DÉTECTION (G9, H4, I5, N12)
(06 / 11 / 2019, © Monfort, Dicostat2005, 2005-2019)

Le terme de **détection** possède deux sens usuels.

(i) En **théorie du signal**, il s'agit de **repérer**, ou d'estimer, un **signal** (pur) au sein d'un **message**, sachant que ce dernier (signal perturbé) est la résultante du signal pur et d'un **bruit** perturbateur. Ce problème peut être rapproché du problème de **reconnaissance des formes**.

(ii) Dans un **modèle statistique**, la **détection d'une aberration**, ou « point aberrant », vise à repérer et à éliminer une (ou plusieurs) observation(s) qui ne suivent pas le **schéma** théorique, celui retenu ou celui supposé.

Ainsi, dans ce deuxième sens, et dans le cadre du **modèle de régression linéaire** multiple :

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

estimé par la **méthode des moindres carrés ordinaires**, la détection peut consister, eg :

(a) à repérer des coordonnées extrêmes du vecteur résiduel $u^\wedge = y - y^\wedge$ (cf **élément extrémal**, **statistique de rang**, **valeur extrême**) ;

(b) à étudier la diagonale de la **matrice** $H = X\,(X'\,X)^{-1}\,X'$, qui vérifie $y^\wedge = H\,y$ (ie l'élément $h_{\alpha\beta}$ est la dérivée partielle de $y_\alpha{}^\wedge$ pr à y_β , donc mesure l'influence de y_β sur $y_\alpha{}^\wedge$) (cf aussi **courbe d'influence**) ;

(c) à calculer les **résidus** « **studentisés** » (ie modifiés par la **transformation de STUDENT**) :

(2) $t_n = u_n{}^\wedge / \{\sigma^\wedge\,(1 - h_{nn})^{1/2}\}$,

où $(\sigma^2)^\wedge = ||u^\wedge||^2 / (N - K)$, $V\,u^\wedge = \sigma^2\,(I_N - H)$, $\sigma^\wedge = \{(\sigma^2)^\wedge\}^{1/2}$, et à admettre l'**hypothèse statistique** $t_n \sim \mathcal{S}_{N-K}$ (**loi de STUDENT**) pour éliminer, par test, les coordonnées extrêmes de $t = (t_1 ,..., t_N)$;

(d) à appliquer divers tests : eg le **test de COOK**.