

SONDAGE PAR QUOTAS (M5)

(16 / 03 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) Soit $\Omega = \{\omega_1, \dots, \omega_M\}$ un ensemble fini (**population**), $(\mathcal{X}, \mathcal{B})$ et $(\mathcal{Y}, \mathcal{C})$ deux **espaces d'observation** et $A = \{a_1, \dots, a_N\}$ un **échantillon aléatoire** extrait de Ω .

On associe aux **unités de sondage** (éléments ou individus) $\omega \in \Omega$ deux **variables** (ou « **caractères** ») ξ et η selon $(\xi, \eta) : \Omega \mapsto (\mathcal{X}, \mathcal{Y})$. Autrement dit, on suppose que l'on peut observer sur chaque élément ω les valeurs $\xi(\omega)$ et $\eta(\omega)$. Les caractères peuvent éventuellement être « multiples » ou vectoriels, eg $\xi = (\xi_1, \dots, \xi_K)$ et $\eta = (\eta_1, \dots, \eta_G)$.

On note alors $(X_m, Y_m) = (\xi(\omega_m), \eta(\omega_m))$ ($m = 1, \dots, M$) les « valeurs » observées sur les unités de Ω , $(x_n, y_n) = (\xi(a_n), \eta(a_n))$ ($n = 1, \dots, N$) les « valeurs » observées sur les unités de A , et $((x_1, y_1), \dots, (x_N, y_N))$ l'**échantillon aléatoire** résultant de ces dernières. On note aussi resp :

$$(1) \quad P_M \text{ ou } P^{(\xi, \eta)} = M^{-1} \sum_{m=1}^M \delta_{(X_m, Y_m)}$$

la **loi de probabilité** (ou « distribution ») du couple (ξ, η) dans la population Ω , et :

$$(2) \quad P_N = N^{-1} \sum_{n=1}^N \delta_{(x_n, y_n)}$$

la **loi empirique** définie par A , où δ_a désigne la **fonction de DIRAC** placée en a , X_m , Y_m , x_n et y_n désignant resp (par commodité) X_m , Y_m , x_n et y_n .

(ii) On appelle **sondage par quotas** un sondage fondé sur l'idée suivante : si la loi empirique P_N définie par A est « voisine » de la loi théorique P_M définie sur Ω pour la variable (marginale) ξ , alors elle est encore voisine de cette même loi pour la variable (marginale) η . La variable ξ est appelée **variable de contrôle**, ou **caractère de contrôle**.

En pratique, on cherche à stratifier la population Ω à l'aide de la **variable de contrôle** ξ (variable de stratification) (cf **sondage stratifié**). Si $\Pi_\Omega = \{\Omega_1, \dots, \Omega_H\}$ représente la stratification (**partition** de Ω) adoptée, les **rapports** M_h / M sont appelés **quotas**, ou **proportions**, d'effectifs par strate, avec $M_h = \text{card } \Omega_h$. Si N est la taille de l'échantillon A , les quotas (absolus) d'**unités de sondage** à observer sont donc :

$$(3) \quad N_h = (M_h / M) \cdot N, \quad \forall h = 1, \dots, H.$$

Une **base de sondage** préalable est donc inutile pour procéder à un sondage par quotas.

(iii) Dans la **méthode des quotas**, le **plan de sondage** Π n'est que partiellement déterminé : Π doit « charger » les échantillons A représentatifs de Ω pour le seul caractère ξ . On ne peut donc pas calculer les **caractéristiques** des **statistiques**

calculées à l'aide du caractère η . Ainsi, dans le cas où $\mathcal{Y} = \mathbf{R}^G$, on ne peut pas évaluer $E_{\Pi} \bar{y}_N$, où $\bar{y}_N$ est la **moyenne empirique** (vectorielle) calculée à l'aide de l'**échantillon** observé $y = (y_1, \dots, y_N)$.

(iv) Les quotas peuvent être des **quotas « croisés »**, auquel cas la stratification Π_{Ω} précédente s'effectue conjointement pour différentes valeurs (ou modalités) des variables $\xi_1, \dots, \xi_K$ (coordonnées de ξ), ou des **quotas « marginaux »**, auquel cas la stratification s'effectue séparément pour différentes valeurs (ou modalités) de chacune des variables $\xi_1, \dots, \xi_K$.

(v) La validité de la **méthode des quotas** suppose notamment :

- (a) que la liste ξ des variables de contrôle soit bien spécifiée ;
- (b) que la distribution des variables de contrôle ξ soit bien connue dans Ω ;
- (c) que ces variables soient corrélées avec les variables d'intérêt étudiées η ;
- (d) que toutes les variables soient mesurées sans erreurs ;

(e) que le plan de sondage ne charge que des échantillons représentatifs de la population Ω pour les caractères de contrôle ξ .

(vi) La **méthode des quotas** peut relever d'une interprétation bayésienne (cf **école bayésienne**, **théorie bayésienne**), ce qui lui fournit un fondement probabiliste (indirect).