B . N_A$.

Si A est prélevé de façon exhaustive (ie sans remise des unités), on montre que :

(a) si B est prélevé avec remise, $E\ M^{**} = M$ (**estimateur sans biais**) ;

(b) si B est prélevé sans remise, $E\ M^{**} = (N_A + 1)^{-1} . N_A . M$ (estimateur biaisé). Ce biais est donc faible si $N_A >> 0$.

(v) Le rapport $L = N_B' / N_A$, qui intervient dans la définition des estimateurs M* ou M**, est appelé **indice de F.C. LINCOLN**.

(vi) Ces méthodes de capture et relâche s'étendent, non seulement à des cardinaux d'unités, mais aussi à des valeurs correspondantes (**variables** mesurées sur les unités).

AInsi, dans le cas d'un **caractère** scalaire, ie d'une **va** $\eta : \Omega \mapsto \mathbf{R}$, le total $T_M = \Sigma_{\omega \in \Omega} \eta(\omega)$ de ce caractère dans la population peut être estimé à l'aide eg la va (ou **statistique**) :

(5) $T_M^* = \{\Sigma_{\omega \in B'} \eta(\omega)\}^{-1} . \{\Sigma_{\omega \in B} \eta(\omega)\} . \Sigma_{\omega \in A} \eta(\omega)$,

où B' est l'ensemble des éléments de B qui sont marqués (avec Card B' = N_B').

(vii) En **théorie des sondages**, il est parfois utile de procéder par sondage (capture) et re-sondage (recapture), eg pour assurer un **chaînage temporel** du **dispositif d'observation** (cas d'échantillons partiellement renouvelés) (cf **renouvellement**, **sondage renouvelé**, **système d'observation**).