TEST DE SHAPIRO - WILK (C7, F6, I2)
(21 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

Le **test de SHAPIRO - WILK** est un **test de normalité** fondé sur le **rapport** de deux **estimateurs** de la **variance** qui sont des **estimateurs sans biais** sous l'hypothèse nulle, et biaisé sous l'alternative.

(i) Soit $(\mathcal{X}, \mathcal{B}, P^{\xi})^{\otimes N}$ un **modèle d'échantillonnage** (**modèle image**) dans lequel P^{ξ} est la **loi de probabilité** d'une **variable parente** $\xi : \Omega \mapsto \mathbf{R}$ dont les **copies** génèrent un **échantillon indépendant équidistribué** $X = (X_1, ..., X_N)$.

On note $X^{(.)} = (X^{(1)}, ..., X^{(N)})$ l'échantillon ordonné associé à X (cf **statistique d'ordre**) et $\mathcal{N} = \{\mathcal{N}_1 (\mu, \sigma^2) : \forall (\mu, \sigma^2) \in \mathbf{R} \times \mathbf{R}_+^*\}$ la **famille** des **lois gaussiennes** (scalaires).

Pour tester la **normalité** de X, ie pour tester l'**hypothèse de base** :

(0) $H_0 : P^{\xi} \in \mathcal{N}$,

on peut procéder comme suit :

(a) tirage d'un **vecteur aléatoire** gaussien $U = (U_1, ..., U_N) \sim \mathcal{N}_N (0, I_N)$, dont on déduit l'échantillon ordonné $U^{(.)} = (U^{(1)}, ..., U^{(N)})$ ainsi que le vecteur des **scores normaux** $a_N (n) = E\, U^{(.)} (n = 1, ..., N)$. La **matrice de dispersion** $\Sigma^{(.)}$ de $U^{(.)}$ n'est pas diagonale (cf **matrice diagonale**) car les **va** $U^{(n)}$ ne sont pas indépendantes ;

(b) **estimation**, par la **méthode des moindres carrés généralisés**, du **modèle de régression** simple :

(1) $X^{(.)} = \mu \, . \, e_N + \sigma \, . \, U^{(.)} + u$,

dans laquelle $X^{(.)}$ et $U^{(.)}$ sont disposés en vecteurs colonnes. En effet, comme les $X^{(n)}$ ne sont pas indépendants, V u n'est pas diagonale et l'**estimateur des mcg** de (μ, σ^2) vaut alors :

(2) $\hat{\mu}_g = \bar{X}_N = N^{-1} \sum_{n=1}^{N} X_n = e_N' X / e_N' e_N$ (**moyenne empirique**),

$(\hat{\sigma^2})_g = \{(\mu^{(.)})' (\Sigma^{(.)})^{-1} \mu^{(.)}\}^{-1} (\mu^{(.)})' (\mu^{(.)})' (\Sigma^{(.)})^{-1} X^{(.)}$,

où $\mu^{(.)} = E\, U^{(.)}$ et $\Sigma^{(.)} = V\, U^{(.)}$;

(c) calcul de la **statistique de S.S. SHAPIRO - M.B. WILK** (cf **statistique de test**) :

(3) $W_N^2 = T_N^2 / S_N^2$,

dans laquelle :

$$S_N^2 = X' P X,$$

(4)

$$T_N^2 = \{\sum_{n=1}^{[N/2]} a_N(n) (X^{(N-n+1)} - X^{(n)})\}^2.$$

Le numérateur s'interprète comme l'estimateur du **coefficient de régression** de $X^{(.)}$ sur $U^{(.)}$, et le dénominateur comme la somme des carrés des écarts des X_n par rapport à leur **moyenne empirique** (P étant la **matrice de centrage** par rapport à cette dernière).

(ii) Le **test de S.S. SHAPIRO - M.B. WILK** est fondé sur W_N^2 et admet des **régions critiques** de la forme :

(5) $w = \{W_N^2 < q_{1-\alpha}\}$,

où $\alpha \in [0, 1[$ est le **risque de première espèce** retenu, et $q_{1-\alpha}$ le **quantile** d'ordre $1 - \alpha$ de la loi de W_N^2.

A distance finie (ie $N << +\infty$), cette loi est tabulée en fonction de α et de N. En effet, on observe, par **simulation**, que $E\, W_N^2$ est plus grande sous l'hypothèse de **normalité** H_0 que sous une alternative non gaussienne ; de même, $V\, W_N^2$ est plus petite sous l'hypothèse H_0 que sous une alternative non gaussienne.

(iii) Sous l'hypothèse H_0 , on montre que :

(a) le **couple aléatoire** (S_N^2, T_N^2) est indépendant ;

(b) $W_N^2 < 1$;

(c) en posant $W_N^2 = w_N(X)$:

1) $w_N(X - \alpha) = w_N(X)$ (invariance par changement d'origine $\alpha \in \mathbf{R}^N$) ;

2) $w_N(X / \beta) = w_N(X)$ (invariance par changement d'échelle $\beta > 0$) ;

3) la loi $\mathcal{L}(W_N^2)$ de W_N^2 ne dépend pas de la **vraie valeur** $(\mu^*, (\sigma^2)^*)$ du paramètre (μ, σ^2).

La **statistique de test** W_N^2 est souvent notée W_N (ou simplement W) et ce test est aussi appelé **W-test (de normalité)**.

(iv) La **statistique de test** de SHAPIRO-WILK se calcule aussi selon :

(6) $W_N = c_N . \{S_N^2 / T_N^2\}$,

avec :

(4)' $S_N^2 = (\sigma^2)_g\hat{}$,

$T_N^2 = \Sigma_{n=1}^N (X_n - \bar{X}_N\}^2 = ||X - \bar{X}_N \, . \, e_N||^2$,

où $c_N > 0$ désigne une **constante** de proportionnalité (cf (2)).