STATISTIQUE DE WELCH (I2, I6)
(29 / 08 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) On considère le problème à deux échantillons (cf **problème à plusieurs échantillons**) dans lequel on teste l'**hypothèse de base** H_0 : $\mu_2 = \mu_1$ (égalité des **moyennes**) contre l'**hypothèse alternative** bilatérale H_a : $\mu_2 \neq \mu_1$.

Si les variances théoriques à comparer, σ_1^2 et σ_2^2, sont différentes, la **statistique de L. WELCH** est définie selon :

(1) $W_{N(1)N(2)} = D / S^2$,

(où l'on désigne N_i par N(i)), avec :

(a) $D = \bar{X}_1 - \bar{X}_2$ (différence des moyennes), où $\bar{X}_i = N^{-1} \Sigma_{n(i)=1}^{N(i)} X_{n(i)}$ est la **moyenne empirique** associée à l'**échantillon aléatoire** $X^i = (X_{i,1}..., X_{i,N(i)})$ (i = 1, 2) ;

(b) $S^2 = N_1^{-1} s_1^2 + N_2^{-1} s_2^2$ est la **variance empirique** d'ensemble et $s_i^2 = (N_i - 1)^{-1} \Sigma_{n(i)=1}^{N(i)} (X_{i,n(i)} - \bar{X}_i)^2$ la **variance** (« corrigée » de son **biais**) de l'échantillon i.

(ii) Sous des hypothèses classiques, le test (asymptotique) de H_0 se base sur la **propriété asymptotique** suivante :

(2) $\mathscr{L}(W_{N(1)N(2)}) \to_{\min(N(1),N(2)) \to \infty} \mathscr{S}_{[\nu]}$ (**loi de STUDENT** à [ν] degrés de liberté),

avec :

(3) $\nu = S^4 / M_4$,

expression dans laquelle :

(4)
$$S^4 = (S^2)^2,$$
$$M_4 = \{N_1^2 (N_1 - 1)\}^{-1} s_1^4 + \{N_2^2 (N_2 - 1)\}^{-1} s_2^4,$$

où [u] dénote la **partie entière** de u.

Le test de H_0 fondé sur $W_{N(1)N(2)}$ est plus « robuste » que le test (fondé sur la statistique) de STUDENT (cf **robustesse**). Lorsque $N_2 = N_1$, il se réduit cependant à ce dernier (cf aussi **problème de BEHRENS-FISHER**).