TEST DE MALLOWS (J1, J6)
(25 / 11 / 2019, © Monfort, Dicostat2005, 2005-2019)

Le **test de MALLOWS** est utilisé en association avec une **régression pas à pas** (choix de **variable exogène**) ou avec la **sélection entre modèles**.

(i) Soit :

(1) $y = b_1 e_N + X b + u$, avec $E u = 0$, $V u = \sigma^2 I_N$,

un **modèle de régression linéaire** multiple avec terme constant ($x_1 = e_N$) et K - 1 variables exogènes « vraies » $\xi_2 ,..., \xi_k$ observées selon la **matrice** $X = [x_2 ,..., x_K] \in M_{N,K-1} (\mathbf{R})$ (cf **espace d'observation**). On suppose que la matrice « augmentée » $[e_N , X]$ est de plein rang K.

Soit $\mathscr{L} \subset \{2 ,..., K\}$. On admet que le « vrai modèle » engendrant les observations est l'un des modèles suivants (indexés par $\mathscr{L}$) :

(2) $y = b_1 e_N + X_{\mathscr{L}} b_{\mathscr{L}} + u_{\mathscr{L}}$,

et l'on note $(b_1\hat{}, b_{\mathscr{L}}\hat{})$ l'**estimateur des moindres carrés ordinaires** de $(b_1 , b_{\mathscr{L}})$ défini à partir du modèle (2).

Si $\mathscr{L}_0 \subset \{2,..., K\}$ est donné (suite d'exogènes privilégiée), le test de C.L. MALLOWS consiste à tester l'**hypothèse de base** :

(3) $H_0 : \mathscr{L} = \mathscr{L}_0$,

contre l'**hypothèse alternative** :

(4) $H_1 : \mathscr{L} \neq \mathscr{L}_0$.

On appelle **statistique de C.L. MALLOWS** la statistique $C_{N,\mathscr{L}}$ définie par le **rapport des variances** suivant :

(3) $C_{N,\mathscr{L}} = \{(\sigma^2)\hat{}\}^{-1} . \Sigma_{n=1}^{N} (y_n - b_1\hat{} e_N - X_{\mathscr{L}} b_{\mathscr{L}}\hat{})^2 - (N - L) + L$,

dans lequel :

(a) le numérateur est (sous H_0) un **estimateur sans biais** de $(N - L) \sigma^2$;

(b) le dénominateur est l'estimateur habituel des moindres carrés ordinaires de σ^2 et $L = \text{Card } \mathscr{L}$.

Sous H_0 , $C_{N,\mathscr{L}}$ est « proche » de l'entier L ; sinon, $C_{N,\mathscr{L}} >> L$.

On appelle alors **test de C.L. MALLOWS** le test de l'hypothèse H_0 précédente fondé sur la statistique $C_{N,\mathscr{L}}$.

(ii) Sous des hypothèses classiques, et avec un seuil critique $\alpha \in]0, 1[$, on établit que la **région critique** du test est de la forme :

(4) $\quad w = [C_{N,\mathscr{L}} \geq (2\,L - K - 1)\,.\,K\,.\,q_{1-\alpha}]$,

dans laquelle $q_{1-\alpha}$ est le **quantile** d'ordre $1 - \alpha$ de la **loi de FISHER-SNEDECOR** $\mathscr{F}_{K,N-K-1}$ à K et N-K-1 **degrés de liberté**.