APPARIEMENT (B2, F)
(23 / 09 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) On appelle **appariement**, ou **rencontre** (en anglais : **match**), entre deux (ou plusieurs) **unités statistiques** la situation selon laquelle ces unités possèdent (au sens strict ou, parfois, au sens large) les mêmes propriétés ou les mêmes **caractéristiques** (cf **problème des rencontres**).

En général, les propriétés sont définies à l'avance et les unités sont extraites (de façon aléatoire ou non) de **populations** distinctes (ie globalement différentes entre elles).

Ainsi, soit $A = \{a_1 ,..., a_N\}$ et $B = \{b_1 ,..., b_M\}$ deux **échantillons** ainsi extraits.

On dit que l'unité $a_n \in A$ est appariée avec l'unité $b_m \in B$ ssi :

(a) soit $B \neq A$ et $b_m = a_n$ (**appariement sur identités** des unités) ;

(b) soit $B = A$ et $m = n$ (**appariement sur identifiants** des unités) ou même $b_m = a_n$ (**appariement sur identités** des unités).

(ii) Le plus souvent, on mesure, sur chaque unité, une ou plusieurs **variables** ζ (**variables numériques** ou **variables qualitatives**), à valeurs dans un ensemble $\mathcal{Z}$.

On dit alors que deux (ou plusieurs) unités sont des **unités appariées**, ou sont des **unités accouplées**, ou qu'il y a **rencontre** entre elles, ou encore qu'elles se **rencontrent**, ssi elles sont égales (ou, parfois, seulement « comparables ») d'un point de vue défini par la considération de certaines des variables précédentes.

Ainsi, soit $A = \{a_1 ,..., a_N\}$ et $B = \{b_1 ,..., b_M\}$ deux **échantillons** extraits de populations données, $Z^a{}_n = \zeta (a_n)$ la valeur observée pour ζ sur l'unité a_n et $Z^b{}_m = \zeta (b_m)$ la valeur observée pour ζ sur l'unité b_m .

On dit que l'unité $a_n \in A$ est appariée avec l'unité $b_m \in B$ ssi :

(a) de façon stricte : $m = n$ et $Z^b{}_m = Z^a{}_n$ (**appariement sur identifiants et caractéristiques** des unités) ;

(b) de façon plus large : $Z^b{}_m = Z^a{}_n$ (**appariement sur caractéristiques** des unités) ;

(c) de façon encore moins restrictive : $d (Z^b{}_m , Z^a{}_n) < \varepsilon$ (distance « négligeable » entre valeurs), où d est une **distance** sur $\mathcal{Z}$ et $\varepsilon > 0$ un nombre tq $\varepsilon << +\infty$.

(iii) Dans le cadre du **problème à plusieurs échantillons** (échantillons d'**observations**), on note ces derniers resp $X_i = (X_{i,1} ,..., X_{i,N(i)})$, $\forall$ i $\in N_k$*, où $X_{i,n(i)} = \xi_i (a_{n(i)})$ est une observation effectuée sur l'unité $a_{n(i)}$ extraite de la i-ième population ($\forall$ i $\in N_k$*).

On dit alors eg que les unités $a_{n(1)}$,..., $a_{n(k)}$ sont appariées ssi :

(1) $X_{1,n(1)} = \ldots = X_{k,n(k)}$.

De tels problèmes se posent souvent lorsque les ξ_i sont des **variables qualitatives**. Ils conduisent à la définition de **lp** variées, alors dites **lois d'appariement**.