REVUE DE STATISTIQUE APPLIQUÉE

E. MORICE

# Tests de normalité d'une distribution observée

*Revue de statistique appliquée*, tome 20, nº 2 (1972), p. 5-35
<http://www.numdam.org/item?id=RSA_1972__20_2_5_0>

© Société française de statistique, 1972, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (http://www.sfds.asso.fr/publicat/rsa.htm) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
*Article numérisé dans le cadre du programme*
*Numérisation de documents anciens mathématiques*
http://www.numdam.org/

# TESTS DE NORMALITÉ D'UNE DISTRIBUTION OBSERVÉE

E. MORICE

*L'auteur passe en revue, dans l'article qui suit, les principaux tests de normalité qui sont proposés dans la littérature statistique, en essayant de rendre clair leur mode d'emploi, d'une part, et ce qu'on sait (peu de chose en vérité) de leurs propriétés en termes de puissance. Peut-on faire plus, et donner des recommandations à l'usager, quant au choix du test qu'il aura intérêt à employer, dans telle situation pratique ?*

*Il ne nous semble pas que ce soit aisé. Reprenons ici l'essentiel des leçons qu'apporte en ce domaine la pratique de la statistique.*

*Lorsqu'on a affaire à de très grands échantillons, on arrivera pratiquement toujours à mettre en évidence une discordance avec la loi normale – et cela quel que soit le test utilisé. D'un autre côté, lorsqu'on a affaire à des échantillons assez faibles, l'accord avec la loi normale paraîtra presque toujours admissible. C'est dans le domaine intermédiaire, quant à l'effectif des échantillons – disons, le plus souvent, entre une à deux dizaines et une à deux centaines d'observations – qu'on peut rechercher quelque assurance dans une méthode de test apportant des garanties raisonnables à l'encontre des deux espèces classiques d'erreurs. On sait combien il est aisé de contrôler l'erreur de première espèce, on sait aussi qu'il en va tout différemment pour l'erreur de seconde espèce, qui n'est même pas définie tant qu'on n'a pas précisé une hypothèse alternative.*

*D'ailleurs, dans la pratique courante, les raisons qui font choisir tel test sont très terre à terre, et tiennent souvent à la présentation des données, à la commodité, ou aux calculs déjà effectués par ailleurs, et qu'on peut utiliser à bon compte. De telles motivations en valent bien d'autres – pourvu qu'on prenne garde toutefois à ne pas utiliser un test* $\beta_1$ *de Pearson si les seules alternatives raisonnables sont des distributions symétriques, ou un test utilisant l'étendue lorsque les perturbations que l'on peut craindre, par rapport à la loi normale, affectent de préférence les valeurs centrales des observations !*

*Cela dit, on peut s'inspirer du peu que l'on sait sur la puissance des tests courants, et c'est à cette fin que nous présentons le contenu de cet article. On peut aussi consoler l'usager en lui rappelant qu'il est toujours possible, en théorie, d'établir "sur mesure" le test optimal à l'encontre de telle hypothèse alternative complètement spécifiée. Le lemme de Neyman et Pearson nous apprend, d'une manière incontestable, à faire cette étude. On ne sera pas toujours conduit à des calculs simples, mais les ordinateurs sont là. Mais surtout, en contre partie, une alternative bien précise peut rarement traduire correctement ce que sait et ce que craint le praticien, et pour cette raison une telle étude ne pourra que très exceptionnellement être envisagée par le statisticien.*

Les tests classiques de comparaison de moyennes ou de variances à une valeur spécifiée ou de comparaison de plusieurs moyennes ou variances impliquent généralement l'hypothèse préalable d'échantillons tirés de populations normales. Si, dans certains cas, par exemple étude de moyennes d'échantillons d'effectifs non très petits, on peut ne pas tenir compte de cette hypothèse, on sera souvent amené en présence d'une population inconnue à se demander si l'hypothèse de normalité peut être considérée comme valable, d'où l'utilité de tests de normalité ou plutôt, comme le disait K. Pearson, de "tests de non normalité", permettant éventuellement d'être en mesure de juger que la population échantillonnée est probablement une population normalement distribuée.

De nombreux tests ont été proposés, pouvant se répartir en deux catégories.

## I - TESTS BASES SUR LA COMPARAISON DE LA DISTRIBUTION OBSERVEE A UNE DISTRIBUTION NORMALE $\mathcal{N}(m, \sigma)$ DONT LES PARAMETRES ONT ETE ESTIMES A PARTIR DES OBSERVATIONS GENERALEMENT PAR LA METHODE DES MOMENTS :

$$m_r = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}, \quad x = \frac{1}{n} \sum_{i=1}^{n} x_i$$

### 1 - Test d'adéquation de $\chi^2$

Son utilisation suppose pratiquement que l'on dispose d'un échantillon suffisamment important pour que l'on puisse répartir les n observations par classes.

Si $X_i$ est le centre de la classe $(x_{i-1}, x_i)$, $X_i = \frac{1}{2}(x_{i-1} + x_i)$ et $n_i$ l'effectif de cette classe, $\sum_{i=1}^{k} n_i = n$, k étant le nombre des classes, les paramètres de la loi normale ajustée seront estimés par

$$m = \frac{\sum_{i=1}^{k} n_i X_i}{n}$$

$$\sigma^2 = \frac{\sum_{1}^{k} (n_i X_i^2) - \frac{(\sum_{1}^{k} n_i X_i)^2}{n}}{n - 1}$$

Le groupement de toutes les observations d'une classe sur sa valeur centrale entraine une erreur systématique qui pourra être approximativement corrigée par la correction de Sheppard, qui donnerait :

$$\sigma'^2 = \sigma^2 - \frac{a^2}{12}$$

correction négligeable dans le cas de classes égales de faible amplitude a (pratiquement dès que le nombre des classes dépasse 10).

L'effectif théorique $n'_i$ de la classe $(x_{i-1}, x_i)$ peut alors être calculé à partir de la fonction de répartition $F(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$ :

$$n'_i = n\,[F(u_i) - F(u_{i-1})]$$

avec $u_i = \frac{x_i - m}{\sigma}$

On calcule ensuite

$$\chi^2 = \sum_1^k \frac{(n_i - n'_i)^2}{n'_i}$$

distribué approximativement comme la variable $\chi^2$ avec $\nu = k - 3$ degrés de liberté pour k classes utilisées dans la comparaison.

Le test conduit au rejet de l'hypothèse de normalité au risque $\alpha$ si la valeur trouvée est supérieure à la valeur $\chi^2_{1-\alpha}$ donnée par la table de $\chi^2$ ; $\alpha = \Pr\,[\chi^2 > \chi^2_{1-\alpha}]$.

On admet généralement que ce test est utilisable à condition qu'aucun effectif théorique n' ne soit très petit (au moins 4 ou 5 unités)(1), ce qui peut exiger pour son emploi de réunir certaines classes aux extrémités de la distribution, diminuant alors le nombre de degrés de liberté ; son résultat peut dépendre du choix des limites de classes, laissé à l'arbitraire de l'utilisateur.

2 - Test de Geary [8] [9]

Il est basé sur la distribution théorique du rapport a de l'écart moyen à l'écart-type, rapport dont la valeur théorique dans le cas d'une loi normale est $\left(\frac{2}{\pi}\right)^{1/2} = 0,7979$.

La loi de distribution de la variable

$$a = \frac{\sum_1 |x_i - \bar{x}|}{[n \sum_1 (x_i - \bar{x})^2]^{1/2}}$$

a été étudiée par Geary qui a calculé certains fractiles de cette loi. Le test conduira au rejet de l'hypothèse de normalité au risque $\alpha$ si la valeur calculée de a est extérieure à l'intervalle $a_{\alpha/2}$ $a_{1-\alpha/2}$ donné par les tables.

---

(1) Cochran [4], estime que cette condition est trop sévère, particulièrement dans le cas d'une distribution unimodale et donne comme condition : $n'_i$ soit au moins de 1 ou 2 unités.

On trouvera en annexe un extrait de ces tables. Pour l'emploi de ce test on notera que le calcul de l'écart moyen n'exige pas le calcul individuel de toutes les valeurs $|x_i - \bar{x}|$.

Plus simplement, après avoir classé les observations individuelles par rapport à la moyenne on a :

$$\Sigma\ |x_i - \bar{x}| = (n_j - n_k)\ \bar{x} + \Sigma x_k - \Sigma x_j$$

avec :

$x_j$, observations inférieures à $\bar{x}$, en nombre $n_j$
$x_k$, observation supérieure à $\bar{x}$, en nombre $n_k$

Le test de Geary est surtout sensible au caractère trop ou insuffisamment aplati de la distribution.

3 - Test de K. Pearson [17]

Il est basé sur la comparaison des valeurs calculées à partir de l'échantillon, des rapports

$$\beta_1 = \frac{m_3}{m_2^{3/2}} \quad \text{et} \quad \beta_2 = \frac{m_4}{m_2^2}, \qquad m_r = 1/n \sum_{i=1}^{n} (x_i - \bar{x})^r$$

à leurs valeurs théoriques dans le cas d'une loi normale soit :

$$\beta_1 = 0 \qquad\qquad \beta_2 = 3$$

Des tables donnant quelques fractiles de ces distributions, respectivement pour $n > 25$ et $n > 50$, ont été publiées dans Biometrika Tables [17] , ainsi que des abaques dans la revue Biometrika (1936 p. 304-7).

Le test conduira au rejet de l'hypothèse de normalité au risque $\alpha$, si $|\beta_1| > \beta_{1,(1-\alpha/2)}$ ou si $\beta_2$ est extérieur à l'intervalle $(\beta_{2,(\alpha/2)}, \beta_{2,(1-\alpha/2)})$.

On trouvera en annexe un extrait de ces tables.

Le test de K. Pearson est sensible à la fois au manque de symétrie caractérisé par $\beta_1$ et au degré d'aplatissement caractérisé par $\beta_2$. Par contre, en raison des moments d'ordre élevé qui interviennent dans les calculs, sa conclusion risque d'être faussée par la présence d'observations aberrantes (anormalement grandes ou petites). Ce problème des observations aberrantes devra être examiné avant le calcul de $\beta_1$ et $\beta_2$ ;

Exemple

Un échantillon de $n = 200$ a donné :

$$\beta_1 = -\ 0{,}51 \qquad \beta_2 = 2{,}83$$

Au risque $\alpha = 2\ \%$, les fractiles correspondants sont :

| | | |
|---|---|---|
| pour $\beta_1$ : | - 0,403 | + 0,403 |
| pour $\beta_2$ : | 2,37 | 3,98 |

Le test global conduit à conclure que la distribution n'est pas normale en raison de son asymétrie significative (avec une queue de distribution plus étalée vers les petites valeurs) bien qu'elle ne soit pas significativement plus aplatie qu'une distribution normale.

4 - Test de David - Hartley - Pearson [6]

Il est basé sur le rapport de l'étendue à l'estimation de l'écart-type de l'échantillon

$$u = \frac{W}{s} = \frac{x_{max} - x_{min}}{\left[ \frac{1}{n-1} \Sigma (x - \bar{x})^2 \right]^{1/2}}$$

Des tables des fractiles de la distribution de u ont été calculées pour n = 3....20.. 30.....100.....1000, pour les valeurs $\alpha$ ou $1 - \alpha = 0,005$ - 0,01 - 0,025 - 0,005 - 0,10. (voir table V)

Le test conduit au rejet de l'hypothèse de normalité au risque $\alpha$ si la valeur calculée de u est extérieure à l'intervalle $u_{\alpha/2}$, $u_{1-\alpha/2}$ donné par les tables.

5 - Test de Shapiro - Wilk [19] [20]

Il est basé sur l'étude du rapport du carré d'une certaine combinaison linéaire des étendues incomplètes (c'est-à-dire des différences des observations ordonnées : $x_n - x_1$, $x_{n-1} - x_2$ ....., $x_{n-j+1} - x_j$, $j < n/2$) à la somme des carrés des écarts à la moyenne des n observations.

La statistique utilisée est :

$$W = \frac{\left[ \sum_{j=1}^{k} a_j \, d_j \right]^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

avec :

$$d_j = x_{n-j+1} - x_j \qquad \begin{cases} k = \frac{n}{2} \text{ si n est pair} \\ k = \frac{n-1}{2} \text{ si n est impair} \end{cases}$$

Les coefficients $a_j$ étant donnés en fonction de n par une table, on dispose d'autre part d'une autre table donnant, en fonction de n, quelques fractiles usuels de la distribution de W.

On trouvera en annexe un extrait de ces tables.

De même que dans le test de Geary, il s'agit encore du rapport des carrés de deux quantités associées toutes deux à l'estimation de la dispersion (le numérateur dépendant d'une fonction linéaire des étendues tronquées).

Le test conduira à rejeter l'hypothèse de normalité au risque $\alpha$ si la valeur calculée de W est extérieure à l'intervalle $(W_{\alpha/2}, W_{1-\alpha/2})$.

6 - Test de R. d'Agostino [1]

Tenant compte, lui aussi, d'une certaine fonction des observations ordonnées $x_{1n}, \ldots x_{in}, \ldots, x_{nn}$, le test d'Agostino peut être considéré comme un prolongement du test de Shapiro-Wilk, dont la table est limitée à $n = 50$.

Il est basé sur les statistiques

$$T = \sum_{i=1}^{n} \left[ i - \frac{1}{2}(n+1) \right] x_{in}, \qquad S^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - x)^2$$

$$D = \frac{T}{n^2 S}, \qquad Y = \frac{D - 0,2821}{0,03} \sqrt{n}$$

Une table donne quelques fractiles usuels, ($\alpha$ ou $1 - \alpha = 0,5 - 1 - 2,5 - 5 - 10$ % = de la distribution de Y pour $n = 50(1) \ldots 100(50) \ldots 1000$.

## II - METHODES BASEES SUR LA COMPARAISON DE LA DISTRIBUTION CUMULATIVE OBSERVEE A LA FONCTION DE REPARTITION D'UNE LOI NORMALE AJUSTEE AUX OBSERVATIONS

### II.1 - Estimation des fréquences cumulées

Ces méthodes impliquent le calcul ou l'estimation de la fréquence cumulée des observations jusqu'à $X = x_i$ pour les valeurs observées ou, éventuellement, pour quelques unes d'entre elles (limites de classes).

S'il s'agit d'un échantillon important permettant la répartition des observations en un nombre suffisant de classes (au moins 10 à 15 classes), la fréquence des observations inférieures à $X = x_i$ est $S_n(x_i) = n_i/n$ avec

- $n_i$ nombre des observations inférieures ou égales à $X = x_i$,
- n effectif de l'échantillon.

Mais dans le cas d'observations individuelles ordonnées

$$x_1 < x_2 < \ldots\ldots < x_n$$

en trop petit nombre pour permettre une répartition par classes, on sera obligé d'estimer $S_n(x_i)$, car pour $X = x_i$ la valeur $n_i$, effectif cumulé jusqu'à et y compris $X = x_i$ est une variable aléatoire, variable d'un échantillon à un autre, pour une même population. Il en est de même de $S_n(x_i)$, qu'il est nécessaire d'estimer statistiquement en fonction de l'information donnée par l'échantillon.

On est donc confronté au problème de l'estimation d'une valeur centrale d'une variable aléatoire $S_n(x_i)$ compte tenu de sa loi de distribution qui est connue (on montre qu'elle ne dépend que de i et de n).

De même la fonction de répartition F(x), loi normale ajustée aux observations, devra être déterminée à partir de celle-ci, ses paramètres m et $\sigma$ étant estimés par :

$$m = \bar{x} = \frac{\sum_{1}^{n} x_i}{n}, \qquad \sigma^2 = s^2 = \frac{\sum_{1}^{n} x_j^2 - n\,\bar{x}^2}{n - 1}$$

En ce qui concerne l'estimation de $S_n(x)$, diverses solutions ont été proposées :

a) $S_n^*(x_i) = i/n$

Cette estimation, particulièrement simple, ne tient pas compte du caractère continu de la variable aléatoire X.

b) $S_n^*(x_i) = \frac{i - 0,5}{n}$

Cette solution proposée par Chernoff [3] en 1954 consiste à considérer que la courbe cumulative empirique est jalonnée par les milieux des segments verticaux du graphique en escalier correspondant aux observations.

c) $S_n^*(x_i) = \frac{i}{n + 1}$

Cette valeur est l'espérance mathématique de la variable aléatoire $S_n(x_i)$, dont la loi de densité est une loi Beta de paramètres i et n - i + 1 , [2].

d) $S_n^*(x_i) = M[S_n(x_i)] = \frac{i\, F_{0,50}}{(n - i + 1) + i\, F_{0,50}}$

$[\nu_1 = 2i,\ \nu_2 = 2(n - i + 1)]$

$F_{0,50}$ étant la valeur médiane de la variable F de Fisher-Snedecor avec les degrés de liberté $\nu_1$ et $\nu_2$.

Cette valeur $S_n(x_i)$ est la valeur médiane de la variable aléatoire $S_n(x_i)$, valeur qui a une égale probabilité de ne pas être atteinte ou d'être dépassée. Elle a été proposée par Johnson [12] sous le nom de "median rank". Ses valeurs ont été tabulées, on trouvera en annexe un extrait de cette table, que l'on pourrait prolonger à l'aide de la formule ci-dessus et d'une table des valeurs de $F_{0,50}(\nu_1, \nu_2)$ (Tables de Hald [11]).

e) $S_n^*(x_i) = F(\xi_i)$

Dans cette formule, F désigne la fonction de répartition de la variable normale réduite u et $\xi_i = E(u_i)$, $u_i$ étant la $i^{ème}$ observation ordonnée d'un échantillon aléatoire de n variables normales réduites.

Cette solution a été proposée par E. Pearson. Une table des valeurs de $\xi_{in}$ est donnée dans Biometrika Tables [17]. On en trouvera un extrait en annexe.

Les quatre dernières solutions, identiques pour $i = \frac{n + 1}{2}$ (n impair , restent très voisines pour des valeurs de i situées dans la région centrale de l'intervalle (i, n), mais elles diffèrent systématiquement aux extrémités.

Comparaison des estimations de $S_n(x_i)$

1/ n = 9

| | a | b | c | d | e |
|---|---|---|---|---|---|
| i = 1 | 0,111 | 0,055 | 0,100 | 0,074 | 0,069 |
| 5 | 0,555 | 0,500 | 0,500 | 0,500 | 0,500 |
| 9 | 1 | 0,944 | 0,900 | 0,926 | 0,931 |

2/ n = 49

| | a | b | c | d | e |
|---|---|---|---|---|---|
| i = 1 | 0,204 | 0,010 | 0,020 | 0,013 | 0,014 |
| 25 | 0,510 | 0,500 | 0,500 | 0,500 | 0,500 |
| 49 | 1 | 0,990 | 0,980 | 0,987 | 0,986 |

Bien que les trois dernières estimations soient définies théoriquement de manière précise, aucun critère ne permet de dire que systématiquement et généralement l'une d'entre elles est supérieure aux autres.

Chacune présente pour l'utilisateur des avantages et des inconvénients (voir § II.2).

La formule donnant la valeur médiane (d), permet en y remplaçant 0,50 par $\alpha/2$ et $1 - \alpha/2$ de définir un intervalle de confiance à $1 - \alpha$ pour la variable aléatoire $S_n(x_{in})$, à partir des tables de la distribution de la variable F.

Pour une valeur donnée de $\alpha$ et pour un effectif donné n, l'amplitude des intervalles de confiance est maximale pour i voisin de n/2 (pour $i = \frac{n+1}{2}$ si n est impair) ; pour i fixé, elle décroit lorsque n augmente.

$S_{0,05}$

| n \ i | 1 | $\frac{n+1}{2}$ | n |
|---|---|---|---|
| 9 | 0,0057 | 0,25 | 0,717 |
| 39 | 0,0013 | 0,372 | 0,926 |
| 499 | 0,0001 | 0,463 | 0,994 |

$S_{0,95}$

| n \ i | 1 | $\frac{n+1}{2}$ | n |
|---|---|---|---|
| 9 | 0,28 | 0,75 | 0,994 |
| 39 | 0,074 | 0,628 | 0,987 |
| 499 | 0,0060 | 0,537 | 0,9999 |

$S_{0,95} - S_{0,05}$

| n \ i | 1 | $\frac{n+1}{2}$ | n |
|---|---|---|---|
| 9 | 0,274 | 0,50 | 0,27 |
| 39 | 0,073 | 0,26 | 0,06 |
| 499 | 0,006 | 0,07 | 0,006 |

Ces marges donnent une idée de l'incertitude de l'interprétation graphique, à l'aide d'une droite (voir ci-après), d'un ensemble de points $[x_i, S_n^*(x_{in})]$ fournis par un petit échantillon.

## II.2 - Comparaison de la distribution cumulative $S_n^*(x)$ à la distribution normale ajustée aux observations

### 1/ Test de Kolmogorov - Lilliefors [14]

Ce test résulte d'une adapatation des tests classiques de Kolmogorov Smirnov (comparaison de deux distributions cumulatives observées et comparaison d'une distribution cumulative à une distribution normale donnée à priori).

La statistique utilisée est

$$D = \max |F^*(x) - S_n^*(x)|$$

valeur maximale des écarts entre la fonction de répartition $F^*_n(x)$ de la loi normale ajustée aux observations avec $\sigma^2$ estimé par $s^2 = \frac{\sum_1^n (x_i - x)^2}{n - 1}$ et m estimé par $\bar{x} = \frac{\sum_1^n x_i}{n}$ et la distribution cumulative définie par $S_n^*(x_i) = i/n$(1).

L'étude de la loi de la variable D ne pouvant être faite par voie analytique, a été réalisée par la méthode de Monte-Carlo : pour chaque valeur de n, 1000 échantillons issus de populations normales ont été tirés permettant l'estimation des fractiles de la distribution de D pour n = 4, ..., 30 et au-delà.

On trouvera en annexe un extrait de la table donnée par Lilliefors.

L'hypothèse de normalité est rejetée au risque $\alpha$ choisi de rejeter une hypothèse vraie, si la valeur trouvée de D est supérieure à la valeur donnée par la table en fonction de n et $\alpha$.

Les valeurs critiques du test de Lilliefors sont, pour $\alpha = 0,05$, environ 1,55 fois plus petites que les valeurs du test de Kolmogorov-Smirnov relatif à la comparaison d'une distribution observée et d'une loi normale donnée a priori.

2/ Test de Kuiper, Louter et Koerts [15]

Avec les mêmes définitions de $S_n^*(x)$ et de $F^*(x)$, la statistique utilisée par Kuiper est :

$$V = \max_x [S_n^*(x) - F^*(x)] - \min_x [S_n^*(x) - F^*(x)]$$

elle tient compte séparément, aussi bien des différences positives que des différences négatives, le dernier terme de V étant en général négatif.

La table des fractiles de la distribution de V a aussi été déterminée par la méthode de Monte-Carlo à l'aide de 10 000 échantillons pour chaque système de valeurs de n et $\alpha$.

Pour un même couple de valeurs de n et $\alpha$, le rapport des fractiles V/D reste voisin de 1,6.

## II.3 - Méthodes graphiques

Toutes ces méthodes sont des adaptations de la méthode graphique de Henry : possibilité d'ajustement graphique d'une droite à un ensemble de n points définis par leurs abscisses (valeurs observées) et leurs ordonnées (es-

---

(1) Pour un petit nombre d'observations, la distribution cumulative définit un diagramme en escalier, et on a :

$$S_n^*(x) = 0 \text{ pour } x < x_1, \quad S_n^*(x) = \frac{i}{n} \text{ pour } x_i < x < x_{i+1} \; ; \quad S_n^*(x) = 1 \text{ pour } x > x_n$$

timations des valeurs correspondantes de la fonction de répartition (ou de la variable normale réduite ordonnée) qui correspond aux observations $x_i$)(1).

Ces méthodes ne diffèrent que par le choix, qui reste arbitraire, de l'estimation de l'ordonnée $S_n(x)$, et aussi par le choix du papier utilisé pour construire le graphique.

- Si l'on utilise les estimations (b), (c) ou (d) il sera souvent commode d'utiliser du papier ayant une échelle de probabilité normale. Dans ce cas les estimations $\frac{i - 0,5}{n}$ et $\frac{i}{n + 1}$ de calcul facile, permettront de placer immédiatement les points $[x_i, S_n^*(x_i)]$, à l'aide de l'échelle fonctionnelle en $S_n^*(x)$.

Par contre, l'utilisation de la valeur médiane nécessitera de recourir préalablement à la table de Johnson.

Par ailleurs l'emploi du papier à échelle de probabilité normale, oblige à utiliser l'échelle fonctionnelle graphique choisie par l'éditeur et qui peut paraitre ne pas convenir au dessinateur. On peut alors utiliser du papier millimétré ordinaire, permettant de choisir à sa guise les deux échelles métriques en remplaçant l'échelle de probabilité normale par une échelle métrique en u, en faisant correspondre à chaque valeur $S_n^*(x_i)$ la valeur $u_i$ définie par une table de la loi normale, c'est-à-dire par

$$\int_{-\infty}^{u_i} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = S_n^*(x_i) \quad (2)$$

- Si l'on utilise la variable normale ordonnée $\xi_i$, ce qui implique l'usage d'une table spéciale, le graphique des points $(x_i, \xi_i)$, se fera simplement sur papier millimétré, mais si l'on est intéressé ultérieurement par l'estimation des valeurs de la fonction de répartition ajustée graphiquement, il faudra, pour chaque valeur de $\xi$, calculer

$$F^*(x) = \int_{-\infty}^{\xi} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$

qui peut être lue directement sur le graphique à échelle de probabilité normale.

Avec cette variable normale ordonnée, on peut évidemment aussi utiliser le papier à échelle de probabilité normale si on prend comme ordonnée $S_n^*(x_i) = F(\xi_i)$ déduite de la table de la loi normale.

En fait, compte tenu du caractère quelque peu subjectif du tracé d'une droite d'estimation, lorsqu'il parait possible, ce choix entre les diverses

---

(1) Toutes les remarques ci-après sont évidemment valables dans d'autres cas d'anamorphose linéaire : distributions lognormales de Weibull, Gumbel, Frechet .pour lesquelles des échelles fonctionnelles permettent aussi de représenter la fonction de répartition par une droite.

(2) Pour divers couples (i, n), $1 < i, n < 40$, une table calculée par M. Dumas [8], donne les valeurs, exprimées en mm, de $50\sqrt{2}\, u_i$, permettant l'usage de blocs de papier millimétré de format courant.

estimations des fréquences cumulées n'a que peu d'importance, les points représentatifs correspondant à une même valeur de $x_i$ étant très rapprochés, sauf aux extrémités (pour i voisin de 1 ou de n).

Cependant on peut noter que, d'une manière générale, on a :

$$\frac{i - 0,5}{n} < F(\xi_i) < M[S_n(x_i)] < \frac{i}{n + 1} \quad \text{pour } i < \frac{n + 1}{2}$$

et

$$\frac{i - 0,5}{n} > F(\xi_i) > M[S_n(x_i)] > \frac{i}{n + 1} \quad \text{pour } i > \frac{n + 1}{2}$$

d'où il peut résulter, pour le dessinateur de légères différences dans l'appréciation de la pente de la droite ajustée, pente qui caractérise l'estimation approximative de l'écart-type de la distribution.

A titre d'exemple pour n = 20 les estimations ci-dessus envisagées sont

| | b | e | d | c |
|---|---|---|---|---|
| i = 1 | 0,025 | 0,031 | 0,034 | 0,048 |
| i = 10 | 0,475 | 0,475 | 0,476 | 0,476 |
| i = 20 | 0,975 | 0,969 | 0,966 | 0,952 |

La droite d'ajustement ayant pu être éventuellement tracée, on en déduira, conformément au schéma ci-après (figure 1), les estimations de m et $\sigma$. Ces estimations, et particulièrement celle de $\sigma$, n'ont évidemment, dans la plupart des cas, qu'une valeur approximative.

Lorsqu'il n'y a pas linéarité approximative du graphique, son allure générale peut donner une indication quant aux divergences qui peuvent exister par rapport à une distribution normale : une concavité tournée vers le haut ou vers le bas caractérise une distribution dissymétrique dont le sommet est déplacé respectivement vers les grandes ou les petites valeurs. Elle peut être l'indice de la validité possible d'un ajustement par une distribution log-normale dont l'étude graphique pourra se faire de la même manière en utilisant un papier à échelle de probabilité normale en ordonnée et à échelle logarithmique en abscisse. On pourra, éventuellement, envisager d'autres changements de variables tels que $y = \log (x + a)$, $y = \sqrt{x}$, ....

Quel que soit le mode d'estimation des fréquences cumulées $S_n (x_i)$ -et ceci n'a pratiquement que peu d'importance- la simple méthode graphique, proposée par Henry, consiste à tracer une droite $\Delta$ paraissant compenser "au mieux" les différences d'ordonnées entre les n points ainsi déterminés $[x_i \; S_n (x_i)]$ et les points de même abscisse de la droite $\Delta$.

Même si cette compensation graphique parait convenablement réalisée, il faut se demander si les écarts constatés sont compatibles avec l'hypothèse de normalité et peuvent être imputés aux hasards de l'échantillonnage.

Si l'on admet que la droite de Henry représente rigoureusement la loi de probabilité cherchée et si $F_i$ est l'ordonnée du point observation d'abscisse $x_i$ de la droite, on peut, si n est suffisamment grand, approximativement

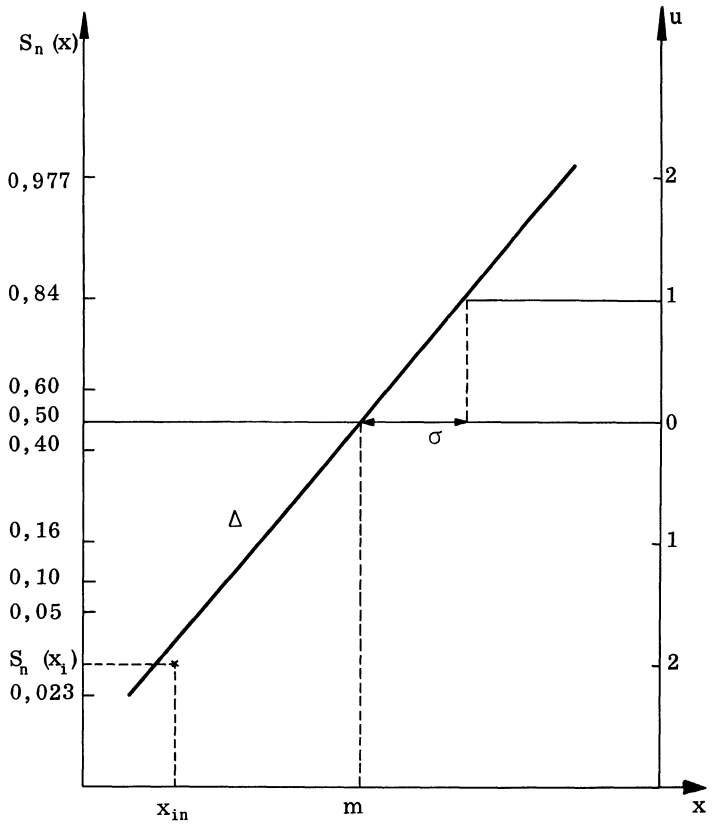



Figure 1

considérer que sur n observations, on en a trouvé i qui sont inférieures à $x_i$.

Dans ces conditions l'écart $\frac{i}{n} - F_i$ suit approximativement une loi normale de moyenne nulle et d'écart-type $\sigma = [F_i\ (1 - F_i)/n]^{1/2}$, ceci au moins dans la région où $n\ F_i(1 - F_i)$ est suffisamment grand pour que l'approximation normale d'une loi binomiale soit valable (par exemple supérieur à 10).

On pourrait alors associer à chaque abscisse $x_i$ un intervalle $F_i \pm u_{1-\frac{\alpha}{2}}\sigma$ définissant, pour l'ensemble des points observés, une région d'acceptation de l'hypothèse de normalité au risque $\alpha$ de rejeter une hypothèse vraie (u = variable normale réduite).

M. Dumas [8] a, pour sa méthode graphique, signalée ci-dessus, calculé les limites de ces intervalles pour $\alpha = 0,50$.

Haag [11] avait préconisé cette méthode dans le cas particulièrement favorable de l'étude de la normalité de tirs balistiques d'artillerie de 60 coups tirés dans des conditions aussi identiques que possible.

Il a d'ailleurs signalé les faiblesses de la méthode, particulièrement lorsque n est petit, en raison du caractère subjectif du tracé de la droite

D et de la non validité de l'approximation normale pour les observations extrêmes pour lesquels n F(1 - F) est petit.

Le test de Lilliefors, plus objectif, implique en principe le calcul des valeurs $F^*(x_i)$ de la distribution normale ajustée $\mathcal{N}(\bar{x}, s^2)$, après l'estimation de x et $s^2$, mais il est facile d'y faire correspondre une solution graphique en traçant la droite $\Delta$ définie, sur un graphique à échelle de probabilité normale par les deux points $(\bar{x}, 50\ \%)$ et $(\bar{x} + s, 84{,}2\ \%)$.

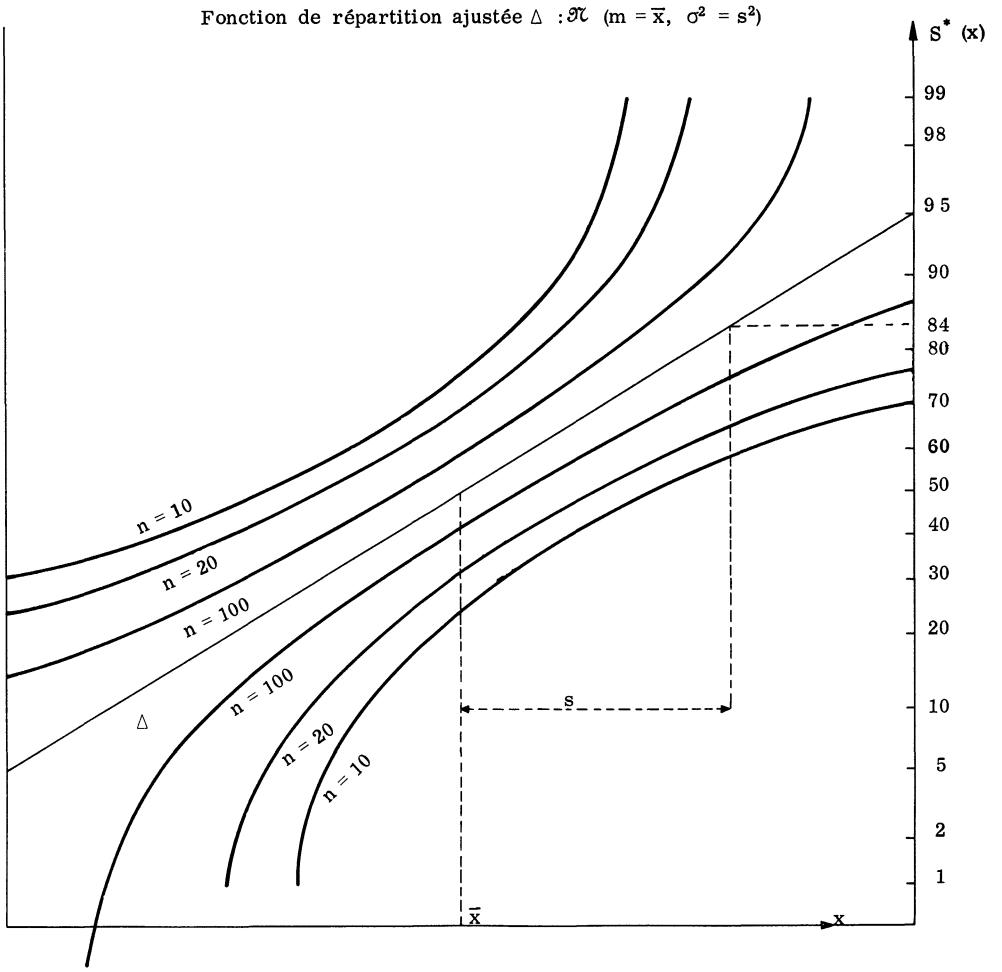
On notera d'ailleurs que si le problème de la normalité se pose c'est en général parce que l'on a envisagé d'utiliser cette normalité dans des problèmes d'estimation et de test, relatifs soit à des moyennes, soit à des variances, problèmes qui impliqueront le calcul de $\bar{x}$ et de $s^2$.

On peut alors, sur le graphique, mesurer, compte tenu de l'échelle normale, les écarts $|F^*(x) - S_n^*(x)|$ entre les points observations et la droite $\Delta$ et comparer le plus grand d'entre eux aux valeurs critiques données par la table de Lilliefors pour le risque choisi.

Figure n° 2 Test de Kolmogorov-Lilliefors

Limites des régions d'acceptation au risque 10 % de l'hypothèse de normalité (n = 10, 20 et 100)

Fonction de répartition ajustée $\Delta$ : $\mathcal{N}$ $(m = \bar{x},\ \sigma^2 = s^2)$



On peut aussi délimiter une zone d'acceptation de l'hypothèse de normalité, au risque choisi, $\alpha$, en traçant les courbes définies à partir de la droite $\Delta$, à l'aide des différences critiques données par la table. Il suffira pratiquement de faire cette transformation pour cinq points répartis sur la partie utile de la droite $\Delta$ (figure 2).

Pratiquement, il semble bien que si l'on s'intéresse seulement à la réponse qualitative à la question "normalité ou non normalité" et si l'on veut éviter le petit calcul de $\overline{x}$ et $s^2$ (!) -au moins dans les cas où le simple tracé graphique de la droite $\Delta$ ne parait pas devoir donner lieu à des variantes très différentes suivant l'appréciation du dessinateur- l'on puisse utiliser pratiquement les critères déduits de la table de Lilliefors, relativement à cette droite.

## III - EXEMPLES

Les diverses méthodes envisagées ci-dessus sont appliquées à un même exemple : étude de la normalité d'une distribution définie par un échantillon de 20 observations. (Cet échantillon est constitué par 20 valeurs consécutives prises dans la table des variables normales aléatoires de la "Rand Corporation", pour une population normale $\mathcal{N}(m = 2,\ \sigma = 1)$.

Ces 20 valeurs ordonnées sont :

0,790 - 0,843 - 1,286 - 1,331 - 1,383 - 1,411 - 1,757 - 1,874 - 1,985 - 2,006 - 2,009 - 2,355 - 2,481 - 2,560 - 2,980 - 3,129 - 3,284 - 3,287 - 3,328 - 3,532.

Les différentes méthodes envisagées donnent les résultats ci-après :

1/ Test de Geary (Table n° 4)

$$\overline{x} = 2,181, \quad \Sigma|x_i - \overline{x}| = 2x + 29,636 - 16,677 = 14,620$$

$$\sum_1^{20} (x_i - \overline{x})^2 = 14,232, \qquad a = \frac{14,620}{[284,64]^{1/2}} = 0,866$$

valeur intérieure à l'intervalle bilatéral d'acceptation au risque 0,05 (soit 0,71 - 0,89).

2/ Test de David (Table n° 5)

$$u = \frac{W}{s} = \frac{2,742}{0,865} = 3,17$$

valeur intérieure à l'intervalle bilatéral d'acceptation au risque 0,05 (soit 3,10 - 4,67).

3/ Test de Shapiro-Wilk (Table n° 6)

$$W = \frac{\left(\sum_{1}^{10} a_j \, d_j\right)^2}{\sum_{1}^{20} (x_i - \bar{x})^2} = \frac{18,373}{14,232} = 0,939$$

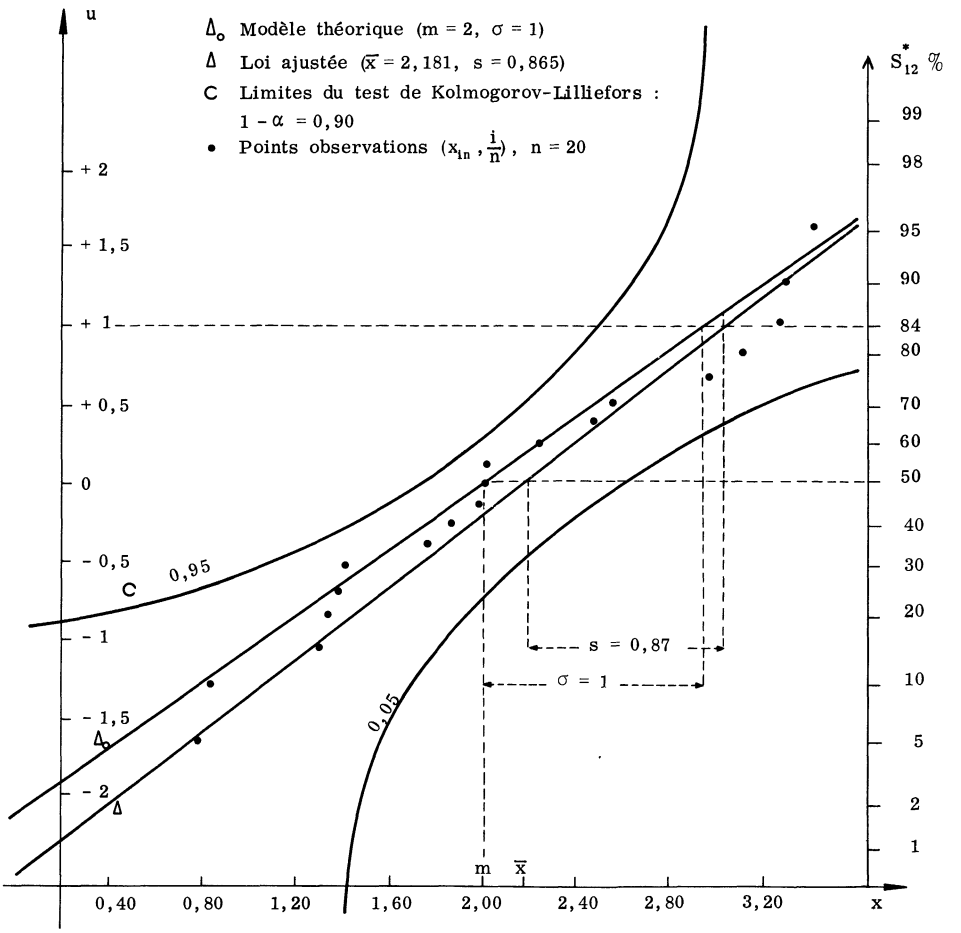
valeur intérieure à l'intervalle bilatéral d'acceptation au risque 0,05 (soit 0,837 - 0,986).

4/ Test de Lilliefors (Table n° 7)

$$D = \mathrm{Max}_x \, |F^* (x) - S^* (x)| = 0,111$$

Figure n° 3

Etude graphique de la normalité d'une distribution
(échantillon tiré d'une population normale (m = 2, σ = 1)



valeur inférieure à la limite supérieure de l'intervalle d'acceptation au risque unilatéral 0,025 (soit 0,216).

5/ Test de Kuiper-Louter (Table n° 7)

$$V = \mathrm{Max}_x\,[S^*(x) - F^*(x)] - \mathrm{Min}_x\,[S^*(x) - F^*(x)] = 0{,}111 - (-\,0{,}072 = 0{,}183$$

valeur inférieure à la limite supérieure de l'intervalle d'acceptation au risque unilatéral 0,025 (soit 0,333).

6/ Etude graphique du test de Lilliefors

Sur le graphique à échelle normale (figure 3), on a marqué les 19 points ($x_i$, i/20), (i = 1 ..... 19, le point i = 20 ne pouvant évidemment pas être représenté), et on a tracé la droite Δ correspondant à la distribution normale définie à partir des observations par $\overline{x}$ = 2,181, s = 0,865.

On a de plus tracé les courbes limitant les zones d'acceptation aux risques 5 % relatifs à la valeur absolue des écarts $F^*$ (x) - S (x).

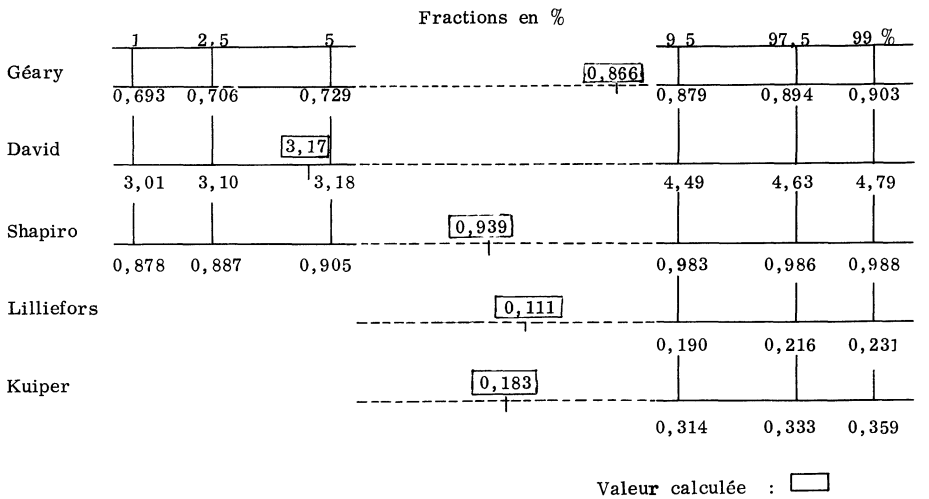



Figure 4 - Comparaison des statistiques calculées à leurs lois de probabilité

La figure 4 montre la position des valeurs numériques trouvées, dans les divers tests, par rapport à certains fractiles des distributions des statistiques utilisées ; on constate, dans le cas présent, que tous ces tests conduisent à la même conclusion qualitative : acceptation de l'hypothèse de normalité. Mais ces conclusions correspondent à des risques réels différents qui dépendent de la position des valeurs calculées sur l'échantillon, dans les lois de probabilité des statistiques utilisées.

Ces statistiques, définies en fonction de certaines combinaisons particulières des observations ne sont pas sensibles de la même manière aux fluctuations aléatoires d'un échantillon tiré au hasard d'une population normale (m, $\sigma^2$), et il peut fort bien arriver, dans un cas particulier que, pour un risque de première espèce fixé, l'utilisation de ces tests donne des résultats légèrement contradictoires en apparence.

## CONCLUSION

Comment tester la normalité d'une distribution et quelle méthode convient-il d'utiliser est une question souvent posée par le praticien et, particulièrement, par l'ingénieur responsable des décisions prises en matière de contrôle statistique, et, fréquemment, à partir d'échantillons d'effectif assez petit.

Bien que les techniques d'étude des petits échantillons aient été introduites dans l'analyse statistique depuis près de 50 ans, le problème reste ouvert et le restera longtemps quelles que soient les nouvelles méthodes proposées et la facilité des calculs numériques résultant des moyens modernes.

D'ailleurs, la réponse au test de normalité n'est généralement pas le but final de l'opération : on désire souvent savoir si l'on peut valablement utiliser les lois de Laplace-Gauss, de Student-Fisher, de $\chi^2$ ou de Fisher-Snedecor dans des tests ultérieurs d'estimation ou de comparaison de moyennes ou de variances.

Si le nombre des observations utilisées est grand, le test de normalité peut conduire à conclure à un écart significatif de non-normalité, alors que, cependant, les lois précédentes pourront être utilisées sans erreur appréciable dans l'étude des paramètres estimés sur le(s) échantillon(s).

Si le nombre des observations est petit, il peut arriver qu'aucun test ne soit capable de mettre en évidence de façon significative la non-normalité bien que dans un pareil cas l'emploi des distributions envisagées ci-dessus soit sérieusement erroné.

Mais que signifient les mots "grand" ou "petit" ? Là encore aucune réponse formelle ne peut être donnée : cela dépend de la nature de l'hypothèse alternative qui peut être envisagée et du but poursuivi. Dans certains cas pratiques, par exemple en contrôle statistique des fabrications, on peut généralement admettre que les lois dérivées de la loi normale peuvent être valablement utilisées si $n > 10$ (tests relatifs aux moyennes) ou $n > 30$ (tests relatifs aux variances).

Cette dernière remarque montre aussi que dans certains cas il peut être préférable de tenir compte de l'information préalable possédée sur la population dont provient l'échantillon plutôt que des résultats d'un simple test sur un petit échantillon.

Enfin, il importe de noter qu'un test de normalité est un test composite non paramétrique, aussi bien en ce qui concerne l'hypothèse à tester que l'hypothèse alternative - généralement non précisée même qualitativement.

Il en résulte que la puissance d'un test particulier ne peut être ni formulée, ni calculée (sauf par simulation dans le cas d'une alternative particulière bien précisée) et qu'évidemment il ne peut pas exister de test uniformément le plus puissant contre toutes les hypothèses alternatives possibles.

Théoriquement on peut considérer qu'une distribution normale est complètement caractérisée par la nullité de ses moments d'ordre impair et par les relations liant ses moments d'ordre pair à la variance. Pratiquement

ceci conduit à faire appel aux autres premiers moments et à utiliser parallèlement les deux tests de K. Pearson ($\beta_1$ : symétrie, et $\beta_2$ : aplatissement).

Les valeurs calculées de $\beta_1$ et $\beta_2$ sur un échantillon de n observations risquent, en raison du calcul de moments d'ordre élevé, d'être très sensibles à la présence d'observations très éloignées de la moyenne, particulièrement pour de petits échantillons : il en résulte que les tables de la distribution de $\beta_1$ n'ont été publiées jusqu'à une date récente que pour $n > 25$ et celles de $\beta_2$ pour $n > 50$. (1)

L'emploi simultané de $\beta_1$ et $\beta_2$ exige donc des échantillons importants.

Il en est de même du test $\chi^2$ en raison de la nécessité de grouper les observations en classes. Sa sensibilité dépend du nombre des classes, lié à l'intervalle de classe, arbitrairement choisi.

L'étude purement graphique de la normalité n'est pas un test au sens statistique du mot, de plus on peut lui reprocher son caractère subjectif, Cependant ce ne sont pas des raisons suffisantes pour en rejeter systématiquement l'emploi : en plus de l'indication qualitative, certes peu précise qu'elle fournit, cette méthode est, même pour de petits échantillons, un guide utile pour décider si une certaine fonction de la variable x, par exemple log x ou $\sqrt{x}$, a une distribution plus voisine de la normale que la variable x elle-même.

Il semble d'ailleurs, comme il a été dit plus haut, que sous réserve du calcul préalable de $\overline{x}$ et $s^2$, on puisse associer aisément la méthode graphique au test de Lilliefors de manière à préciser le risque de première espèce (rejet de la normalité alors qu'elle est vraie), associé à la décision.

Pour les risques de première espèce définies par $\alpha = 0,01 - 0,05 - 0,10$, les puissances des tests de Lilliefors et de Kuiper ont été comparées par la méthode de simulation, par Louter [15] pour des échantillons d'effectif n = 10, 20 et 30 et pour quelques alternatives spécifiées (uniforme, $\chi^2(\nu = 3)$, t de Student ($\nu = 3$), exponentielle, F de Fisher-Snedecor ($\nu_1 = 3$, $\nu_2 = 5$) et Cauchy).

Pour ces hypothèses alternatives, en nombre trop limité pour en tirer une conclusion générale, on constate que les puissances des deux tests sont en général très voisines avec un léger avantage pour le test de Kuiper, avantage qui ne compense pas l'impossibilité d'associer à ce dernier des limites graphiques d'intervalles de confiance.

Il convient de plus de remarquer que ces deux tests n'utilisent pas toute l'information fournie par les observations puisqu'ils sont basés uniquement sur les valeurs extrêmes des différences $F^*(x) - S_n^*(x)$.

Les deux tests de Shapiro-Wilk et d'Agostino ont évidemment le mérite d'utiliser toute l'information fournie par les observations, ainsi que les deux tests de K. Pearson, ce qui n'est pas le cas ni pour le test de David, ni pour le test de Geary.

---

(1) Des tables de la distribution de $b_2$ ont été recemment calculées par simulation pour n = 7(1) .... 12, 15(5) .... 50.
R. d'Agostino et G. Tietjen : Biometrika 1971 n° 3 p. 670.

Diverses tentatives de comparaison de ces tests ont été faites, par simulation, par Louter et Koerts, par Shapiro, Wilk et Chen et par d'Agostino.

Malheureusement les puissances ainsi calculées ne correspondent pas toutes ni aux mêmes effectifs n, ni aux mêmes alternatives. Pour quelques alternatives particulières, on peut cependant noter les résultats suivants, pour $n = 50$. Les tests de Kolmogorov-Lilliefors (K.L.) et de Kuiper sont, en général, moins puissants que les tests de Shapiro-Wilk (W), d'Agostino (Y), Pearson ($b_2$) et David (u).

- Alternative : loi uniforme

Le test Y est moins puissant que les tests W, $\beta_2$ et u qui ont sensiblement la même puissance voisine de 0,95 (le test $\beta_1$ ayant dans ce cas une puissance voisine de zéro).

- Alternative : loi de $\chi^2$, $\nu = 2$

Les tests W, $\beta_1$ et $\chi^2$ ont sensiblement une même puissance très élevée, supérieure à celle des tests Y, $b_2$ et K.L., très nettement supérieure à celle du test u.

- Alternative : loi de t, $\nu = 2$

Les cinq tests Y, W, $\beta_1$, $\beta_2$ et u ont des puissances relativement élevées, nettement supérieures à celles des tests $\chi^2$ et K.L.

- Alternative : loi log-normale ($\mu = 0$, $\sigma^2 = 1$)

Les cinq tests Y, W, $\beta_1$, $\beta_2$, $\chi^2$ et K.L. sont très puissants, nettement plus puissants que le test u.

Bien sûr, ces indications très partielles, relativement à un problème aux multiples aspects, ne permettent pas de donner au statisticien une règle systématique. Peut-être, cependant, pourront-elles, compte tenu de l'information qu'il possède a priori, des raisons qu'il a de se poser la question, de l'usage qu'il veut faire de la conclusion du test, de l'importance du risque matériel ou financier -au sens courant du mot- associé à une acceptation erronée de la normalité, et de la commodité de la méthode, l'aider à choisir celle (ou celles) qui lui paraissent le mieux convenir à son problème.

Puissances en % de quelques tests de normalité, $\alpha = 0,10$[1]
(Echantillons d'effectifs n = 10, 20, 50, méthode de simulation)

Test utilisé

| Alternative | W | | | Y | $\beta_1$ | | | $\beta_2$ | | | $\chi^2$ | | | u | | | K.L. | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | n=10 | 20 | 50 | 50 | 10 | 20 | 50 | 10 | 20 | 50 | 10 | 20 | 50 | 10 | 20 | 50 | 10 | 20 | 50 |
| Uniforme | 15 | 39 | 96 | 70 | 5 | 0 | 1 | 20 | 38 | 93 | 14 | 18 | 22 | 32 | 53 | 99 | 14 | 19 | 26 |
| Exponentielle $\chi^2(\nu = 2)$ | 57 | 89 | * | 90 | 43 | 82 | * | 27 | 44 | 81 | 31 | 43 | 96 | 15 | 12 | 21 | 32 | 37 | 65 |
| $\chi^2(\nu = 10)$ | 24 | 40 | 74 | 30 | 19 | 35 | 71 | 12 | 25 | 37 | 14 | 14 | 28 | 13 | 12 | 19 | 15 | 19 | 34 |
| Log-normale ($\mu = 0$, $\sigma^2 = 1$) | 72 | 96 | * | 99 | 60 | 91 | * | 46 | 67 | 90 | 80 | 98 | * | 13 | 14 | 43 | 28 | 57 | * |
| Student $t(\nu = 2)$ | 41 | 58 | 84 | 93 | 42 | 54 | 75 | 38 | 58 | 92 | 15 | 22 | 59 | 19 | 44 | 87 | 15 | 23 | 43 |
| Student $t(\nu = 10)$ | 13 | 18 | 31 | 28 | 17 | 18 | 28 | 10 | 15 | 28 | 9 | 9 | 12 | 13 | 14 | 24 | 9 | 11 | 9 |
| Weibull $f(x) = 2x \exp(-x^2)$ | 18 | 24 | 59 | 14 | 16 | 23 | 51 | 11 | 14 | 18 | 11 | 12 | 17 | 11 | 10 | 14 | 12 | 15 | 21 |
| Weibull $f(x) = 0,5\, x^{-0,5} \exp(-x^{0,5})$ | * | * | * | * | 76 | 99 | * | 54 | 83 | 98 | 94 | 99 | * | 19 | 17 | 44 | 58 | * | * |

* : $\sim$ 100 %

(1) Extrait de Shapiro, Wilk et Chen [12] et d'Agostino [1] pour le test Y.

Nota - Il y a, en général, un bon accord entre les valeurs données par Shapiro-Wilk et d'Agostino

## BIBLIOGRAPHIE


[1] d'AGOSTINO R. - An omnibus test of normality for moderate and large scale samples (Biometrika 58, 1971, n° 2 p. 841/48).

[2] CHARTIER F. et MORICE E. - Relations entre quelques lois de probabilité (Revue de Statistique Appliquée 1967 n° 4 p. 17/33).

[3] CHERNOFF, HERMAN and LIEBERMAN - Use of normal probability paper (Jal Am. Statistical Association 1954 p. 722/85).

[4] COCHRAN - The $\chi^2$ test of goodness of fit (Annals of Mathematics Statistics 1952 Vol. 23, n° 3 p. 315).

[5] DAGNELIE P. - A propos de l'emploi du test de Kolmogorov-Smirnov comme test de normalité (Biométrie-Praximétrie 1968 IX, n° 1 p. 3/13).

[6] DAGNELIE P. - Théorie et méthodes statistiques (Editions Duculot, Gembloux, Tome II p. 60/78).

[7] DAVID H.A., HARTLEY H. and PEARSON E. - The distribution of the ratio in a single normal sample of range to standard deviation (Biometrika 1954 p. 482).

[8] DUMAS M. - Les épreuves sur échantillons (Edition C.N.R.S., Paris 1955) (Document A.F.N.O.R. X 06.185).

[9] GEARY R.C. - Tests de la normalité (Ann. Institut Poincaré 1956 p. 35/65).

[10] GEARY R.C. - The ratio of mean deviation to the standard deviation (Biométrika 1936 p. 295/307).

[11] HAAG - Traité du calcul des probabilités : applications au tir (Gauthier-Villars 1926).

[12] HALD A. - Statistical Tables and formulas (John Wiley).

[13] JOHNSON - The median ranks of sample values in their population (Ind. Math. 1951 n° 2).

[14] KIMBALL A.W. - Plotting positions on probability paper (Jal Am. Stat. Ass. 1960 p. 545/60).

[15] LILLIEFORS H.W. - On the Kolmogorov-Smirnov test for normality with mean and variance unknown (Jal Am. Stat. Ass. 1967 p. 393/402).

[16] LOUTER S.A. and KOERTS J. - On the Kuiper test for normality with mean and variance unknown (Stat. Neerl. 1970 p. 83/87).

[17] MASSEY F.J. - The Kolmogorov-Smirnov test for goodness of fit (Jal Am. Stat. Ass. 1951 p. 68).

[18] PEARSON E. and HARTLEY H. - Biometrika Tables for Statisticians (Cambridge Univ. Press 1966).

[19] PEARSON E. - Some comments on the assumption of normality, involved in the I.S.O. documents (Document "International Standardization Organisation" non publié).

[20] SHAPIRO S.A. and WILK M.B. - An analysis of variance test for normality (Biometrika 1965 p. 591/611).

[21] SHAPIRO S.S., WILK M.B. and CHEN H.J. - A comparative study of various tests for normality (Jal Am. Stat. Ass. 1968 p. 1343/72).

[22] SHAPIRO S.S. and WILK M.B. - A comparative study of various tests of normality (Proceed. Am. Math. Stat. Ass. Chicago, 1964).

[23] SMIRNOV - Tables for estimating the goodness of fit of empirical distributions (Ann. Math. Stat. 1948 p. 271/81).

[24] SNEYERS R. - Du test de validité d'un ajustement basé sur les fonctions de l'ordre des observations (Pub. Inst. Royal Météo. Belgique, série B, 1963 p.17).

[25] SRINIVASAN - An approach to testing the goodness of fit of incompletely specified distributions (Biometrika 57, 1970, n° 3 p. 605/11).

[26] VANDEWIELE G. and de WITTE P. - A test of goodness of fit (Stat. Neerl. 1966 p. 87/106).

[27] VESSEREAU A. - Sur les conditions d'application du critérium $\chi^2$ de Pearson (Revue de Statistique Appliquée 1958 n° 2 p. 83/95).

## ANNEXE (Tables)

| | |
|---|---|
| I | Valeur médiane de la fonction de répartition de variables ordonnées |
| $II_1$, $II_2$ | Distribution de $\lvert m_3/(m_2)^{3/2}\rvert$ et $m_4/(m_2)^2$ |
| III | Espérances mathématiques de variables normales réduites |
| IV | Distribution du rapport de Geary |
| V | Distribution du rapport de David |
| VI | Coefficients pour le calcul du rapport de Shapiro-Wilk |
| VI | Distribution du rapport de Shapiro-Wilk |
| VII | Distribution des statistiques de Lilliefors et de Kuiper |
| VIII | Distribution de la statistique de R. d'Agostino |

Tables publiées avec l'autorisation des auteurs et des éditeurs.

Table 1

Valeur médiane de $S_n(x_i)$

(Fonction de répartition estimée à partir des valeurs ordonnées d'un échantillon de n observations)

Effectif de l'échantillon = n

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | .5000 | .2929 | .2063 | .1591 | .1294 | .1091 | .0943 | .0830 | .0741 | .0670 |
| 2 | | .7071 | .5000 | .3864 | .3147 | .2655 | .2295 | .2021 | .1806 | .1632 |
| 3 | | | .7937 | .6136 | .5000 | .4218 | .3648 | .3213 | .2871 | .2594 |
| 4 | | | | .8409 | .6583 | .5782 | .5000 | .4404 | .3935 | .3557 |
| 5 | | | | | .8706 | .7345 | .6352 | .5596 | .5000 | .4519 |
| 6 | | | | | | .8909 | .7705 | .6787 | .6065 | .5481 |
| 7 | | | | | | | .9057 | .7979 | .7129 | .6443 |
| 8 | | | | | | | | .9170 | .8194 | .7406 |
| 9 | | | | | | | | | .9529 | .8368 |
| 10 | | | | | | | | | | .9330 |

Effectif de l'échantillon = n

| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | .0611 | .0561 | .0519 | .0483 | .0452 | .0424 | .0400 | .0378 | .0358 | .0341 |
| 2 | .1489 | .1368 | .1266 | .1178 | .1101 | .1034 | .0975 | .0922 | .0874 | .0831 |
| 3 | .2366 | .2175 | .2013 | .1873 | .1751 | .1644 | .1550 | .1465 | .1390 | .1322 |
| 4 | .3244 | .2982 | .2760 | .2568 | .2401 | .2254 | .2125 | .2009 | .1905 | .1812 |
| 5 | .4122 | .3789 | .3506 | .3263 | .3051 | .2865 | .2700 | .2553 | .2421 | .2302 |
| 6 | .5000 | .4956 | .4253 | .3958 | .3700 | .3475 | .3275 | .3097 | .2937 | .2793 |
| 7 | .5878 | .5404 | .5000 | .4653 | .4350 | .4085 | .3850 | .3641 | .3453 | .3283 |
| 8 | .6756 | .6211 | .5747 | .5347 | .5000 | .4695 | .4425 | .4184 | .3968 | .3774 |
| 9 | .7364 | .7018 | .6494 | .6042 | .5650 | .5305 | .5000 | .4728 | .4484 | .4264 |
| 10 | .8511 | .7825 | .7240 | .6737 | .6300 | .5915 | .5575 | .5272 | .5000 | .4755 |
| 11 | .9389 | .8632 | .7987 | .7432 | .6949 | .6525 | .6150 | .5816 | .5516 | .5245 |
| 12 | | .9439 | .8734 | .8127 | .7599 | .7135 | .6725 | .6359 | .6032 | .5732 |
| 13 | | | .9481 | .8822 | .8249 | .7746 | .7300 | .6903 | .6547 | .6226 |
| 14 | | | | .9517 | .8899 | .8356 | .7875 | .7447 | .7063 | .6717 |
| 15 | | | | | .9548 | .8966 | .8450 | .7991 | .7579 | .7207 |
| 16 | | | | | | .9576 | .9025 | .8535 | .8095 | .7698 |
| 17 | | | | | | | .9600 | .9078 | .8610 | .8188 |
| 18 | | | | | | | | .9622 | .9126 | .8678 |
| 19 | | | | | | | | | .9642 | .9169 |
| 20 | | | | | | | | | | .9659 |

Johnson - Industrial Mathematics - 1951

Table $2_1$

Tests de Pearson (Biometrika Tables)

| Distribution de $\lvert m_3/(m_2)^{3/2} \rvert$ | | | Distribution de $m_4/(m_2)^2$ | | | | |
|---|---|---|---|---|---|---|---|
| | Fractiles | | Fractiles | | | | |
| n | 0,95 | 0,99 | 0,01 | 0,05 | 0,95 | 0,99 | n |
| 25 | 0,71 | 1,06 | 1,95 | 2,15 | 3,99 | 4,88 | 50 |
| 30 | 0,66 | 0,98 | 2,08 | 2,27 | 3,87 | 4,59 | 75 |
| 35 | 0,62 | 0,92 | 2,18 | 2,35 | 3,77 | 4,39 | 100 |
| 40 | 0,59 | 0,87 | 2,24 | 2,40 | 3,71 | 4,24 | 125 |
| 45 | 0,56 | 0,83 | 2,29 | 2,45 | 3,65 | 4,13 | 150 |
| 50 | 0,53 | 0,79 | 2,37 | 2,51 | 3,57 | 3,98 | 200 |
| 60 | 0,49 | 0,72 | 2,42 | 2,55 | 3,52 | 3,87 | 250 |
| 70 | 0,46 | 0,67 | 2,46 | 2,59 | 3,47 | 3,79 | 300 |
| 80 | 0,43 | 0,61 | 2,50 | 2,62 | 3,44 | 3,72 | 350 |
| 90 | 0,41 | 0,60 | 2,52 | 2,64 | 3,41 | 3,67 | 400 |
| 100 | 0,39 | 0,57 | 2,55 | 2,66 | 3,49 | 3,63 | 450 |
| 125 | 0,35 | 0,51 | 2,57 | 2,67 | 3,37 | 3,60 | 500 |
| 150 | 0,32 | 0,46 | 2,58 | 2,69 | 3,35 | 3,57 | 550 |
| 175 | 0,29 | 0,43 | 2,60 | 2,70 | 3,34 | 3,54 | 600 |
| 200 | 0,28 | 0,40 | 2,61 | 2,71 | 3,33 | 3,52 | 650 |
| 300 | 0,23 | 0,33 | 2,62 | 2,72 | 3,31 | 3,50 | 700 |
| 400 | 0,20 | 0,29 | 2,64 | 2,73 | 3,30 | 3,48 | 750 |
| 500 | 0,18 | 0,26 | 2,65 | 2,74 | 3,29 | 3,46 | 800 |
| 600 | 0,16 | 0,23 | 2,66 | 2,74 | 3,28 | 3,45 | 850 |
| 800 | 0,14 | 0,20 | 2,66 | 2,75 | 3,28 | 3,43 | 900 |
| 1000 | 0,13 | 0,18 | 2,67 | 2,76 | 3,27 | 3,42 | 950 |
| | | | 2,68 | 2,76 | 3,26 | 3,41 | 1000 |

Table $2_2$

Distribution de $b_2 = m_4 | (m_2)^2$ pour les petites valeurs de n

Fractiles de la distribution de $b_2$

| n | 0,01 | 0,05 | 0,15 | 0,99 |
|---|---|---|---|---|
| 7 | 1,25 | 1,41 | 3,55 | 4,29 |
| 8 | 1,31 | 1,46 | 3,70 | 4,53 |
| 9 | 1,35 | 1,53 | 3,86 | 4,82 |
| 10 | 1,39 | 1,56 | 3,95 | 5,00 |
| 12 | 1,46 | 1,64 | 4,05 | 5,20 |
| 15 | 1,55 | 1,72 | 4,13 | 5,30 |
| 20 | 1,65 | 1,82 | 4,17 | 5,36 |
| 25 | 1,72 | 1,91 | 4,16 | 5,30 |
| 20 | 1,79 | 1,98 | 4,11 | 5,21 |
| 35 | 1,84 | 2,03 | 4,10 | 5,13 |
| 40 | 1,89 | 2,07 | 4,06 | 5,04 |
| 45 | 1,93 | 2,11 | 4,00 | 4,94 |
| 50 | 1,95 | 2,15 | 3,99 | 4,88 |

Extrait de R. d'Agostino et G. Tietjen - Biometrika 1971 n° 3 p. 670

Table 3

Espérances mathématiques de variables normales ordonnées

$$\xi(i,\ n) = E\left[\frac{x_{(i,n)} - m}{\sigma}\right]$$

H. Hartley and E. Pearson - Biometrika Tables

| i \ n | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.564 | 0.846 | 1.029 | 1.163 | 1.267 | 1.352 | 1.424 | 1.485 | 1.539 | 1.586 | 1.629 |
| 2 | | 0.000 | 0.297 | 0.945 | 0.642 | 0.757 | 0.852 | 0.932 | 1.001 | 1.062 | 1.116 |
| 3 | | | | .000 | .202 | .353 | .473 | .572 | 0.656 | 0.729 | 0.793 |
| 4 | | | | | | .000 | .153 | .275 | .376 | .462 | .537 |
| 5 | | | | | | | 0.000 | | 0.123 | 0.225 | 0.312 |
| 6 | | | | | | | | | | .000 | .103 |

| i \ n | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.668 | 1.703 | 1.736 | 1.766 | 1.794 | 1.820 | 1.844 | 1.867 | 1.89 | 1.91 | 1.93 | 1.95 | 1.97 |
| 2 | 1.164 | 1.208 | 1.248 | 1.285 | 1.319 | 1.350 | 1.380 | 1.408 | 1.43 | 1.46 | 1.48 | 1.50 | 1.52 |
| 3 | 0.850 | 0.901 | 0.948 | 0.990 | 1.029 | 1.066 | 1.099 | 1.131 | 1.16 | 1.19 | 1.21 | 1.24 | 1.26 |
| 4 | .603 | .662 | .715 | .763 | 0.807 | 0.848 | 0.886 | 0.921 | 0.95 | 0.98 | 1.01 | 1.04 | 1.07 |
| 5 | 0.388 | 0.456 | 0.516 | 0.570 | 0.619 | 0.665 | 0.707 | 0.745 | 0.78 | 0.82 | 0.85 | 0.88 | 0.91 |
| 6 | .190 | .267 | .335 | .396 | .451 | .502 | .548 | .590 | .63 | .67 | .70 | .73 | .76 |
| 7 | .000 | .088 | .165 | .234 | .295 | .351 | .402 | .448 | .49 | .53 | .57 | .60 | .64 |
| 8 | | | .000 | .077 | .146 | .208 | .264 | .315 | .36 | .41 | .45 | .48 | .52 |
| 9 | | | | | .000 | .069 | .131 | .187 | .24 | .29 | .33 | .37 | .41 |
| 10 | | | | | | | 0.000 | 0.062 | 0.12 | 0.17 | 0.22 | 0.26 | 0.30 |
| 11 | | | | | | | | | .00 | .06 | .11 | .16 | .20 |
| 12 | | | | | | | | | | | .00 | .05 | .10 |
| 13 | | | | | | | | | | | | | .00 |

| i \ n | 26 | 28 | 30 | 32 | 34 | 36 | 38 | 40 | 42 | 44 | 46 | 48 | 50 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.98 | 2.01 | 2.04 | 2.07 | 2.09 | 2.12 | 2.14 | 2.16 | 2.18 | 2.20 | 2.22 | 2.23 | 2.25 |
| 2 | 1.54 | 1.58 | 1.62 | 1.65 | 1.68 | 1.70 | 1.73 | 1.75 | 1.78 | 1.80 | 1.82 | 1.84 | 1.85 |
| 3 | 1.29 | 1.33 | 1.36 | 1.40 | 1.43 | 1.46 | 1.49 | 1.52 | 1.54 | 1.57 | 1.59 | 1.61 | 1.63 |
| 4 | 1.09 | 1.14 | 1.18 | 1.22 | 1.25 | 1.28 | 1.32 | 1.34 | 1.37 | 1.40 | 1.42 | 1.44 | 1.46 |
| 5 | 0.93 | 0.98 | 1.03 | 1.07 | 1.11 | 1.14 | 1.17 | 1.20 | 1.23 | 1.26 | 1.28 | 1.31 | 1.33 |
| 6 | .79 | .85 | 0.89 | 0.94 | 0.98 | 1.02 | 1.05 | 1.08 | 1.11 | 1.14 | 1.17 | 1.19 | 1.22 |
| 7 | .67 | .73 | .78 | .82 | .87 | 0.91 | 0.94 | 0.98 | 1.01 | 1.04 | 1.07 | 1.09 | 1.12 |
| 8 | .55 | .61 | .67 | .72 | .76 | .81 | .85 | .88 | 0.91 | 0.95 | 0.98 | 1.00 | 1.03 |
| 9 | .44 | .51 | .57 | .62 | .67 | .72 | .75 | .79 | .83 | .86 | .89 | 0.92 | 0.95 |
| 10 | 0.34 | 0.41 | 0.47 | 0.53 | 0.58 | 0.63 | 0.67 | 0.71 | 0.75 | 0.78 | 0.81 | 0.84 | 0.87 |
| 11 | .24 | .32 | .38 | .44 | .50 | .54 | .59 | .63 | .67 | .71 | .74 | .77 | .80 |
| 12 | .14 | .22 | .29 | .36 | .41 | .47 | .51 | .56 | .60 | .64 | .67 | .70 | .74 |
| 13 | .05 | .13 | .21 | .28 | .34 | .39 | .44 | .49 | .53 | .57 | .60 | .64 | .67 |
| 14 | | .04 | .12 | .20 | .26 | .32 | .37 | .42 | .46 | .50 | .54 | .58 | .61 |
| 15 | | | 0.04 | 0.12 | 0.18 | 0.24 | 0.30 | 0.35 | 0.40 | 0.44 | 0.48 | 0.52 | 0.55 |
| 16 | | | | .04 | .11 | .17 | .23 | .28 | .33 | .38 | .42 | .46 | .49 |
| 17 | | | | | .04 | .10 | .16 | .22 | .27 | .32 | .36 | .40 | .44 |
| 18 | | | | | | .03 | .10 | .16 | .21 | .26 | .30 | .34 | .38 |
| 19 | | | | | | | .03 | .09 | .15 | .20 | .25 | .29 | .33 |
| 20 | | | | | | | | 0.03 | 0.09 | 0.14 | 0.19 | 0.24 | 0.28 |
| 21 | | | | | | | | | .03 | .09 | .14 | .18 | .23 |
| 22 | | | | | | | | | | .03 | .08 | .13 | .18 |
| 23 | | | | | | | | | | | .03 | .08 | .13 |
| 24 | | | | | | | | | | | | .03 | .08 |
| 25 | | | | | | | | | | | | | 0.03 |

Table 4

Fractiles de la distribution du rapport de Geary

$$a = \frac{\text{Ecart moyen}}{\text{Ecart-type}} = \frac{\Sigma |x_i - \bar{x}|}{[n\ \Sigma (x_i - \bar{x})^2]^{1/2}}$$

| n | E(a) | 1 % | 5 % | 95 % | 99 % |
|---|---|---|---|---|---|
| 10 | 0,819 | 0,664 | 0,714 | 0,911 | 0,940 |
| 15 | 0,813 | 0,680 | 0,722 | 0,892 | 0,918 |
| 20 | 0,809 | 0,693 | 0,729 | 0,879 | 0,903 |
| 25 | 0,806 | 0,702 | 0,735 | 0,870 | 0,892 |
| 30 | 0,805 | 0,710 | 0,739 | 0,864 | 0,884 |
| 35 | 0,804 | 0,716 | 0,743 | 0,859 | 0,878 |
| 40 | 0,803 | 0,720 | 0,746 | 0,855 | 0,873 |
| 45 | 0,802 | 0,725 | 0,749 | 0,851 | 0,869 |
| 50 | 0,902 | 0,728 | 0,751 | 0,849 | 0,865 |
| 60 | 0,801 | 0,734 | 0,755 | 0,844 | 0,860 |
| 70 | 0,801 | 0,738 | 0,758 | 0,841 | 0,856 |
| 80 | 0,800 | 0,742 | 0,760 | 0,838 | 0,852 |
| 90 | 0,800 | 0,745 | 0,762 | 0,836 | 0,849 |
| 100 | 0,800 | 0,748 | 0,764 | 0,835 | 0,846 |
| 200 | 0,799 | 0,762 | 0,774 | 0,825 | 0,832 |
| 300 | 0,799 | 0,769 | 0,778 | 0,819 | 0,826 |
| 400 | 0,798 | 0,772 | 0,781 | 0,816 | 0,822 |
| 500 | 0,798 | 0,775 | 0,782 | 0,814 | 0,820 |
| 600 | 0,798 | 0,777 | 0,784 | 0,813 | 0,818 |
| 800 | 0,798 | 0,780 | 0,786 | 0,810 | 0,815 |
| 1000 | 0,798 | 0,782 | 0,787 | 0,809 | 0,813 |
| ∞ | 0,7979 | | | | |

(Geary, Biometrika 1936)

Table 5

$$\text{Test de David : } n = \frac{w}{s} = \frac{x_n - x_1}{\left[\frac{1}{n-1}\,\Sigma(x - \bar{x})^2\right]^{1/2}}$$

Fractiles de la distribution de u

| n | Frochles en % | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.5 | 1.0 | 2.5 | 5.0 | 10.0 | 90 | 95 | 97.5 | 99 | 99.5 | n |
| | | | | | | 1.997 | 1.999 | 2.000 | 2.000 | 2.000 | 3 |
| | | | | | | 2.409 | 2.429 | 2.439 | 2.445 | 2.447 | 4 |
| | | | | | | 2.712 | 2.753 | 2.782 | 2.803 | 2.813 | 5 |
| | | | | | | 2.949 | 3.012 | 3.056 | 3.095 | 3.115 | 6 |
| | | | | | | 3.143 | 3.222 | 3.282 | 3.338 | 3.369 | 7 |
| | | | | | | 3.308 | 3.399 | 3.471 | 3.543 | 3.585 | 8 |
| | | | | | | 3.449 | 3.552 | 3.634 | 3.720 | 3.772 | 9 |
| 10 | 2.47 | 2.51 | 2.59 | 2.67 | 2.77 | 3.57 | 3.685 | 3.777 | 3.875 | 3.935 | 10 |
| 11 | 2.53 | 2.58 | 2.66 | 2.74 | 2.84 | 3.68 | 3.80 | 3.903 | 4.012 | 4.079 | 11 |
| 12 | 2.59 | 2.65 | 2.73 | 2.80 | 2.91 | 3.78 | 3.91 | 4.01 | 4.134 | 4.208 | 12 |
| 13 | 2.65 | 2.70 | 2.78 | 2.86 | 2.97 | 3.87 | 4.00 | 4.11 | 4.244 | 4.325 | 13 |
| 14 | 2.70 | 2.75 | 2.83 | 2.91 | 3.02 | 3.95 | 4.09 | 4.21 | 4.34 | 4.431 | 14 |
| 15 | 2.75 | 2.80 | 2.88 | 2.96 | 3.07 | 4.02 | 4.17 | 4.29 | 4.43 | 4.53 | 15 |
| 16 | 2.80 | 2.85 | 2.93 | 3.01 | 3.13 | 4.09 | 4.24 | 4.37 | 4.51 | 4.62 | 16 |
| 17 | 2.84 | 2.90 | 2.98 | 3.06 | 3.17 | 4.15 | 4.31 | 4.44 | 4.59 | 4.69 | 17 |
| 18 | 2.88 | 2.94 | 3.02 | 3.10 | 3.21 | 4.21 | 4.38 | 4.51 | 4.66 | 4.77 | 18 |
| 19 | 2.92 | 2.98 | 3.06 | 3.14 | 3.25 | 4.27 | 4.43 | 4.57 | 4.73 | 4.84 | 19 |
| 20 | 2.95 | 3.01 | 3.10 | 3.18 | 3.29 | 4.32 | 4.49 | 4.63 | 4.79 | 4.91 | 20 |
| 30 | 3.22 | 3.27 | 3.37 | 3.46 | 3.58 | 4.70 | 4.89 | 5.06 | 5.25 | 5.39 | 30 |
| 40 | 3.41 | 3.46 | 3.57 | 3.66 | 3.79 | 4.96 | 5.15 | 5.34 | 5.54 | 5.69 | 40 |
| 50 | 3.57 | 3.61 | 3.72 | 3.82 | 3.94 | 5.15 | 5.35 | 5.54 | 5.77 | 5.91 | 50 |
| 60 | 3.69 | 3.74 | 3.85 | 3.95 | 4.07 | 5.29 | 5.50 | 5.70 | 5.93 | 6.09 | 60 |
| 80 | 3.80 | 3.93 | 4.05 | 4.15 | 4.27 | 5.51 | 5.73 | 5.93 | 6.18 | 6.35 | 80 |
| 100 | 4.02 | 4.09 | 4.20 | 4.31 | 4.44 | 5.68 | 5.90 | 6.11 | 6.36 | 6.54 | 100 |
| 150 | 4.30 | 4.36 | 4.47 | 4.59 | 4.72 | 5.96 | 6.18 | 6.39 | 6.64 | 6.84 | 150 |
| 200 | 4.50 | 4.56 | 4.67 | 4.78 | 4.90 | 6.15 | 6.38 | 6.59 | 6.85 | 7.03 | 200 |
| 500 | 5.06 | 5.13 | 5.25 | 5.37 | 5.49 | 6.72 | 6.94 | 7.15 | 7.42 | 7.60 | 500 |
| 1000 | 5.50 | 5.57 | 5.68 | 5.79 | 5.92 | 7.11 | 7.33 | 7.54 | 7.80 | 7.99 | 1000 |

(David, Hartley et Pearson, Biometrika 1954)

Table $6_a$

Test de Shapiro-Wilk : $W = \dfrac{\left(\sum_1^k a_j\, d_j\right)^2}{\sum_1^n (x_i - \bar{x})^2}$

$$d_j = x_{n-j+1} - x_j \qquad k = \begin{cases} \dfrac{n}{2} \text{ si } n \text{ pair} \\ \dfrac{n-1}{2} \text{ si } n \text{ impair} \end{cases}$$

Table des coefficients $a_i$ (n = 2, ... 50)

| j \ n | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.7071 | 0.7071 | 0.6872 | 0.6646 | 0.6431 | 0.6233 | 0.6052 | 0.5888 | 0.5739 |
| 2 | | .0000 | .1677 | .2413 | .2806 | .3031 | .3164 | .3244 | .3291 |
| 3 | | | | .0000 | .0875 | .1401 | .1743 | .1976 | .2141 |
| 4 | | | | | | .0000 | .0561 | .0947 | .1224 |
| 5 | | | | | | | | .0000 | .0399 |

| j \ n | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.5601 | 0.5475 | 0.5359 | 0.5251 | 0.5150 | 0.5056 | 0.4968 | 0.4486 | 0.4808 | 0.4734 |
| 2 | .3315 | .3325 | .3325 | .3318 | .3306 | .3290 | .3273 | .3253 | .3232 | .3211 |
| 3 | .2260 | .2347 | .2412 | .2460 | .2495 | .2521 | .2540 | .2533 | .2561 | .2565 |
| 4 | .1429 | .1586 | .1707 | .1802 | .1878 | .1939 | .1988 | .2027 | .2059 | .2085 |
| 5 | .0695 | .0922 | .1099 | .1240 | .1353 | .1447 | .1524 | .1587 | .1641 | .1686 |
| 6 | 0.0000 | 0.0303 | 0.0539 | 0.0727 | 0.0880 | 0.1005 | 0.1109 | 0.1197 | 0.1271 | 0.1334 |
| 7 | | | .0000 | .0240 | .0433 | .0593 | .0725 | .0837 | .0932 | .1013 |
| 8 | | | | | .0000 | .0196 | .0359 | .0496 | .0612 | .0711 |
| 9 | | | | | | | .0000 | .0613 | .0303 | .0422 |
| 10 | | | | | | | | | .0000 | .1040 |

| j \ n | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.4643 | 0.4390 | 0.4542 | 0.4493 | 0.4450 | 0.4407 | 0.4366 | 0.4328 | 0.4291 | 0.4254 |
| 2 | .3815 | .3156 | .3126 | .3098 | .3069 | .3043 | .3018 | .2992 | .2968 | .2944 |
| 3 | .2578 | .2571 | .2563 | .2554 | .2543 | .2533 | .2522 | .2510 | .2499 | .2487 |
| 4 | .2199 | .2131 | .2139 | .2145 | .2148 | .2151 | .2152 | .2151 | .2150 | .2148 |
| 5 | .1736 | .1764 | .1787 | .1807 | .1822 | .1836 | .1848 | .1857 | .1864 | .1870 |
| 6 | 0.1399 | 0.1443 | 0.1480 | 0.1512 | 0.1539 | 0.1563 | 0.1584 | 0.1601 | 0.1616 | 0.1630 |
| 7 | .1092 | .1150 | .1201 | .1245 | .1283 | .1316 | .1346 | .1372 | .1395 | .1415 |
| 8 | .0804 | .0878 | .0941 | .0997 | .1046 | .1089 | .1128 | .1162 | .1192 | .1219 |
| 9 | .0530 | .0618 | .0696 | .0764 | .0823 | .0876 | .0923 | .0965 | .1002 | .1036 |
| 10 | .0263 | .0368 | .0459 | .0539 | .0610 | .0672 | .0728 | .0778 | .0822 | .0862 |
| 11 | 0.0000 | 0.0122 | 0.0228 | 0.0321 | 0.0403 | 0.0476 | 0.0540 | 0.0598 | 0.0650 | 0.0697 |
| 12 | | | .0000 | .0107 | .0200 | 0.284 | .0358 | .0424 | .0483 | .0537 |
| 13 | | | | | .0000 | .0094 | .0178 | .0253 | .0320 | .0381 |
| 14 | | | | | | | .0000 | .0084 | .0159 | .0227 |
| 15 | | | | | | | | | .0000 | .0076 |

## Table $6_a$ (suite)

| j \ n | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.4420 | 0.4188 | 0.4156 | 0.4127 | 0.4096 | 0.4068 | 0.4040 | 0.4015 | 0.3989 | 0.3964 |
| 2 | .2021 | .2898 | .2876 | .2854 | .2834 | .2813 | .2794 | .2774 | .2755 | .2737 |
| 3 | .2475 | .2463 | .2451 | .2439 | .2427 | .2415 | .2403 | .2391 | .2380 | .2368 |
| 4 | .2145 | .2141 | .2137 | .2132 | .2127 | .2121 | .2116 | .2110 | .2104 | .2098 |
| 5 | .1874 | .1878 | .1880 | .1882 | .1883 | .1883 | .1883 | .1881 | .1880 | .1878 |
| 6 | 0.1641 | 0.1651 | 0.1660 | 0.1667 | 0.1673 | 0.1678 | 0.1683 | 0.1686 | 0.1689 | 0.1691 |
| 7 | .1433 | .1449 | .1463 | .1475 | .1487 | .1496 | .1505 | .1513 | .1520 | .1526 |
| 8 | .1243 | .1265 | .1284 | .1301 | .1317 | .1331 | .1344 | .1356 | .1366 | .1376 |
| 9 | .1066 | .1093 | .1118 | .1140 | .1160 | .1170 | .1196 | .1211 | .1225 | .1237 |
| 10 | .0899 | .0931 | .0961 | .0988 | .1013 | .1036 | .1056 | .1075 | .1092 | .1108 |
| 11 | 0.0739 | 0.0777 | 0.0812 | 0.0844 | 0.0873 | 0.0900 | 0.0924 | 0.0947 | 0.0967 | 0.0986 |
| 12 | .0585 | .0629 | .0669 | .0706 | .0739 | .0770 | .0798 | .0824 | .0848 | .0870 |
| 13 | .0435 | .0485 | .0530 | .0572 | .0610 | .0645 | .0677 | .0706 | .0733 | .0759 |
| 14 | .0280 | .0344 | .0395 | .0441 | .0484 | .0523 | .0559 | .0592 | .0622 | .0651 |
| 15 | .0144 | .0206 | .0262 | .0314 | .0361 | .0404 | .0444 | .0481 | .0515 | 0.546 |
| 16 | 0.0000 | 0.0068 | 0.0131 | 0.0187 | 0.0239 | 0.0287 | 0.0331 | 0.0372 | 0.0409 | 0.0444 |
| 17 | | | .0000 | .0062 | .0119 | .0172 | .0220 | .0264 | .0305 | .0343 |
| 18 | | | | | .0000 | .0057 | .0110 | .0158 | .0203 | .0244 |
| 19 | | | | | | | .0000 | .0053 | .0101 | .0146 |
| 20 | | | | | | | | | .0000 | .0049 |

| j \ n | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.3940 | 0.3917 | 0.3894 | 0.3872 | 0.3850 | 0.3830 | 0.3808 | 0.3789 | 0.3770 | 0.3751 |
| 2 | .2719 | .2701 | .2864 | .2667 | .2651 | .2635 | .2620 | .2604 | .2589 | .2574 |
| 3 | .2357 | .2345 | .2334 | .2323 | .2313 | .2302 | .2291 | .2281 | .2271 | .2260 |
| 4 | .2091 | .2085 | .2078 | .2072 | .2065 | .2058 | .2052 | .2045 | .2038 | .2032 |
| 5 | .1876 | .1874 | .1871 | .1808 | .1865 | .1862 | .1859 | .1855 | .1851 | .1847 |
| 6 | 0.1693 | 0.1694 | 0.1695 | 0.1695 | 0.1695 | 0.1695 | 0.1695 | 0.1693 | 0.1692 | 0.1691 |
| 7 | .1531 | .1535 | .1539 | .1542 | .1545 | .1548 | .1550 | .1551 | .1553 | .1554 |
| 8 | .1384 | .1392 | .1398 | .1405 | .1410 | .1415 | .1420 | .1423 | .1427 | .1430 |
| 9 | .1249 | .1259 | .1269 | .1278 | .1286 | .1293 | .1300 | .1306 | .1312 | .1317 |
| 10 | .1123 | .1136 | .1149 | .1160 | .1170 | .1180 | .1189 | .1197 | .1205 | .1212 |
| 11 | 0.1004 | 0.1020 | 0.1035 | 0.1049 | 0.1062 | 0.1073 | 0.1085 | 0.1093 | 0.1105 | 0.1113 |
| 12 | .0891 | .0909 | .0927 | .0943 | .0959 | .0972 | .0986 | .0998 | .1010 | .1020 |
| 13 | .0782 | .0804 | .0824 | .0842 | .0860 | .0876 | .0892 | .0906 | .0919 | .0932 |
| 14 | .0677 | .0701 | .0724 | .0745 | .0765 | .0783 | .0801 | .0817 | .0832 | .0846 |
| 15 | .0575 | .0602 | .0628 | .0651 | .0673 | .0694 | .0713 | .0731 | .0748 | .0764 |
| 16 | 0.0476 | 0.0506 | 0.0534 | 0.0560 | 0.0584 | 0.0607 | 0.0628 | 0.0648 | 0.0667 | 0.0685 |
| 17 | .0379 | .0411 | .0442 | .0471 | .0497 | .0522 | .0546 | .0568 | .0588 | .0608 |
| 18 | .0283 | .0318 | .0352 | .0383 | .0412 | .0439 | .0465 | .0489 | .0511 | .0532 |
| 19 | .0188 | .0227 | .0263 | .0296 | .0328 | .0357 | .0385 | .0411 | .0436 | .0459 |
| 20 | .0094 | .0136 | .0175 | .0211 | .0245 | .0277 | .0307 | .0335 | .0361 | .0386 |
| 21 | 0.0000 | 0.0045 | 0.0087 | 0.0126 | 0.0163 | 0.0197 | 0.0229 | 0.0259 | 0.0288 | 0.0314 |
| 22 | | | .0000 | .0042 | .0081 | .0118 | .0153 | .0185 | .0215 | .0244 |
| 23 | | | | | .0000 | .0039 | .0076 | .0111 | .0143 | .0174 |
| 24 | | | | | | | .0000 | .0037 | .0071 | .0104 |
| 25 | | | | | | | | | .0000 | .0035 |

Table $6_b$

Test de Shapiro-Wilk

Fractiles de la distribution de W

| n | 0.01 | 0.02 | 0.05 | 0.10 | 0.50 | 0.90 | 0.95 | 0.98 | 0.99 |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 0.753 | 0.756 | 0.767 | 0.789 | 0.959 | 0.998 | 0.999 | 1.000 | 1.000 |
| 4 | .687 | .707 | .748 | .792 | .935 | .987 | .992 | .996 | .997 |
| 5 | .686 | .715 | .762 | .806 | .927 | .979 | .986 | .991 | .993 |
| 6 | 0.713 | 0.743 | 0.788 | 0.826 | 0.927 | 0.974 | 0.981 | 0.986 | 0.989 |
| 7 | .730 | .760 | .803 | .838 | .928 | .972 | .979 | .985 | .988 |
| 8 | .749 | .778 | .818 | .851 | .932 | .972 | .978 | .984 | .987 |
| 9 | .764 | .791 | .829 | .859 | .935 | .972 | .978 | .984 | .986 |
| 10 | .781 | .806 | .842 | .869 | .938 | .972 | .978 | .983 | .986 |
| 11 | 0.792 | 0.817 | 0.850 | 0.876 | 0.940 | 0.973 | 0.979 | 0.984 | 0.986 |
| 12 | .803 | .828 | .859 | .883 | .943 | .973 | .979 | .984 | .986 |
| 13 | .814 | .837 | .866 | .889 | .945 | .974 | .979 | .984 | .986 |
| 14 | .825 | .846 | .874 | .895 | .947 | .975 | .980 | .984 | .986 |
| 15 | .835 | .855 | .881 | .901 | .950 | .975 | .980 | .984 | .987 |
| 16 | 0.844 | 0.863 | 0.887 | 0.906 | 0.952 | 0.976 | 0.981 | 0.985 | 0.987 |
| 17 | .851 | .869 | .892 | .910 | .954 | .977 | .981 | .985 | .987 |
| 18 | .858 | .874 | .897 | .914 | .956 | .978 | .982 | .986 | .988 |
| 19 | .863 | .879 | .901 | .917 | .957 | .978 | .982 | .986 | .988 |
| 20 | .868 | .884 | .905 | .920 | .959 | .979 | .983 | .986 | .988 |
| 21 | 0.873 | 0.888 | 0.908 | 0.923 | 0.960 | 0.980 | 0.983 | 0.987 | 0.989 |
| 22 | .878 | .892 | .911 | .926 | .961 | .980 | .984 | .987 | .989 |
| 23 | .881 | .895 | .914 | .928 | .962 | .981 | .984 | .987 | .989 |
| 24 | .884 | .898 | .916 | .930 | .963 | .981 | .984 | .987 | .989 |
| 25 | .888 | .901 | .918 | .931 | .964 | .981 | .985 | .988 | .989 |
| 26 | 0.891 | 0.904 | 0.920 | 0.933 | 0.965 | 0.982 | 0.985 | 0.988 | 0.989 |
| 27 | .894 | .906 | .923 | .935 | .965 | .982 | .985 | .988 | .990 |
| 28 | .896 | .908 | .924 | .936 | .966 | .982 | .985 | .988 | .990 |
| 29 | .898 | .910 | .926 | .937 | .966 | .982 | .985 | .988 | .990 |
| 30 | .900 | .912 | .927 | .939 | .967 | .983 | .985 | .988 | .900 |
| 31 | 0.902 | 0.914 | 0.929 | 0.940 | 0.967 | 0.983 | 0.986 | 0.988 | 0.990 |
| 32 | .904 | .915 | .930 | .941 | .968 | .983 | .986 | .988 | .990 |
| 33 | .906 | .917 | .931 | .942 | .968 | .983 | .986 | .989 | .990 |
| 34 | .908 | .919 | .933 | .943 | .969 | .983 | .986 | .989 | .990 |
| 35 | .910 | .920 | .934 | .944 | .969 | .984 | .986 | .989 | .990 |
| 36 | 0.912 | 0.922 | 0.935 | 0.945 | 0.970 | 0.984 | 0.986 | 0.989 | 0.990 |
| 37 | .914 | .924 | .936 | .946 | .970 | .984 | .987 | .989 | .990 |
| 38 | .916 | .925 | .938 | .947 | .971 | .984 | .987 | .989 | .990 |
| 39 | .917 | .927 | .939 | .948 | .971 | .984 | .987 | .989 | .991 |
| 40 | .919 | .928 | .940 | .949 | .972 | .985 | .987 | .989 | .991 |
| 41 | 0.920 | 0.929 | 0.941 | 0.950 | 0.972 | 0.985 | 0.987 | 0.989 | 0.991 |
| 42 | .922 | .930 | .942 | .951 | .972 | .985 | .987 | .989 | .991 |
| 43 | .923 | .932 | .943 | .951 | .973 | .985 | .987 | .990 | .991 |
| 44 | .924 | .933 | .944 | .952 | .973 | .985 | .987 | .990 | .991 |
| 45 | .926 | .934 | .945 | .953 | .973 | .985 | .988 | .990 | .991 |
| 46 | 0.927 | 0.935 | 0.945 | 0.953 | 0.974 | 0.985 | 0.988 | 0.990 | 0.991 |
| 47 | .928 | .936 | .946 | .954 | .974 | .985 | .988 | .990 | .991 |
| 48 | .929 | .937 | .947 | .954 | .974 | .985 | .988 | .990 | .991 |
| 49 | .929 | .937 | .947 | .955 | .974 | .985 | .988 | .990 | .991 |
| 50 | .930 | .938 | .947 | .955 | .974 | .985 | .988 | .990 | .991 |

Extrait de Shapiro and Wilk - Biometrika - 1965.

Table 7

Tests de Lilliefors et de Kuiper

| Niveaux de signification | | | | | |
|---|---|---|---|---|---|
| | Test de Lilliefors | | Test de Kuiper | | |
| n | 1 % | 5 % | 1 % | 5 % | n |
| 6 | 0,364 | 0,319 | 0,606 | 0,535 | 6 |
| 7 | 0,348 | 0,300 | 0,567 | 0,503 | 7 |
| 8 | 0,331 | 0,285 | 0,535 | 0,476 | 8 |
| 9 | 0,311 | 0,271 | 0,508 | 0,450 | 9 |
| 10 | 0,294 | 0,258 | 0,484 | 0,429 | 10 |
| 11 | 0,284 | 0,249 | 0,468 | 0,412 | 11 |
| 12 | 0,275 | 0,242 | 0,449 | 0,395 | 12 |
| 13 | 0,268 | 0,234 | 0,436 | 0,385 | 13 |
| 14 | 0,261 | 0,227 | 0,423 | 0,373 | 14 |
| 15 | 0,257 | 0,220 | 0,410 | 0,360 | 15 |
| 16 | 0,250 | 0,213 | 0,397 | 0,348 | 16 |
| 17 | 0,245 | 0,206 | 0,387 | 0,341 | 17 |
| 18 | 0,239 | 0,200 | 0,377 | 0,331 | 18 |
| 19 | 0,235 | 0,195 | 0,368 | 0,324 | 19 |
| 20 | 0,231 | 0,190 | 0,359 | 0,314 | 20 |
| 22 | 0,220 | 0,186 | 0,345 | 0,304 | 22 |
| 25 | 0,203 | 0,180 | 0,326 | 0,285 | 25 |
| 30 | 0,187 | 0,161 | 0,301 | 0,263 | 30 |
| > 30 | $\frac{1,031}{\sqrt{n}}$ | $\frac{0,886}{\sqrt{n}}$ | $\frac{1,647}{\sqrt{n}}$ | $\frac{1,450}{\sqrt{n}}$ | > 30 |

(Lilliefors J.A.SA 1967) (Louter et Koerts Stat. Neerl. 1970)

Table 8

Test de R.d'. Agostino

Fractiles de la distribution de Y

| | 0,01 | 0,05 | 0,95 | 0,99 |
|---|---|---|---|---|
| 50 | - 3,44 | - 2,22 | 0,92 | 1,14 |
| 60 | - 3,36 | - 2,18 | 0,99 | 1,24 |
| 70 | - 3,29 | - 2,15 | 1,04 | 1,31 |
| 80 | - 3,24 | - 2,12 | 1,08 | 1,37 |
| 90 | - 3,10 | - 2,09 | 1,11 | 1,43 |
| 100 | - 3,15 | - 2,07 | 1,14 | 1,47 |
| 150 | - 3,01 | - 2,00 | 1,23 | 1,62 |
| 200 | - 2,92 | - 1,16 | 1,29 | 1,71 |
| 500 | - 2,71 | - 1,85 | 1,42 | 1,94 |
| 1000 | - 2,60 | - 1,79 | 1,49 | 2,05 |

Extrait de R. d'Agostino, Biometrika 1971 n° 2 p. 341