TEST DE NORMALITÉ (C7, I2)
(11 / 10 / 2019, © Monfort, Dicostat2005, 2005-2019)

(i) La **loi normale** joue un rôle important en **calcul des probabilités** et en **Statistique** :

(a) comme loi limite (cf **loi asymptotique**), en raison du **théorème de la limite centrale** ;

(b) du fait de ses commodités (relatives) de calcul ;

(c) parce qu'elle est étroitement associée à la **méthode des moindres carrés** et à ses variantes.

Elle intervient souvent dans les hypothèses « stochastiques » qui sont à la base de certaines **procédures statistiques**. Il est donc utile de pouvoir vérifier si ces hypothèses de **normalité** sont vérifiées, ou simplement « plausibles » (ie approximativement vérifiées) : cette vérification est l'objet d'un **test de normalité**.

(ii) Il existe deux catégories de test de normalité :

(a) les **tests « empiriques »** et, notamment graphiques : **droite de HENRI**, **droite de PEARSON**, etc. Ces tests sont simples et rapides à utiliser (cf **test rapide**) mais peu rigoureux ;

(b) les **tests « statistiques »** proprement dits sont très variés :

$(b)_1$ **tests d'adéquation** à la loi normale (cf **adéquation**) : **test du chi-deux**, **test de KOLMOGOROV-SMIRNOV**, **test de SHAPIRO-WILK**, etc ;

$(b)_2$ tests spécifiques, utilisant généralement une propriété caractéristique (ou typique) de la loi normale.

Un test d'adéquation est un test très général, et parfois qualifié de **« test omnibus »**, mais il est souvent moins discriminants entre les lois.

(ii) En pratique, on effectue souvent, préalablement au test, une **transformation des données** (ou **observations**) ou une élimination de valeurs aberrantes (cf **aberration**). Ce type de traitements permet, dans certaines **situations statistiques**, de rendre positives les conclusions du test de normalité.

(iii) D'un point de vue formel, soit $X = (X_1 ,..., X_N)$ un **échantillon aléatoire** issu d'un **vecteur aléatoire** $\xi : \Omega \mapsto \mathbf{R}^K$, **variable parente** dont la loi est notée P^ξ.

Un **test de normalité** revient à vérifier (valider) l'**hypothèse de base** :

(1) $H_0 : P^\xi = \mathcal{N}_K (\mu, \Sigma)$ (**loi normale multidimensionnelle**)

au vu de la **loi empirique** P_N ou de la **fr empirique** F_N .

Lorsque l'**hypothèse alternative** n'est pas toujours connue ou spécifiée (alternative « générale »), on peut mettre en oeuvre des tests « omnibus » (eg le **test du chi-deux**). Dans d'autres situations, l'alternative spécifie une loi particulière (eg la loi de **loi de CAUCHY** ou la **loi de STUDENT**) ou, plus largement, une **famille** de lois concurrentes.

Souvent, l'hypothèse H_0 est remplacée seulement par une **hypothèse restreinte**, tq :

(2) $H_0' : \; c\,(P^{\xi}) \; = \; \gamma_0$,

dans laquelle c est une application caractéristique et γ_0 une **caractéristique légale** (paramètre d'intérêt) associé à la loi normale.

Ainsi, dans la cas scalaire (K = 1), on peut s'intéresser au **coefficient d'aplatissement** $\gamma_0 = \gamma_2 = \beta_2 - 3 = 0$ de la loi $\mathcal{N}_1\,(\mu\,,\,\sigma^2)$, et tester l'hypothèse :

(3) $H_0' : \; \beta_2 \; = \; 3$ (ie $\gamma_0 = 0$).