TEST DE HAGA (F6, I2, C9)
(09 / 10 / 2020, © Monfort, Dicostat2005, 2005-2020)

Le **test de HAGA** est un **test rapide** d'**homogénéité** entre deux **loi de probabilité** (ou deux **populations**) réalisé à partir d'un **échantillon** issu de chacune d'elles (cf **problème à plusieurs échantillons**). Ce test est basé sur une **statistique des extrêmes** (cf aussi **rapport des extrêmes**, **valeur extrême**).

(i) Soit $X^i = (X_{i1}, ..., X_{iN(i)})$ $(i = 1, 2)$ deux échantillons et $N = N_1 + N_2$ (où $N(i)$ désigne par commodité N_i). On note X l'échantillon « regroupé » et R sa **statistique de rang**.

On appelle alors **test de T. HAGA** le test basé sur la **statistique de test** suivante, appelée **statistique de T. HAGA** :

(1) $\quad H_{N(1)N(2)} = A + B - (A' + B'),$

dans laquelle :

$$A = \{\min_{n=1}^{N(1)} R_n\} - 1, \quad B = N - \{\max_{n=N(1)+1}^{N} R_n\},$$

(2)

$$A' = \{\min_{n=N(1)+1}^{N} R_n\} - 1, \quad B' = N - \{\max_{n=1}^{N(1)} R_n\}.$$

L'interprétation est la suivante :

(a) A est le nombre de coordonnées de l'échantillon X^2 inférieures au plus petit des X_{1n} $(n = 1, ..., N_1)$;

(b) B est le nombre de coordonnées de l'échantillon X^1 supérieures au plus grand des X_{2n} $(n = 1, ..., N_2)$;

(c) A' est le nombre de coordonnées de X^1 inférieures au plus petit des X_{2n} $(n = 1, ..., N_2)$;

(d) B' est le nombre de coordonnées de X^2 supérieures au plus grand des X_{1n} $(n = 1, ..., N_1)$.

(ii) La **loi de probabilité** de $H_{N(1)N(2)}$ est tabulée. Elle permet de tester l'**hypothèse** d'homogénéité (stricte), ie l'**hypothèse de base** :

(3) $\quad H_0 : P^{\xi(2)} = P^{\xi(1)},$

entre les lois générant les échantillons, contre des **alternatives de position**. Le test semble adapté aux lois « tronquées » (cf **troncature**) ou à des lois sans **queue** (eg **loi uniforme**).