REVUE DE STATISTIQUE APPLIQUÉE

A. VESSEREAU

Essais interlaboratoires pour l'estimation de la fidélité des méthodes d'essais

Revue de statistique appliquée, tome 22, nº 1 (1974), p. 5-48
<http://www.numdam.org/item?id=RSA_1974__22_1_5_0>

© Société française de statistique, 1974, 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/

ESSAIS INTERLABORATOIRES POUR L'ESTIMATION DE LA FIDÉLITÉ DES MÉTHODES D'ESSAIS

A. VESSEREAU

TABLE DES MATIERES

Pages

1. GENERALITES 7

1.1 Définition de la répétabilité et de la reproductibilité 7
1.2 Essai interlaboratoire pour la détermination de la répétabilité et de la reproductibilité 8
1.3 Interprétation statistique 9

2. CAS D'UN SEUL NIVEAU

2.1 Hypothèses et Modèle 10
2.2 Nombre de laboratoires et nombre de répétitions 10
2.3 Différentes modalités de l'essai – Notations 11

2.3.1 Expérience comportant deux répétitions sur échantillons identiques 11
2.3.2 Variante de l'expérience précédente (deux niveaux A et B légèrement différents) 12
2.3.3 Cas général 12

2.4 Test de l'hypothèse de normalité de l'erreur intralaboratoire. 13

2.4.1 Méthode graphique 13
2.4.2 Test de Shapiro et Wilk 13

2.5 Test de l'hypothèse d'homogénéité des variances intralaboratoires 13

2.5.1 Test de Hartley 14
2.5.2 Test de Cochran 14
2.5.3 Test de Bartlett 15
2.5.4 Décision à prendre à la suite de ces tests 16

2.6 Test de l'hypothèse de normalité de l'erreur totale 16
2.7 Estimation des variances intra et interlaboratoires et de la variance totale 17

Pages

2.7.1 Expérience comportant deux répétitions sur échantillons identiques. 17
2.7.2 Variante de l'expérience précédente (deux niveaux A et B légèrement différents). 18
2.7.3 Cas général. 18
2.8 Estimation de la répétabilité et de la reproductiblité. . . . 19
2.9 Exemple. 20

3. CAS DE PLUSIEURS NIVEAUX LORSQUE REPETABILITE ET REPRODUCTIBILITE SONT INDEPENDANTES DU NIVEAU. . .
3.1 Hypothèses et modèle. 25
3.2 Nombre de niveaux, de laboratoires, et de répétitions. 26
3.3 Différentes modalités de l'essai – Notations. 26
3.4 Test des hypothèses à chaque niveau. 27
3.5 Test des hypothèses selon lesquelles répétabilité et reproductibilité sont indépendantes du niveau. 28
3.5.1 Test concernant la répétabilité. 28
3.5.2 Test concernant la reproductibilité. 28
3.5.3 Décision à prendre à la suite de ces tests. 28
3.6 Estimation des variances intra et interlaboratoires et de la variance totale valable pour tous les niveaux. 29
3.6.1 Expérience comportant, à chaque niveau, deux répétitions sur échantillons identiques. 29
3.6.2 Variante de l'expérience précédente (deux niveaux légèrement différents à chaque niveau moyen). 29
3.6.3 Cas général. 29
3.7 Estimations de la répétabilité et de la reproductibilité valables pour tous les niveaux. 30
3.8 Exemple. 32

4. CAS GENERAL
4.1 Analyse globale. 36
4.2 Analyse par niveau. 37
4.3 Exemple. 37

5. QUELQUES CAS D'UTILISATION PRATIQUE DE LA REPETABILITE ET DE LA REPRODUCTIBILITE. . . . . .
5.1 Résultats obtenus dans un seul laboratoire. 43
5.1.1 Comparaison de deux ou plusieurs échantillons. 43
5.1.2 Intervalle de confiance attaché à une moyenne de résultats. 43

Pages

5.2 Résultats obtenus dans plusieurs laboratoires. 44

5.2.1 Accord entre deux ou plusieurs résultats. 44

5.2.2 Intervalle de confiance attaché à une moyenne de résultats. 44

TABLES STATISTIQUES

Table 1. – Table pour le test de Hartley. 45

Table 2. – Table pour le test de Cochran. 46

Table 3. – Table pour le test de Bartlett. 47

REFERENCES

1 – GENERALITES

1.1 – **Définition de la répétabilité et de la reproductibilité**

L'erreur à caractère aléatoire qui affecte les résultats obtenus en utilisant une méthode d'essai aussi parfaitement définie que possible caractérise la fidélité de cette méthode. Cette erreur peut être envisagée sous deux aspects qui portent le nom de "répétabilité" et de "reproductibilité".

Après de nombreuses discussions, les experts de la Commission des Méthodes Statistiques de l'Organisation Internationale de Normalisation (I.S.O.) se sont mis (provisoirement) d'accord sur les définitions ci-après.

Fidélité

Etroitesse de l'accord entre les résultats obtenus en appliquant le procédé expérimental à plusieurs reprises dans des conditions déterminées.

Le procédé est d'autant plus fidèle que la partie aléatoire des erreurs expérimentales qui affectent les résultats est moindre.

Répétabilité

Qualitativement

Etroitesse de l'accord entre les résultats successifs obtenus avec la même méthode sur une matière identique soumise à l'essai dans les mêmes conditions (même opérateur, même appareil, même laboratoire et court intervalle de temps).

Note : Les paramètres représentatifs de la dispersion de la population qui peut être associée aux résultats sont qualifiés du terme "répétabilité". Exemple : "écart-type de répétabilité", "variance de répétabilité".

Quantitativement

La répétabilité s'exprime par la valeur au-dessous de laquelle est située, avec une probabilité spécifiée, la valeur absolue de la différence entre deux

résultats individuels obtenus dans les conditions ci-dessus. En l'absence d'indication, le niveau de probabilité est de 95 %.

Reproductibilité

Qualitativement

Etroitesse de l'accord entre les résultats individuels obtenus avec la même méthode sur une matière identique soumise à l'essai mais dans des conditions différentes (opérateurs différents, appareils différents, laboratoires différents et/ou époques différentes).

Note : Les paramètres représentatifs de la dispersion de la population qui peut être associée aux résultats sont qualifiés du terme "reproductibilité". Exemple : "écart-type de reproductibilité", "variance de reproductibilité".

Quantitativement

La reproductibilité s'exprime par la valeur au-dessous de laquelle est située, avec une probabilité spécifiée, la valeur absolue de la différence entre deux résultats individuels obtenus dans les conditions ci-dessus. En l'absence d'indication, le niveau de probabilité est de 95 %.

Il résulte de ces définitions que si σ_r et σ_R sont les écarts-types de répétabilité et de reproductibilité, et, dans la mesure où les distributions auxquelles s'appliquent ces écarts-types sont normales ou approximativement normales, les expressions "quantitatives" de la répétabilité (r) et de la reproductibilité (R) sont respectivement :

$$r = 1{,}96\sqrt{2}\ \sigma_r \ \# \ 2{,}8\ \sigma_r$$

$$R = 1{,}96\sqrt{2}\ \sigma_R \ \# \ 2{,}8\ \sigma_R$$

1.2 – Essai interlaboratoires pour la détermination de la répétabilité et de la reproductibilité

La répétabilité et la reproductibilité se déterminent au moyen d'un essai interlaboratoires.

Des échantillons identiques ou aussi identiques que possible d'un produit –celui-ci étant éventuellement représenté à plusieurs "niveaux"– sont adressés, aux fins de détermination de la grandeur intéressée aux laboratoires participant à l'essai. Chaque laboratoire effectue, à chaque niveau, un même nombre de déterminations, ou "répétitions" (souvent, ce nombre est réduit au minimum, à savoir n = 2).

Il est indispensable que les conditions de l'essai fassent l'objet d'un protocole très détaillé ; celui-ci est souvent établi à la suite d'un essai préliminaire, portant sur un nombre réduit de laboratoires, et généralement appelé "essai pilote". L'objet du présent article étant essentiellement l'interprétation statistique des résultats, nous n'insisterons pas sur cet aspect –cependant très important– des essais interlaboratoires. Le lecteur pourra trouver des détails à ce sujet dans le chapitre 1 de la référence [1].

1.3 – Interprétation statistique

A l'occasion des réunions des experts de l'I.S.O. dont il a été parlé plus haut, plusieurs techniques ont été présentées pour l'interprétation statistique des essais interlaboratoires. Celles-ci sont décrites dans différents documents ou articles, notamment :

[1] AFNOR – NF X 06-041. – Fidélité d'une méthode d'essai. Guide pour les essais interlaboratoires (1970).

[2] ASTM STP 335. – Manual for conducting an Interlaboratory Study of a Test Method (1963).

[3] The INSTITUTE OF PETROLEUM. – Recommended Practice for determining precision data for methods on petroleum products and lubricants. Londres 1958.

[4] John MANDEL. – Repeatability and Reproducibility (comporte 7 références). Journal of Quality Technology Vol 4, n° 2 (1972).

En fait, nous avons pu constater, sur plusieurs exemples, que les méthodes les plus "sophistiquées", donnent, dans la pratique, des estimations de la répétabilité et de la reproductibilité très voisines de celles obtenues par les méthodes les plus simples. Ce sont celles-ci qui sont présentées dans les chapitres ci-après. Elles présentent selon nous, par rapport aux méthodes théoriquement plus élaborées, les avantages suivants :

– Elles évitent l'estimation des données manquantes, ou rejetées comme aberrantes – estimation présentant un caractère hasardeux,

– Dans le cas de plusieurs niveaux, la répétabilité et la reproductibilité dépendant du niveau, elles ne font pas appel à un changement de variable, dont la justification technique est souvent loin d'être évidente, et qui ne peut donner qu'une "stabilisation" approximative des variances de répétabilité et de reproductibilité,

– Elles n'introduisent pas la notion de "variance d'interaction entre laboratoires et niveaux", qui ne présente pas d'utilité lorsque le but recherché est uniquement l'estimation de la répétabilité et de la reproductibilité.

Dans les chapitres suivants, on exposera successivement les cas ci-après :

– Essais portant sur un seul niveau. Cette situation est assez exceptionnelle, mais elle permet d'exposer en détail les différents tests préalables à l'estimation proprement dite de la répétabilité et de la reproductibilité.

– Essais portant sur plusieurs niveaux, la répétabilité et la reproductibilité étant indépendantes du niveau. On peut dans ce cas rassembler dans une analyse globale les analyses faites à chaque niveau.

– Cas général où répétabilité et reproductibilité dépendent du niveau.

Dans un dernier chapitre on traitera de quelques cas d'utilisation pratique de la répétabilité et de la reproductibilité.

2 - CAS D'UN SEUL NIVEAU

2.1 – Hypothèses et modèle

Les hypothèses sur lesquelles repose l'estimation de la répétabilité et de la reproductibilité sont essentiellement les suivantes.

L'erreur intralaboratoire qui affecte les résultats obtenus dans les conditions de répétabilité est distribuée suivant une loi normale ou approximativement normale de moyenne égale à 0 ; ces erreurs sont sans corrélation entre elles.

La variance de l'erreur intralaboratoire est la même dans tous les laboratoires.

Les différents laboratoires susceptibles d'appliquer la méthode d'essai introduisent un autre terme d'erreur indépendant du précédent, (erreur interlaboratoires) qui obéit également à une loi normale ou approximativement normale.

Il en résulte que l'erreur totale (somme des erreurs intra et interlaboratoires) est distribuée suivant une loi normale ou approximativement normale de moyenne égale à 0.

Le modèle correspondant à ces hypothèses s'écrit :

$$y_{ik} = m + b_i + e_{ik}$$

y_{ik} = l'un des résultats obtenus par le laboratoire (i) dans les conditions de répétabilité

e_{ik} = erreur aléatoire affectant ce résultat.

D'après les hypothèses faites :

e_{ik} appartient à une distribution normale ou approximativement normale de moyenne 0 et de variance σ_r^2 (variance intralaboratoire, ou de répétabilité)

$$E(e_{ik}) = 0 \qquad E(e_{ik}^2) = \sigma_r^2 \quad \text{quel que soit i.}$$

b_i appartient à une distribution normale ou approximativement normale de moyenne 0 et de variance σ_L^2 (variance interlaboratoires).

$$E(b_i) = 0 \qquad E(b_i^2) = \sigma_L^2$$

La variance totale, ou variance de reproductibilité, est

$$\sigma_R^2 = \sigma_r^2 + \sigma_L^2$$

2.2 – Nombre de laboratoires et nombre de répétitions

Les valeurs numériques de la répétabilité et de la reproductibilité obtenues à la suite d'un essai interlaboratoires sont des *estimations* des vraies valeurs de ces grandeurs.

La précision de la reproductibilité estimée dépend du nombre de laboratoires participant à l'essai. Il y a intérêt à ce que ce nombre soit aussi élevé que possible ; selon nous il ne devrait pas être inférieur à 12 – il devrait si possible être de 20 ou plus. S'il y a moins de 12 laboratoires, l'estimation de la reproductibilité devient trop imprécise pour avoir une valeur pratique acceptable.

La précision de la répétabilité estimée dépend à la fois du nombre de laboratoires et du nombre de répétitions effectuées dans les différents laboratoires.

Le nombre de laboratoires étant désigné par p, les conditions ci-dessus : $p \geqslant 12$, et si possible $p \geqslant 20$ permettent d'obtenir pour la répétabilité une estimation dont la valeur pratique est acceptable, même lorsque le nombre de répétitions est réduit au minimum ($n = 2$).

2.3 – Différentes modalités de l'essai – Notations

Lorsque les résultats ont été fournis, il convient de les examiner soigneusement afin d'éliminer ceux qui sont manifestement aberrants (résultats isolés, ou totalité des résultats d'un laboratoire). Lorsqu'à la suite de cette opération (ou parce que certaines déterminations n'ont pas été faites) certains laboratoires ne sont représentés que par un seul résultat, il est préférable de les éliminer (on aura donc toujours $n \geqslant 2$).

On procède éventuellement à un changement d'origine et à un changement d'unité permettant de simplifier les calculs.

2.3.1 *Expérience comportant deux répétitions sur échantillons identiques*

Le cas le plus simple est celui où chacun des p laboratoires effectue deux déterminations, les 2 p échantillons étant identiques ou aussi identiques que possible.

Soient y_{i1} et y_{i2} les résultats obtenus par le laboratoire L_i ($i = 1, 2 \ldots p$). On pose :

$$w_i = |y_{i1} - y_{i2}| \qquad T_i = y_{i1} + y_{i2} \qquad \bar{y}_i = \frac{T_i}{2}$$

$$T = \sum_{i=1}^{p} T_i \qquad N = 2\,p \qquad \bar{y} = \frac{T}{N}$$

L'estimation de la variance intralaboratoire dans le laboratoire L_i est :

$$s_i^2 = \frac{w_i^2}{2}$$

Elle est basée sur $\nu = 1$ degré de liberté.

2.3.2 *Variante de l'expérience précédente* (deux niveaux A et B légèrement différents)

Dans certains cas on préfère confier à chaque laboratoire deux échantillons correspondant à deux niveaux légèrement différents de la grandeur mesurée, respectivement A et B (les mêmes pour tous les laboratoires). Cette façon d'opérer a pour avantage que l'expérimentateur ne peut pas être influencé, lorsqu'il effectue l'essai sur le deuxième échantillon, par le résultat obtenu sur le premier. Elle a pour inconvénient qu'il est impossible d'estimer la variance intralaboratoire de chaque laboratoire, de sorte que les *tests cités au § 2.5 ne peuvent être appliqués.* L'hypothèse d'homogénéité des variances intralaboratoires doit être admise, ou avoir été testée lors d'un essai préalable portant sur l'ensemble des laboratoires.

Soient y_{iA} et y_{iB} les résultats obtenus aux niveaux A et B par le laboratoire L_i. On pose :

$$w_i = | y_{iA} - y_{iB} | \qquad W = \sum_{i=1}^{p} w_i \qquad T_i = y_{iA} + y_{iB} \qquad \bar{y}_i = \frac{T_i}{2}$$

$$T = \sum_{i=1}^{p} T_i \qquad N = 2\,p \qquad \bar{y} = \frac{T}{N}$$

2.3.3 – *Cas général*

Le cas général est celui où, pour le laboratoire $L_i (i = 1, 2, \ldots, p)$ on dispose de n_i résultats y_{ik} $(k = 1, 2, \ldots, n_i)$, l'ensemble des $N = \sum_{i=1}^{n_i}$ résultats provenant d'échantillons identiques ou aussi identiques que possible.

On a ici :

$$T_i = \sum_{k=1}^{n_i} y_{ik} \qquad \bar{y}_i = \frac{T_i}{n_i} \qquad T = \sum_{i=1}^{p} T_i \qquad \bar{y} = \frac{T}{N}$$

L'estimation de la variance intralaboratoire pour le laboratoire L_i est donnée par la formule :

$$s_i^2 = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (y_{ik} - \bar{y}_i)^2 = \frac{1}{n_i - 1} \left[\sum_{k=1}^{n_i} y_{ik}^2 - \frac{T_i^2}{n_i} \right]$$

Elle est basée sur $\nu_i = n_i - 1$ degré de liberté.

Dans le cas particulier où $n_i = n = 2$, $s_i^2 = w_i^2/2$ (avec un seul degré de liberté), w_i étant la différence entre les deux résultats (étendue).

2.4 – Test de l'hypothèse de normalité de l'erreur intralaboratoire

Il s'agit de l'hypothèse selon laquelle la distribution de l'erreur intralaboratoire est normale ou approximativement normale.

Le test de cette hypothèse exige que l'on dispose d'au moins 10 résultats obtenus dans les conditions de répétabilité, et si possible de 20 ou plus. Il doit être effectué antérieurement à l'essai interlaboratoires, par exemple à l'occasion d'un essai pilote.

Parmi les nombreuses méthodes qui sont décrites dans la référence [5] :

E. MORICE. – Tests de normalité d'une distribution observée Revue de Statistique Appliquée, Vol. XX N° 2 (1972).

Les méthodes suivantes sont particulièrement recommandées.

2.4.1 – *Méthode graphique*

On utilise un papier "gausso-arithmétique", papier dont l'axe des abscisses porte une échelle arithmétique, et dont l'axe des ordonnées est gradué selon la fonction de répartition de la loi normale réduite. Les résultats y_k sont rangés par valeurs croissantes ($y_1 < y_2 < \ldots < y_k < \ldots < y_n$). On porte en abscisse les valeurs y_k et en ordonnée les fréquences cumulées correspondantes, c'est-à-dire les quantités $k/(n + 1)$. Si la distribution à laquelle appartiennent les y_k est normale, ces points se situent au voisinage d'une droite.

Cette méthode permet aussi d'examiner si une transformation des résultats (changement de variable) conduit à un meilleur alignement que les résultats eux-mêmes. Les transformations les plus courantes sont des types suivants :

$$\text{Log } y, \sqrt{y}, \ 1/y.$$

2.4.2 – *Test de Shapiro et Wilk* [6]

Ce test est exposé dans l'article cité sous la référence [5], qui contient les tables extraites de [6]. S'il conduit à rejeter l'hypothèse de normalité un changement de variable susceptible de donner un accord satisfaisant avec cette hypothèse peut être recherché.

Dans la suite de cet article, le symbole y désignera soit la variable étudiée soit, éventuellement, la variable transformée.

2.5 – Test de l'hypothèse d'homogénéité des variances intralaboratoires

Il s'agit de l'hypothèse selon laquelle la variance intralaboratoire est la même dans tous les laboratoires. S'il en est ainsi, les p quantités s_i^2 définies aux § 2.3.1 et 2.3.3 fournissent p estimations indépendantes de la variance intralaboratoire σ_r^2.

Il existe plusieurs méthodes permettant de tester l'hypothèse d'homogénéité des quantités s_i^2, toutes dépendant plus ou moins fortement de l'hypothèse de normalité de la distribution de l'erreur intralaboratoire. Trois tests relatifs à l'homogénéité d'un groupe de variances sont décrits ci-après : test de Hartley, test de Cochran, test de Bartlett.

Avant d'exécuter ces tests, il est hautement recommandé de procéder à une représentation graphique de la distribution des quantités s_i^2. Une telle représentation (sous forme de diagramme ou d'histogramme) peut révéler certaines anomalies que certains tests seraient impuissants à déceler (séparation des s_i^2 en deux groupes nettement distincts par exemple).

Il est rappelé qu'aucun des tests, ni la représentation graphique, ne sont applicables lorsqu'on adopte pour l'essai interlaboratoires la modalité décrite au § 2.3.2.

2.5.1 – *Test de Hartley* [7]

Ce test suppose que le nombre de degrés de liberté ν est le même pour toutes les variances.

Soit s_{max}^2 la plus grande et s_{min}^2 la plus petite de k variances

$$s_i^2 \ (i = 1, 2, \ldots, k).$$

On calcule le rapport

$$q = \frac{s_{max}^2}{s_{min}^2}$$

La Table 1 (reproduite de [7]) donne, en fonction de k et de ν la valeur de q qui n'a que 5 chances sur 100 d'être dépassée ($q_{0,95}$) et celle qui n'a qu'une chance sur 100 d'être dépassée ($q_{0,99}$) lorsque l'ensemble des variances est homogène.

Si la valeur trouvée pour q est inférieure à $q_{0,95}$, l'hypothèse d'homogénéité des variances est acceptable ; si elle est comprise entre $q_{0,95}$ et $q_{0,99}$, cette hypothèse peut encore être acceptée ; si elle est supérieure à $q_{0,99}$ l'hypothèse d'homogénéité des variances doit être rejetée.

Le test de Hartley est rarement applicable à l'homogénéité des variances intralaboratoires. Il suppose que le nombre de résultats est le même dans chaque laboratoire et qu'il n'y a pas de quantités s_i^2 égales à 0. Une autre limitation est que la Table 1 n'a été calculée que pour un nombre de variances au plus égal à 12. Si cependant il est appliqué, la Table 1 doit être consultée avec :

$k = p$ (nombre de laboratoires)

$\nu = n - 1$ (n = nombre de résultats dans chaque laboratoire).

2.5.2 – *Test de Cochran* [8]

Ce test suppose encore que le nombre de degrés de liberté ν est le même pour toutes les variances.

Appelant s_{max}^2 la plus grande des quantités s_i^2, on forme

$$g = \frac{s_{max}^2}{\Sigma s_i^2}$$

La table 2 donne, en fonction du nombre k de variances (ici k = p) et du nombre de degrés de liberté ν attaché à chaque variance (ici $\nu = n - 1$), la valeur de g qui n'a que 5 chances sur 100 d'être dépassée ($g_{0,95}$) et celle qui n'a qu'une chance sur 100 d'être dépassée ($g_{0,99}$) lorsque l'ensemble des variances est homogène. L'interprétation est la même que pour le test de Hartley.

Le test de Cochran, qui suppose encore que le nombre de résultats est le même dans chaque variance intralaboratoire, ne permet de détecter que les variances anormalement élevées.

2.5.3 – *Test de Bartlett* [9]

Le nombre de degrés de liberté peut être quelconque dans chacune des k variances s_i^2 ; le nombre de degrés de liberté attaché à s_i^2 est désigné par ν_i.

On calcule la quantité

$$B = \frac{2{,}3026}{c} \left[\nu \log_{10} s^2 - \sum_{i=1}^{k} \log_{10} s_i^2 \right]$$

$$\nu = \sum_{i=1}^{k} \nu_i = \left(\sum_{i=1}^{k} n_i \right) - k \qquad s^2 = \frac{1}{\nu} \sum_{i=1}^{k} \nu_i \, s_i^2$$

$$c = 1 + \frac{1}{3(k-1)} \left[\sum_{i=1}^{k} \left(\frac{1}{\nu_i} \right) - \frac{1}{\nu} \right]$$

Dans l'hypothèse où l'ensemble des variances est homogène, la quantité B est approximativement distribuée en loi de χ^2 à $(k - 1)$ degrés de liberté. La table 3 donne, pour cette variable, la limite $\chi^2_{0,99}$ qui n'a que 5 chances sur 100 d'être dépassée et la limite $\chi^2_{0,99}$ qui n'a qu'une chance sur 100 d'être dépassée. L'interprétation est la même que pour les tests précédents, avec k = p (nombre de laboratoires), $\nu_i = n_i - 1$ (n_i = nombre de résultats dans le laboratoire L_i).

Le coefficient c de la formule de Bartlett n'est pas entièrement satisfaisant pour les valeurs de $\nu = 1, 2, 3$. Des tables plus précises ont été calculées, dont les références sont données dans la bibliographie, sous les numéros [10] et [11].

Le test de Bartlett est le plus général et le plus efficace mais il nécessite des calculs relativement compliqués. Le nombre de résultats dans chaque laboratoire peut être quelconque. La seule limitation est qu'aucune des variances s_i^2 ne doit être nulle : $s_i^2 = 0$ peut se produire si certains laboratoires n'ont fourni qu'un petit nombre de résultats (par exemple 2), et si ces résultats sont égaux.

2.5.4 – *Décision à prendre après exécution des tests*

Lorsque, à la suite de l'un ou l'autre de ces tests, l'hypothèse d'homogénéité des variances intalaboratoires est rejetée, il convient d'en rechercher la cause. Celle-ci peut être :

a) l'existence de un ou plusieurs résultats aberrants

b) une hétérogénéité réelle des variances.

Dans le cas a) les résultats aberrants sont éliminés, le test d'homogénéité des variances étant éventuellement effectué à nouveau après cette élimination.

Dans le cas b) il convient de procéder à un examen critique des conditions dans lesquelles l'essai a été exécuté ; il se peut que les instructions étaient insuffisamment détaillées, ou pouvaient prêter à des interprétations divergentes.

Dans certains cas le ou les tests font "boule de neige" et entraînent l'élimination d'un nombre important de résultats (plus de 10 %). Il y a alors lieu d'être très réservé sur la possibilité de tirer de l'essai des conclusions valables.

2.6 – Test de l'hypothèse de normalité de l'erreur totale

Il s'agit de l'hypothèse selon laquelle la distribution de l'erreur totale, somme des erreurs intra et interlaboratoires, est normale ou approximativement normale.

Le test de normalité s'effectue comme il est indiqué au § 2.4, en opérant sur les moyennes $\overline{y}_i$ (ou les totaux T_i) des laboratoires ayant participé à l'essai interlaboratoires et non éliminés lors des tests d'homogénéité des variances intralaboratoires. En toute rigueur, le nombre de résultats doit être le même dans tous les laboratoires. Les méthodes du § 2.4 peuvent cependant être appliquées si ce nombre est peu différent d'un laboratoire à l'autre (dans le cas contraire, la validité même de l'essai interlaboratoires serait sujette à caution).

Une autre possibilité est d'appliquer le test de Dixon [12] qui porte sur les valeurs extrêmes de la série des moyennes $\overline{y}_i$ rangées par valeurs croissantes. Bien que le test de Dixon soit d'application plus simple, nous préférons le test de Shapiro et Wilk qui fait intervenir la totalité des moyennes $\overline{y}_i$.

Lorsque l'hypothèse de normalité est rejetée, cela peut provenir

a) d'une, ou plusieurs moyennes de laboratoires aberrantes

b) d'un écart réel à la normalité.

Dans le cas a) les résultats aberrants sont éliminés, le test de normalité étant éventuellement effectué à nouveau après cette élimination.

Dans le cas b) on peut rechercher une transformation permettant d'obtenir des valeurs distribuées normalement ou de façon approximativement normale (cf. § 2.4.1).

Dans certains cas le processus d'élimination fait "boule de neige", conduisant à écarter un nombre important de résultats ou de laboratoires ; là encore on doit être très réservé sur la possibilité de tirer de l'essai des conclusions valables.

2.7 – Estimation des variances intra et interlaboratoires

2.7.1 – *Expérience comportant deux répétitions sur échantillons identiques*

C'est le cas décrit au § 2.3.1. Conformément aux notations de ce paragraphe, si y_{i1} et y_{i2} sont les résultats obtenus par le laboratoire

$$L_i \ (i = 1, 2, \ldots, p),$$

on a :

$$w_i = |y_{i1} - y_{i2}| \qquad T_i = y_{i1} + y_{i2} \qquad \bar{y}_i = \frac{T_i}{2} \qquad T = \sum_{i=1}^{p} T_i \qquad \bar{y} = \frac{T}{2p}$$

On calcule les quantités ci-après :

$$Q_r = \sum_{i=1}^{p} \frac{w_i^2}{2} \text{ (somme des carrés intralaboratoire)}$$

$$Q_L = 2 \sum_{i=1}^{p} (\bar{y}_i - \bar{y})^2 = \sum_{i=1}^{p} T_i^2 - \frac{T^2}{2p} \text{ (somme des carrés entre laboratoires)}$$

Un calcul élémentaire montre que les espérances mathématiques de ces quantités sont

$$E[Q_r] = p\,\sigma_r^2$$

$$E[Q_L] = (p - 1)\,(2\,\sigma_L^2 + \sigma_r^2)$$

Posant $M_r = \dfrac{Q_r}{p}$ (carré moyen intralaboratoire)

$M_L = \dfrac{Q_L}{p - 1}$ (carré moyen entre laboratoires)

on en déduit que les estimations de la variance intralaboratoire σ_r^2 (répétabilité), de la variance interlaboratoires σ_L^2 et de la variance totale σ_R^2 (reproductibilité) sont :

$$s_r^2 = M_r \qquad s_L^2 = \frac{M_L - M_r}{2} \qquad s_R^2 = \frac{M_L + M_r}{2} \qquad (1)$$

Exceptionnellement il peut arriver que $M_L < M_r$. On considérera alors que $s_L^2 = 0$ et $s_R^2 = s_r^2$.

2.7.2 – *Variante de l'expérience précédente* (deux niveaux A et B légèrement différents)

C'est le cas décrit au § 2.3.2. Conformément aux notations de ce paragraphe, si y_{iA} et y_{iB} sont les résultats obtenus aux niveaux A et B par le laboratoire L_i, on a :

$$w_i = |y_{iA} - y_{iB}| \qquad W = \sum_{i=1}^{p} w_i$$

La somme des carrés et le carré moyen intralaboratoires s'écrivent alors :

$$Q_r = \sum_{i=1}^{n_i} w_i^2 - \frac{W^2}{2p} \qquad M_r = \frac{Q_r}{p-1}$$

toutes les autres formules du § 2.7.1 (Q_L, M_L, s_r^2, s_L^2, s_R^2) étant inchangées.

2.7.3 – *Cas général*

Le cas général est le cas décrit au § 2.3.3 où, pour le laboratoire

$$L_i \ (i = 1, 2, \ldots, p)$$

on dispose de n_i résultats ($n_i \geqslant 2$), l'ensemble de ces résultats provenant d'échantillons identiques (ou aussi identiques que possible). Cette situation se présente par exemple, lorsqu'il y a plus de 2 répétitions et que certains résultats ont été éliminés à la suite des tests cités aux § 2.4 à 2.6.

Conformément aux notations du § 2.3.3, on pose :

p = nombre de laboratoires restants

n_i = nombre de résultats retenus dans le laboratoire i ($n_i \geqslant 2$)

$N = \sum_{i=1}^{p} n_i$ = nombre total de résultats retenus

$T_i = \sum_{k=1}^{n_i} y_{ik}$ = somme de tous les résultats retenus dans le laboratoire i

$$\bar{y}_i = \frac{T_i}{n_i}$$

$T = \sum_{i=1}^{p} T_i$ = somme de tous les résultats retenus

$$\bar{y} = \frac{T}{N}$$

Les sommes de carrés intra et entre laboratoires s'expriment par les formules suivantes :

$$Q_r = \sum_{i=1}^{p} \sum_{k=1}^{n_i} (y_{ik} - \bar{y}_i)^2 = \sum_{i=1}^{p} \sum_{k=1}^{n_i} y_{ik}^2 - \sum_{i=1}^{p} \frac{T_i^2}{n_i}$$

$$Q_L = \sum_{i=1}^{p} n_i (\bar{y}_i - \bar{y})^2 = \sum_{i=1}^{p} \frac{T_i^2}{n_i} - \frac{T^2}{N}$$

Un calcul classique montre que

$$E[Q_r] = (N - p)\, \sigma_r^2$$

$$E[Q_L] = \left(N - \frac{\Sigma n_i^2}{N}\right) \sigma_L^2 + (p - 1)\, \sigma_r^2$$

En posant :

$$M_r = \frac{Q_r}{N - p} \qquad \text{(carré moyen intralaboratoire)}$$

$$M_L = \frac{Q_L}{p - 1} \qquad \text{(carré moyen entre laboratoires)}$$

$$f = \frac{N(p - 1)}{N^2 - \sum_{i=1}^{p} n_i^2}$$

on en déduit les estimations de la variance intralaboratoire σ_r^2 (répétabilité), de la variance interlaboratoires σ_L^2, et de la variance totale σ_R^2 (reproductibilité) :

$$s_r^2 = M_r \qquad s_L^2 = f(M_L - M_r) \qquad s_R^2 = fM_L + (1 - f)M_r \qquad (2)$$

Dans le cas exceptionnel où l'on trouve $M_L < M_r$, on considère que $s_L^2 = 0$ et $s_R^2 = s_r^2$.

Si les nombres de résultats sont les mêmes dans les p laboratoires, on a :

$$n_i = n = \frac{N}{p} \qquad f = \frac{1}{n}$$

d'où :

$$s_r^2 = M_r \qquad s_L^2 = \frac{M_L - M_r}{n} \qquad s_R^2 = \frac{M_L + (n - 1)M_r}{n} \qquad (3)$$

expressions qui, pour n = 2, se réduisent à celles qui ont été données au § 2.7.1.

2.8 – Estimation de la répétabilité et de la reproductibilité

s_r^2 et s_R^2 sont des estimations des vraies valeurs des variances de répétabilité (σ_r^2) et de reproductibilité (σ_R^2). A condition que le nombre de laboratoires

et le nombre total de résultats soient suffisamment élevés (cf. § 2.2) ces estimations peuvent, sans trop de risque d'erreur, être assimilées aux vraies valeurs. En supposant d'autre part que les hypothèses de normalité sont suffisamment vérifiées, les relations entre s_r^2 et la répétabilité r d'une part, entre s_R^2 et la reproductibilité R d'autre part (r et R ayant les définitions données au § 1.1) sont approximativement :

$$r = 1{,}96\sqrt{2}\, s_r \# 2{,}8\, s_r$$

$$R = 1{,}96\sqrt{2}\, s_R \# 2{,}8\, s_R$$

Si les calculs ont été faits après un changement d'unité, la nouvelle unité étant h fois plus petite que l'unité initiale, on aura :

$$r = \frac{2{,}8\, s_r}{h} \qquad R = \frac{2{,}8\, s_R}{h} \tag{4}$$

2.9 – **Exemple**

L'exemple ci-après est emprunté à [1].

L'essai porte sur 36 laboratoires avec 2 résultats par laboratoire sur produits identiques.

On se trouve donc dans le cas simple décrit au § 2.3.1 et interprété au § 2.7.1.

Trois laboratoires (n° 18, 23 et 27) n'ont pas fourni les résultats. On a donc au départ p = 33, n = 2. Les données originales ont subi un changement d'origine (origine initiale diminuée de 70) et un changement d'unité (multiplication par 10).

N.B. Il est évident que les résultats fournis par le laboratoire n° 10 sont aberrants. Ce laboratoire a cependant été conservé afin de permettre une illustration aussi complète que possible des tests décrit dans les § 2.4 à 2.6.

Test de l'hypothèse de normalité de l'erreur intralaboratoire

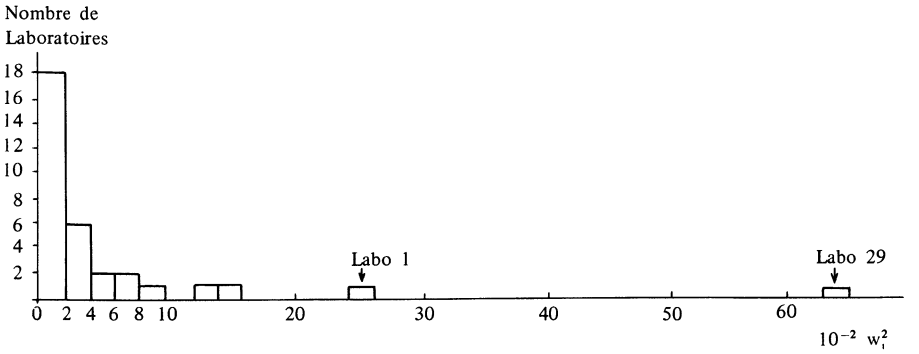
Lors de l'essai pilote ayant précédé l'essai interlaboratoires, on a constaté, aussi bien par la méthode graphique que par le test de Shapiro et Wilk, que cette hypothèse était acceptable.

Test de l'hypothèse d'homogénéité des variances intralaboratoires

L'histogramme de répartition des quantités $10^{-2}\, w_i^2 = 2 \times 10^{-2}\, s_i^2$ se présente de la façon suivante (étendue des classes égale à 2).

Tableau 1

Laboratoire	Résultats y_{i1}	y_{i2}	Total T_i	Moyenne $\bar{y}_i$	Etendue w_i	w_i^2
1	230	180	410	205	50	2 500
2	290	300	590	295	10	100
3	214	223	437	218,5	9	81
4	284	294	578	289	10	100
5	281	285	566	283	4	16
6	310	305	615	307,5	5	25
7	323	348	671	335,5	25	625
8	316	318	634	317	2	4
9	287	289	576	288	2	4
10	72	108	180	90	36	1 296
11	302	301	603	301,5	1	1
12	288	310	598	299	22	484
13	283	276	559	279,5	7	49
14	249	266	515	257,5	17	289
15	295	292	587	293,5	3	9
16	284	284	568	284	0	0
17	244	250	494	247	6	36
19	250	230	480	240	20	400
20	385	370	755	377,5	15	225
21	241	229	470	235	12	144
22	264	304	568	284	40	1 600
24	299	299	598	299	0	0
25	292	316	608	304	24	576
26	296	303	599	299,5	7	49
28	304	330	634	317	26	676
29	310	390	700	350	80	6 400
30	299	305	604	302	6	36
31	310	290	600	300	20	400
32	287	299	586	293	12	144
33	270	272	542	271	2	4
34	315	284	599	299,5	31	961
35	254	239	493	246,5	15	225
36	310	291	601	300,5	19	361
Total : 1/ sur résultats recueillis						17 820
2/ sans les laboratoires 10 et 29 éliminés à la suite des tests ci-après			17 738			10 124



La dispersion des résultats apparaît anormalement élevée dans le laboratoire 29, et peut-être dans le laboratoire 1.

Le test de Hartley $\left(\frac{s^2_{max}}{s^2_{min}}\right)$ ne peut être appliqué ($p > 12$).

Le test de Cochran avec $w^2_{max} = 6\,400$ et $\Sigma w_i^2 = 17\,820$ donne :

$$g = \frac{6\,400}{17\,820} = 0{,}359$$

Cette valeur est supérieure à la limite $g_{0,99} = 0{,}330$ trouvée dans la table 2 (interpolation entre $p = 30$ et $p = 40$). Il convient dont d'éliminer le laboratoire n° 29.

Le même test, appliqué à la nouvelle valeur $w^2_{max} = 2\,500$, conduit à conserver le laboratoire n° 1.

L'allure de l'histogramme, après élimination des valeurs aberrantes provenant du laboratoire n° 29, ne montre pas la nécessité de faire appel au test de Bartlett : celui-ci ne pourrait d'ailleurs être appliqué qu'en ne tenant pas compte des laboratoires n° 16 et 24 pour lesquels $w = 0$.

Test de l'hypothèse de normalité de l'erreur totale

Dans le tableau 2 ci-après, les 32 laboratoires (laboratoire 29 éliminé) sont classés suivant les moyennes $\bar{y}_i$ croissantes (2ème colonne).

L'histogramme de ces moyennes (classes d'étendue égale à 10) figure immédiatement après le tableau 1.

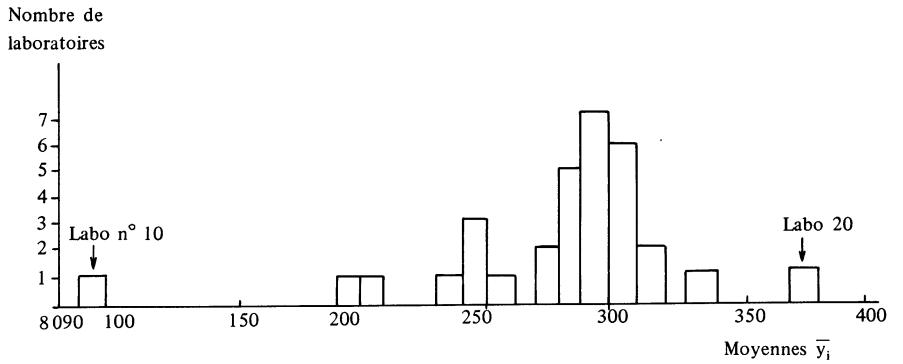


Tableau 2

Laboratoire	$\bar{y}_i$	d_i	a_i
10	90	287,5	0,4188
1	205	130,5	0,2898
3	218,5	98,5	0,2463
21	235	82	0,2141
19	240	67,5	0,1878
35	246,5	57,5	0,1651
17	247	55	0,1449
14	257,5	44	0,1265
33	271	29,5	0,1093
13	279,5	20,5	0,0931
5	283	16,5	0,0777
16	284	15,5	0,0629
22	284	15	0,0485
9	288	11	0,0341
4	289	6	0,0206
32	293	0,5	0,0068
15	293,5		
2	295		
12	299		
24	299		
26	299,5		
34	299,5		
31	300		
36	300,5		
11	301,5		
30	302		
25	304		
6	307,5		
8	317		
28	317		
7	335,5		
20	377,5		

La distribution des moyennes $\bar{y}_i$ paraît s'écarter de la normalité en raison d'une moyenne anormalement faible (laboratoire 10) et peut-être d'une moyenne trop élevée (laboratoire 20).

Appliquant le test de Shapiro et Wilk ([5]et [6]) on calcule les différences $d_i = \bar{y}_{32-i+1} - \bar{y}_i$ (3ème colonne du tableau 2), et on leur affecte les coefficients a_i.

On trouve :

$$b = \sum_{i=1}^{16} a_i d_i = 244,3687$$

$$b^2 = 59.716$$

D'autre part :

$$S^2 = \sum_{i=1}^{32} \bar{y}_i^2 - \frac{(\Sigma \bar{y}_i)^2}{32} = 2\,580\,839 - 2\,508\,240$$

$$= 72\,599$$

$$W = \frac{b^2}{S^2} = 0,822$$

Cette valeur est nettement inférieure à celle qui correspond au seuil 1 % pour n = 32 ($W_{0,99} = 0,904$). L'hypothèse que les 32 valeurs $\bar{y}_i$ appartiennent à une même distribution normale doit être rejetée.

On élimine le laboratoire 10, et on reprend le test de Shapiro et Wilk sur les 31 laboratoires restants. On trouve alors :

$$b = 181,3507$$

$$b^2 = 32.888$$

$$S^2 = 35\,346$$

$$W = \frac{b^2}{S^2} = 0,930$$

Pour p = 31, $W_{0,95} = 0,929$

L'hypothèse que les 31 moyennes $\bar{y}_i$ appartiennent à une même distribution normale est acceptable ; il n'y a pas lieu d'éliminer le laboratoire 20.

En définitive, après élimination des laboratoires 29 et 10, il reste p = 31 laboratoires.

Estimation des variances intra et interlaboratoires (cf. Tableau 1).

$$Q_r = \frac{1}{2} \sum_{i=1}^{31} w_i^2 = \frac{10\,124}{2} = 5\,062$$

$$Q_L = \frac{1}{2} \sum_{i=1}^{31} T_i^2 - \frac{T^2}{62} = \frac{1}{2} [(410)^2 + (590)^2 + \ldots + (601)^2]$$

$$- \frac{(17\,738)^2}{62} = \frac{10\,290\,956}{2} - 5\,074\,785 = 70\,693$$

$$M_r = \frac{5\,062}{31} = 163{,}3$$

$$M_L = \frac{70\,693}{30} = 2\,356{,}4$$

D'après les formules (1) on a :

$$s_r^2 = 163{,}3$$

$$s_L^2 = \frac{1}{2} [2\,356{,}4 - 163{,}3] = 1\,096{,}6$$

$$s_R^2 = \frac{1}{2} [2\,356{,}4 + 163{,}3] = 1\,259{,}9$$

Estimation de la répétabilité et de la reproductibilité de la méthode

Les calculs ont été faits avec une unité 10 fois plus petite que l'unité initiale. On a donc (cf. formules 4) :

$$r = \frac{2{,}8}{10} \sqrt{163{,}3} = 3{,}6$$

$$R = \frac{2{,}8}{10} \sqrt{1\,259{,}9} = 9{,}9$$

3. – CAS DE PLUSIEURS NIVEAUX LORSQUE REPETABILITE ET REPRODUCTIBILITE SONT INDEPENDANTES DU NIVEAU

3.1 – Hypothèses et modèle

Hypothèses – Les hypothèses du § 2.1, supposées satisfaites à chaque niveau, doivent être complétées comme suit :

– la variance intralaboratoire est la même dans tous les laboratoires et pour tous les niveaux,

– la variance interlaboratoires est la même à tous les niveaux.

Le modèle correspondant s'écrit :

$$y_{ijk} = m + b_i + c_j + e_{ijk}$$

y_{ijk} = l'un des résultats obtenus par le laboratoire (i), au niveau (j) dans les conditions de répétabilité.

e_{ijk} = erreur aléatoire affectant ce résultat.

c_j = constante caractérisant le niveau (j).

D'après les hypothèses faites

e_{ijk} appartient à une distribution normale ou approximativement normale de moyenne 0 et de variance σ_r^2 (variance intralaboratoire, ou de répétabilité).

$$E(e_{ijk}) = 0 \qquad E(e_{ijk}^2) = \sigma_r^2$$

quels que soient i et j.

b_i considéré comme aléatoire, appartient à une distribution normale ou approximativement normale de moyenne 0 et de variance σ_L^2 (variance interlaboratoires).

$$E(b_i) = 0 \qquad E(b_i^2) = \sigma_L^2$$

La variance totale ou variance de reproductibilité est :

$$\sigma_R^2 = \sigma_r^2 + \sigma_L^2$$

3.2 – Nombre de niveaux, de laboratoires et de répétitions

Le nombre de niveaux dépend de la plage des valeurs de la grandeur mesurée que l'on désire étudier. Si cette plage est très étendue, il est peu probable que l'hypothèse selon laquelle répétabilité et reproductibilité sont indépendantes du niveau soit satisfaite.

En ce qui concerne le nombre de laboratoires et le nombre de répétitions, les conditions indiquées au § 2.2 demeurent valables. Toutefois comme on a affaire a plusieurs expériences (plusieurs niveaux), le nombre de laboratoires peut, sans grave inconvénient, être abaissé à une valeur inférieure à 12.

3.3 – Différentes modalités de l'essai – Notations

Lorsque les résultats ont été fournis, on procède comme il est indiqué au début du § 2.3 (élimination des résultats manifestement aberrants, changements éventuels d'origine et d'unité).

Les différentes modalités de l'essai sont, pour chaque niveau, celles qui sont indiquées au § 2.3 :

– deux déterminations (n = 2) sur échantillons identiques ou aussi identiques que possible à chaque niveau (cf. § 2.3.1)

– deux déterminations ($n = 2$) sur des échantillons correspondant à des niveaux légèrement différents, encadrant un "niveau moyen" (cf. § 2.3.2)

– cas général où le nombre de déterminations n_{ij} pour les différentes combinaisons L_iM_j, est quelconque ($n_{ij} \geqslant 2$), les échantillons étant, à chaque niveau, identiques ou aussi identiques que possible (cf. § 2.3.3).

Dans le cas général, on aboutit ainsi à un tableau tel que le tableau 3 où certaines cases peuvent être vides ($n_{ij} = 0$). Le nombre de laboratoires et le nombre de niveaux sont désignés respectivement par p et q.

Tableau 3

		Niveaux				
		M_1	.....	M_j	.....	M_q
Laboratoires	L_1		.....		.....	
	...		.....	.....	.....	
	L_i			n_{ij} résultats y_{ijk} de total T_{ij}		
	L_p					
Nombre total de résultats				N_j		
Total des résultats				T_j		

Combinaison L_iM_j, n_{ij} résultats, dont le total est $T_{ij} = \sum_{k=1}^{n_{ij}} y_{ijk}$

Niveau M_j, N_j résultats, dont le total est $T_j = \sum_{i=1}^{p} T_{ij}$

Le nombre de laboratoires effectivement représentés au niveau j est désigné par p_j.

Nombre total de résultats = N

3.4 – Test des hypothèses à chaque niveau

Ces tests s'effectuent comme il est indiqué aux § 2.4, 2.5 et 2.6, et peuvent conduire à éliminer des résultats isolés, ou un laboratoire tout entier à un niveau particulier. L'élimination d'un assez grand nombre de résultats (plus de 10 %) –et notamment si plusieurs cases du tableau deviennent vides– doit conduire à être très réservé sur la possibilité de tirer de l'essai des conclusions valables.

Après les éliminations éventuelles, on désigne encore par p et q le nombre de laboratoires et de niveaux non complètement éliminés, et les notations du § 3.3 (n_{ij}, p_j, N_j, T_j, N) s'appliquent aux seuls résultats conservés. Le nombre de cases vides du tableau des résultats ($n_{.j} = 0$) est désigné par V.

3.5 – Test des hypothèses selon lesquelles répétabilité et reproductibilité sont indépendantes du niveau

3.5.1 – *Test concernant la répétabilité*

Ayant calculé, pour chaque niveau M_j comme il est indiqué au § 2.7, les carrés moyens intralaboratoires $M_{r,j} = s^2_{r,j}$ (j = 1, 2, . . . q) on leur affecte le nombre de degrés de liberté convenable, soit

$\nu_j = p_j$ dans le cas où l'essai est organisé, à chaque niveau, comme il est indiqué au § 2.3.1.

$\nu_j = p_j - 1$ dans le cas où l'essai est organisé, à chaque niveau, comme il est indiqué au § 2.3.2.

$\nu_j = N_j - p_j$ dans le cas général.

On teste l'hypothèse selon laquelle la répétabilité est indépendante du niveau en appliquant aux q variances intralaboratoires $s^2_{r,j}$ le test de Bartlett ou, si les conditions d'application de ces tests sont satisfaites, le test de Hartley ou le test de Cochran.

Dans le cas général (incluant le cas particulier où $n_{ij} = 2$ pour toutes les combinaisons L_iM_j) on peut également effectuer un "test global" portant sur l'ensemble des estimations de la variance intralaboratoire obtenues pour les différentes combinaisons L_iM_j ($\nu_{ij} = n_{ij} - 1$ degrés de liberté).

3.5.2 – *Test concernant la reproductibilité*

Ayant calculé, pour chaque niveau M_j, comme il est indiqué au § 2.7, les carrés moyens entre laboratoires $M_{L,j}$, on leur affecte, dans tous les cas, le nombre de degrés de liberté $\nu_j = p_j - 1$.

On teste l'hypothèse selon laquelle la reproductibilité est indépendante du niveau en appliquant aux q carrés moyens $M_{L,j}$ le ou les tests indiqués au § 3.5.1 pour les variances intralaboratoires.

3.5.3 – *Décision à prendre à la suite de ces tests*

Si l'un ou l'autre des tests conduit au rejet de l'hypothèse testée, on devra procéder comme il est indiqué plus loin au § 4 (cas général où la répétabilité et la reproductibilité dépendent du niveau).

Si au contraire, on peut admettre que la répétabilité et la reproductibilité sont indépendantes du niveau, il est légitime de rassembler additivement les sommes de carré intralaboratoire et entre laboratoires, ainsi que leurs degrés de liberté, ce qui conduit aux calculs décrits dans les § 3.6 et 3.7 ci-après.

3.6 – Estimation des variances intra et interlaboratoires et de la variance totale valables pour tous les niveaux.

On additionne les q sommes de carrés intralaboratoire $Q_{r,j}$ et on divise par le total des degrés de liberté qui leur sont associés. On obtient ainsi le *carré moyen intralaboratoire* M_r.

On additionne les q sommes de carrés entre laboratoires $Q_{L,j}$, et on divise par le total des degrés de liberté qui leur sont associés. On obtient ainsi le *carré moyen entre laboratoires* M_L.

3.6.1 – *Expérience comportant, à chaque niveau, deux répétitions sur échantillons identiques*

Dans ce cas, qui est celui décrit, pour chaque niveau, au § 2.7.1, on a :

$$M_r = \frac{\sum_{j=1}^{q} Q_{r,j}}{pq} \qquad M_L = \frac{\sum_{j=1}^{q} Q_{L,j}}{q(p-1)}$$

On en déduit les estimations de σ_r^2, σ_L^2 et $\sigma_R^2 = \sigma_L^2 + \sigma_r^2$ par les formules suivantes :

$$\begin{cases} s_r^2 = M_r \\ s_L^2 = \dfrac{M_L - M_r}{2} \\ s_R^2 = \dfrac{M_L + M_r}{2} \end{cases} \tag{5}$$

3.6.2 – *Variante de l'expérience précédente* (deux niveaux légèrement différents à chaque niveau moyen)

Dans ce cas (qui est celui décrit pour chaque niveau, au § 2.7.2), on a :

$$M_r = \frac{\sum_{j=1}^{q} Q_{r,j}}{q\,(p-1)}$$

toutes les autres formules du § 3.6.2 étant inchangées.

3.6.3 – *Cas général*

Dans le cas général, pour lequel les notations sont celles du tableau 3, on a :

$$M_r = \frac{\sum_{j=1}^{q} Q_{r,j}}{N - pq + V} \qquad M_L = \frac{\sum_{j=1}^{q} Q_{L,j}}{q(p-1) - V}$$

(V = nombre de "cases vides" du tableau 3).

Ayant calculé, pour chaque niveau, le coefficient

$$f_j = \frac{N_j\,(p_j - 1)}{N_j^2 - \sum_{i=1}^{p} n_{ij}^2}$$

on pose :

$$\frac{1}{\lambda} = \frac{\sum_{j=1}^{q} \frac{p_j - 1}{f_j}}{q(p-1) - V}$$

On en déduit les estimations de σ_r^2, σ_L^2, et $\sigma_R^2 = \sigma_L^2 + \sigma_r^2$ par les formules suivantes :

$$\begin{cases} s_r^2 = M_r \\ s_L^2 = \lambda\,(M_L - M_r) \\ s_R^2 = \lambda M_L + (1-\lambda)\,M_r \end{cases} \tag{6}$$

Lorsque le nombre de résultats est le même pour toutes les combinaisons L_iM_j, on a :

$$n_{ij} = n, \quad V = 0, \quad p_j = p, \quad N_j = pn, \quad f_j = \lambda = \frac{1}{n}$$

Les expressions (6) s'écrivent alors :

$$\begin{cases} s_r^2 = M_r \\ s_L^2 = \dfrac{M_L - M_r}{n} \\ s_R^2 = \dfrac{M_L + (n-1)\,M_r}{n} \end{cases} \tag{7}$$

qui se réduisent aux expressions (5) du § 3.6.1 lorsque n = 2.

3.7 – Estimation de la répétabilité et de la reproductibilité valables pour tous les niveaux

Si les calculs ont été faits après un changement d'unité, la nouvelle unité étant h fois plus petite que l'unité initiale, la répétabilité r et la reproductibilité R de la méthode, sont estimées de la façon suivante :

$$\begin{cases} r = \dfrac{1{,}96\sqrt{2}\,s_r}{h} \# \dfrac{2{,}8\,s_r}{h} \\ R = \dfrac{1{,}96\sqrt{2}\,s_R}{h} \# \dfrac{2{,}8\,s_R}{h} \end{cases} \tag{8}$$

Tableau 4

Laboratoire	Niveau M_1				Niveau M_2				Niveau M_3				Total	
	y	Total	w	w^2	y	Total	w	w^2	y	Total	w	w^2	Total	Σw^2
L_1	31,7 31,2	62,9	0,5	0,25	38,0 37,9	75,9	0,1	0,01	51,5 51,4	102,9	0,1	0,01	241,7	0,27
L_2	30,9 30,9	61,8	0,0	0,00	37,9 38,0	75,9	0,1	0,01	50,4 50,9	101,3	0,5	0,25	239,0	0,26
L_3	30,7 30,9	61,6	0,2	0,04	37,9 38,3	76,2	0,4	0,16	51,6 51,4	103,0	0,2	0,04	240,8	0,24
L_4	31,5 31,1	62,6	0,4	0,16	37,8 37,7	75,5	0,1	0,01	51,2 51,2	102,4	0,0	0,00	240,5	0,17
L_5	31,3 31,6	62,9	0,3	0,09	38,7 38,3	77,0	0,4	0,16	51,6 51,6	103,2	0,0	0,00	243,1	0,25
L_6	31,6 31,4	63,0	0,2	0,04	38,0 38,0	76,0	0,0	0,00	51,5 52,0	103,5	0,5	0,25	242,5	0,29
L_7	30,3 30,2	60,5	0,1	0,01	37,0 37,2	74,2	0,2	0,04	50,5 50,6	101,1	0,1	0,01	235,8	0,06
L_8	31,4 31,4	62,8	0,0	0,00	38,7 38,0	76,7	0,7	0,49	51,7 51,6	103,3	0,1	0,01	242,8	0,50
L_9	32,2 31,6	63,8	0,6	0,36	38,8 38,7	77,5	0,1	0,01	52,4 52,4	104,8	0,0	0,00	246,1	0,37
L_{10}	31,5 31,8	63,3	0,3	0,09	38,1 38,2	76,3	0,1	0,01	51,9 52,1	104,0	0,2	0,04	243,6	0,14
L_{11}	31,2 30,8	62,0	0,4	0,16	38,1 38,0	76,1	0,1	0,01	51,5 51,4	102,9	0,1	0,01	241,0	0,18
L_{12}	31,7 31,9	63,6	0,2	0,04	37,8 37,5	75,3	0,3	0,09	50,6 51,2	101,8	0,6	0,36	240,7	0,49
L_{13}	31,7 31,7	63,4	0,0	0,00	38,0 38,2	76,2	0,2	0,04	52,0 51,6	103,6	0,4	0,16	243,2	0,20
L_{14}	31,7 31,6	63,3	0,1	0,01	38,2 38,3	76,5	0,1	0,01	51,8 51,5	103,3	0,3	0,09	243,1	0,11
L_{15}	30,4 30,6	61,0	0,2	0,04	37,5 37,6	75,1	0,1	0,01	51,5 51,1	102,6	0,4	0,16	238,7	0,21
Total	–	938,5	–	–	–	1 140,4	–	–	–	1 543,7	–	–	3 622,6	–
Σw^2	–	–	–	1,29	–	–	–	1,06	–	–	–	1,39	–	3,74

Ces estimations sont valables dans toute la plage des niveaux étudiés.

3.8 – Exemple

Le tableau ci-après donne, d'après des données fournies par l'AFNOR, les résultats d'un essai interlaboratoires auquel ont participé $p = 15$ laboratoires, avec $q = 3$ niveaux de la grandeur étudiée. Chaque laboratoire a effectué $n = 2$ déterminations, les $2p = 30$ échantillons représentant chaque niveau étant aussi identiques que possible. On se trouve donc, à chaque niveau, dans le cas décrit au § 2.3.1.

Un premier examen des résultats permet de constater qu'aucun de ceux-ci n'est manifestement aberrant.

Test des hypothèses à chaque niveau

Les représentations graphiques indiquées aux § 2.4 à 2.6 (il est inutile de les donner ici) font présumer que ces hypothèses seront vérifiées.

Normalité de l'erreur intralaboratoire

Lors d'essais pilotes, on a vérifié que l'erreur intralaboratoire peut, à chaque niveau, être considérée comme distribuée normalement.

Normalité de l'erreur totale

Le test de Shapiro et Wilk, appliqué aux totaux T_i de laboratoires (totaux de deux mesures) donne les résultats résumés ci-après :

Tableau 5

Coefficients a_i [5]	$d_i = T_{15-i+1} - T_i$		
	M_1	M_2	M_3
0,5150	3,3	3,3	3,7
0,3306	2,6	1,9	2,7
0,2495	1,8	1,4	1,8
0,1878	1,5	1,0	1,1
0,1353	1,3	0,4	0,7
0,0880	0,4	0,3	0,4
0,0433	0,1	0,2	0,3
$b = \Sigma a_i d_i$	3,50528	2,95392	3,59670
b^2	12,28700	8,72564	12,93625
S^2 (1)	13,3933	8,9693	13,4373
$W = \frac{b^2}{S^2}$	0,917	0,973	0,963

(1) Pour chaque niveau, T_i désignant le total des résultats du laboratoire (i)

$$S^2 = \sum_{i=1}^{15} T_i^2 - \frac{(\Sigma T_i)^2}{15}$$

Par exemple, au niveau (1) :

$$S^2 = (62,9)^2 + (61,8)^2 + \ldots + (61,0)^2 - \frac{(938,5)^2}{15}$$

$$= 13,3933$$

Conclusion

A chaque niveau, W est nettement supérieur à $W_{0,95}$ qui est égal à 0,881 [5]. L'hypothèse de normalité de l'erreur totale est acceptable.

Homogénéité des variances intralaboratoires

p étant supérieur à 12 et plusieurs valeurs w_i étant nulles, seul le test de Cochran est applicable. Pour les différents niveaux, les valeurs du rapport

$$g = \frac{s^2_{max}}{\Sigma s_i^2} = \frac{w^2_{max}}{\Sigma w_i^2}$$

sont les suivants :

Niveau M_1 : $g = \frac{0,36}{1,29} = 0,279$

Niveau M_2 : $g = \frac{0,49}{1,06} = 0,462$

Niveau M_3 : $g = \frac{0,36}{1,39} = 0,259$

A chaque niveau, g est inférieur à la valeur $g_{0,95} = 0,471$ lue dans la Table 2 pour $\nu = 1$, $k = 15$. L'hypothèse d'homogénéité, à chaque niveau, des variances intralaboratoires est acceptable.

Test de l'hypothèse selon laquelle la répétabilité est indépendante du niveau

Le test de Cochran, appliqué à l'ensemble des 45 estimations de la variance intralaboratoire (tous niveaux) donne :

$$g = \frac{0,49}{3,74} = 0,131 \qquad g_{0,95} > 0,20$$

L'ensemble des 45 estimations peut être considéré comme homogène ; l'hypothèse que la répétabilité est indépendante du niveau est acceptable.

La conclusion est la même si on applique la méthode du § 3.5.1. Les estimations de la variance intralaboratoire aux différents niveaux sont (cf. Tableau 4, Total des w_i^2 à chaque niveau)

Niveau M_1 : $M_{r,1} = s_{r,1}^2 = \frac{1,29}{30} = 0,043$

Niveau M_2 : $M_{r,2} = s_{r,2}^2 = \frac{1,06}{30} = 0,035$

Niveau M_3 : $M_{r,3} = s_{r,3}^2 = \frac{1,39}{30} = 0,046$

Total 0,124

Test de Cochran $g = \frac{0,046}{0,124} = 0,371$ pour $k = 3$, $\nu = 15$, $g_{0,95} \# 0,54$

Test de Hartley $q = \frac{0,046}{0,035} = 1,31$ pour $k = 3$, $\nu = 15$, $q_{0,95} \# 3,49$

Test de l'hypothèse selon laquelle la reproductibilité est indépendante du niveau

Les carrés moyens entre laboratoires se déduisent du tableau 5 : ce sont les valeurs S^2 divisées par $2(p - 1) = 28$; chacun de ces carrés moyens comporte $\nu = p - 1 = 14$ degrés de liberté

Niveau M_1 : $M_{L,1} = \frac{13,3933}{28} = 0,478$

Niveau M_2 : $M_{L,2} = \frac{8,9693}{28} = 0,320$

Niveau M_3 : $M_{L,3} = \frac{13,4373}{28} = 0,480$

Total 1,278

Test de Cochran : $g = \frac{0,480}{1,278} = 0,376$; pour $k = 3$, $\nu = 14$, $g_{0,95} \# 0,56$

Test de Hartley : $q = \frac{0,480}{0,320} = 1,50$; pour $k = 3$, $\nu = 14$, $q_{0,95} \# 3,49$

Les deux tests concluent que la reproductibilité peut être considérée comme indépendante du niveau.

Estimation des variances intra et interlaboratoires et de la variance totale valables pour tous les niveaux

Les carrés moyens intralaboratoire et entre laboratoires sont

$$M_r = \frac{M_{r,1} + M_{r,2} + M_{r,3}}{3} = 0{,}0413$$

$$M_L = \frac{M_{L,1} + M_{L,2} + M_{L,3}}{3} = 0{,}4260$$

Des formules (5), on déduit :

$$s_r^2 = M_r = 0{,}0413$$

$$s_L^2 = \frac{M_L - M_r}{2} = 0{,}1923$$

$$s_R^2 = \frac{M_L + M_r}{2} = 0{,}2336$$

Estimations de la répétabilité et de la reproductibilité valables pour tous les niveaux

$$r = 2{,}8\ s_r = 2{,}8\sqrt{0{,}0413} = 0{,}6$$

$$R = 2{,}8\ s_R = 2{,}8\sqrt{0{,}2336} = 1{,}4$$

Ces estimations sont valables à l'intérieur de la plage des niveaux étudiés, soit :

du niveau correspondant à la moyenne générale pour $M_1 = \frac{938{,}5}{30} \# 30$

au niveau correspondant à la moyenne générale pour $M_3 = \frac{1\,543{,}7}{30} \# 50$

4 – CAS GENERAL

Lorsque l'essai porte sur une plage étendue de niveaux, on constate généralement que la répétabilité et la reproductibilité croissent avec le niveau. La méthode faisant l'objet du chapitre 3 n'est alors pas applicable.

Les théoriciens de la statistique préconisent souvent (cf. notamment [3]) de procéder à un changement de variable susceptible de "stabiliser" les variances de répétabilité et de reproductibilité, puis de procéder à une "*analyse globale*" sur résultats transformés, avec introduction d'un terme d'interaction entre laboratoires et niveaux.

Nous nous bornerons à indiquer le principe de cette méthode à laquelle nous ne sommes pas favorables (pour les raisons que nous indiquerons), préférant la méthode plus directe de l'"*analyse par niveau*" qui fera l'objet du § 4.2.

4.1 – Analyse globale

On admet qu'il existe une fonction de la vraie valeur m de la grandeur telle que, $\sigma(y)$ désignant, soit l'écart-type de répétabilité, soit celui de reproductibilité, on a :

$$\sigma(y) = f(m) \tag{9}$$

On montre facilement que pour la variable z déduite de la variable y par la transformation fonctionnelle

$$z = \varphi(y) = \lambda \int \frac{dy}{f(y)} \tag{10}$$

(λ étant une constante arbitraire), l'écart-type de répétabilité (de reproductibilité) devient pratiquement indépendant de m.

On procède à une analyse de variance à deux facteurs (laboratoires et niveaux) avec répétitions, sur l'ensemble des résultats transformés z_{ijk} ($i = 1, 2, \ldots, p$; $j = 1, 2, \ldots, q$; $k = 1, 2, \ldots, n$, si aucun des résultats n'est manquant ou n'a été rejeté). Cette analyse permet d'estimer, pour la variable z :

– une variance intralaboratoire (répétabilité) $\sigma_r^2(z)$
– une variance interlaboratoire $\sigma_L^2(z)$
– une variance d'interaction laboratoires x niveaux $\sigma_{L \times M}^2(z)$

et d'en déduire, toujours pour la variable z, la variance de reproductibilité :

$$\sigma_R^2(z) = \sigma_r^2(z) + \sigma_L^2(z) + \sigma_{L \times M}^2(z)$$

Cette méthode, séduisante sur le plan théorique, présente selon nous les inconvénients, ou absences d'avantages ci-après.

– La fonction de transformation f(m) doit être recherchée empiriquement ; comme elle doit s'appliquer à la fois à la répétabilité et à la reproductibilité, son choix résulte généralement d'un compromis qui la rend incertaine.

– L'analyse de la variance n'est relativement aisée que si le tableau des z_{ijk} est complet (npq valeurs) ; les données non fournies ou rejetées comme aberrantes doivent être estimées, ce qui nécessite des calculs supplémentaires assez compliqués s'il y a plusieurs telles données. Le nombre de degrés de liberté affecté aux différentes sommes de carrés de l'analyse de variance doit tenir compte du nombre de ces estimations.

– L'introduction de la variance d'interaction entre laboratoires et niveaux $\sigma_{L \times M}^2$ ne présente pas d'intérêt direct puisque, après avoir obtenu son estimation, on l'ajoute aux estimations de σ_r^2 et de σ_L^2 pour obtenir l'estimation de la variance de reproductibilité σ_R^2.

– Lorsqu'on a obtenu les estimations de $\sigma_r(z)$ et $\sigma_R(z)$ –valables pour tous les niveaux– on en déduit les estimations de $\sigma_r(y)$ et $\sigma_R(y)$ en fonction du niveau m, par la transformation déduite de la relation (10), soit (σ désignant σ_r ou σ_R)

$$\sigma(y) = \frac{\sigma(z)}{\frac{d\varphi}{dm}} = \frac{\sigma(z)}{\lambda} f(m)$$

Cette relation rapprochée de la relation (9), montre que l'on devrait théoriquement trouver, à la fin de l'analyse globale, $\sigma(z) = \lambda$, constante choisie dans la fonction de transformation. L'intérêt théorique de l'analyse globale serait donc de substituer à la relation de départ

$$\sigma(y) = f(m)$$

la relation en principe améliorée :

$$\sigma(y) = \frac{\sigma(z)}{\lambda} f(m) \qquad \text{avec} \qquad \frac{\sigma(z)}{\lambda} \# 1$$

Compte tenu de l'incertitude qui règne sur le choix de la fonction de transformation f(m), il est pour le moins douteux que l'amélioration soit réelle.

– Enfin, pour le praticien, l'analyse globale présente un certain ésotérisme qui risque de le rebuter – ou tout au moins de lui masquer la réalité objective du phénomène étudié.

C'est pourquoi nous recommandons, comme beaucoup plus simple, plus directe, et selon nous tout aussi précise, la méthode de l'analyse par niveau.

4.2 – Analyse par niveau

Il y a peu de choses à dire sur cette méthode qui consiste à appliquer, à chaque niveau, la méthode décrite au chapitre 2. On peut ensuite rechercher, graphiquement ou analytiquement, un ajustement de la répétabilité et de la reproductibilité en fonction du niveau (celui-ci étant défini par la moyenne des résultats obtenus à chaque niveau considéré).

4.3 – Exemple

L'exemple porte sur q = 5 niveaux et p = 9 laboratoires avec n = 2 répétitions. Les données initiales ont été multipliées par 100 afin d'éviter les décimales. Données et éléments de calcul sont présentés dans le tableau 6.

Les moyennes générales par niveau varient dans le rapport de 1 à 5, ce qui laisse présumer que la répétabilité et la reproductibilité dépendront du niveau.

Homogénéité des variances interlaboratoires

Etant admis que, pour un laboratoire et un niveau déterminé, les résultats sont distribués suivant une loi approximativement normale, nous laissons au lecteur le soin de vérifier que l'hypothèse d'homogénéité des variances intralaboratoires (test de Cochran, le seul possible plusieurs des valeurs w étant nulles) est acceptable à chaque niveau pris isolément.

Par contre l'hypothèse d'homogénéité doit être rejetée pour l'ensemble des 45 variances intralaboratoires (tous niveaux). On a en effet

$$g = \frac{s^2_{max}}{\Sigma s_i^2} = \frac{39\,204}{91\,394} = 0{,}429$$

valeur très supérieure à la valeur $g_{0,99} = 0{,}274$ lue dans la table 2 pour $\nu = 1$, $k = 45$.

On aboutit à la même conclusion en appliquant le test de Hartley ou le test de Cochran aux quantités Σw^2 obtenus aux 5 niveaux M_1 à M_5, avec $\nu = 9$, $k = 5$.

$$q = \frac{61\,663}{1\,384} = 44{,}6 > q_{0,99} = 11{,}1 \qquad \text{(cf. Table 1)}$$

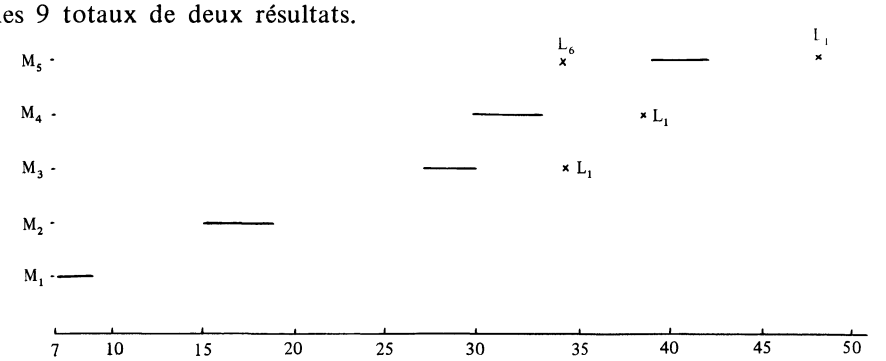
$$g = \frac{61\,663}{91\,394} = 0{,}675 > g_{0,99} = 0{,}485 \qquad \text{(cf. Tableau 2)}$$

(A noter que le test de Hartley, ici applicable, est préférable au test de Cochran, puisqu'il tient compte à la fois de la plus petite et de la plus grande des 5 estimations de variance).

En conclusion, la variance interlaboratoire, que l'on peut considérée comme constante à chaque niveau, varie avec le niveau : elle est fonction croissante de celui-ci.

Test de l'hypothèse de normalité, à chaque niveau, de l'erreur totale

Le graphique ci-après montre comment se répartissent, à chaque niveau, les 9 totaux de deux résultats.



Les traits horizontaux, accompagnés éventuellement des points isolés tels que L_1, L_σ, indiquent la plage où, pour chaque niveau, se répartissent les 9 totaux des laboratoires.

Tableau 6

Laboratoire	Niveau M_1				Niveau M_2				Niveau M_3				Niveau M_4				Niveau M_5				
	y	Total	w	w^2	y	Total	w	w^2	y	Total	w	w^2	y	Total	w	w^2	y	Total	w	w^2	
L_1	444 439	883	5	25	934 934	1 868	0	0	1 740 1 690	3 430	50	2 500	1 923 1 923	3 846	0	0	2 428 2 400	4 828	28	784	
L_2	403 423	826	20	400	842 833	1 675	9	81	1 442 1 450	2 892	8	64	1 606 1 622	3 228	16	256	2 040 1 991	4 031	49	2 401	
L_3	370 370	740	0	0	760 740	1 500	20	400	1 360 1 360	2 720	0	0	1 450 1 510	2 960	60	3 600	1 930 1 970	3 900	40	1 600	
L_4	410 410	820	0	0	893 880	1 773	13	169	1 460 1 420	2 880	40	1 600	1 560 1 550	3 110	10	100	2 030 2 030	4 060	0	0	
L_5	397 404	801	7	49	789 812	1 601	23	529	1 373 1 392	2 765	19	361	1 554 1 578	3 132	24	576	2 053 2 088	4 141	35	1 225	
L_6	375 403	778	28	784	876 924	1 800	48	2 304	1 390 1 406	2 796	16	256	1 642 1 658	3 300	16	256	1 856 1 658	3 514	198	39 204	
L_7	370 380	750	10	100	800 830	1 630	30	900	1 410 1 420	2 830	10	100	1 490 1 600	3 090	110	12 100	1 970 2 050	4 020	80	6 400	
L_8	391 390	781	1	1	804 807	1 611	3	9	1 484 1 484	2 968	0	0	1 541 1 522	3 063	19	361	2 110 2 078	4 188	32	1 024	
L_9	402 40 /	809	5	25	844 817	1 661	27	729	1 424 1 410	2 834	14	196	1 514 1 544	3 058	30	900	2 071 2 166	4 237	95	9 025	Total
Total(1)	–	7 188	–	–	–	15 119	–	–	–	26 115	–	–	–	28 787	–	–	–	36 919	–	–	–
Σw^2	–	–	–	1 384	–	–	–	5 121	–	–	–	5 077	–	–	–	18 149	–	–	–	61 663	91 394
Total(2)	–	7 188	–	–	–	15 119	–	–	–	22 685	–	–	–	24 941	–	–	–	36 919	–	–	
Σw^2	–	–	–	1 384	–	–	–	5 121	–	–	–	2 577	–	–	–	18 149	–	–	–	61 663	

(1) Sur la totalité des résultats

(2) Sans les résultats éliminés à la suite des tests (laboratoire 1, niveaux M_3 et M_4)

Ce graphique montre, de façon assez suggestive, que :

– La dispersion des moyennes (liée à la reproductibilité) a tendance à augmenter avec le niveau,

– Aux niveaux 3, 4 et 5 le laboratoire L_1 peut être suspecté de donner des résultats anormalement élevés, et au niveau M_5, le laboratoire L_6 de donner des résultats anormalement faibles.

Nous laissons au lecteur le soin d'appliquer le test de Shapiro et Wilk aux totaux (ou moyennes) correspondant aux 5 niveaux.

Les conclusions sont les suivantes.

L'hypothèse de normalité doit être rejetée aux niveaux M_3 et M_4 ; à ces niveaux, on est donc conduit à éliminer le laboratoire L_1 (il est inutile, d'après le graphique ci-dessus, de recommencer les tests après élimination de L_1).

L'hypothèse de normalité n'est pas rejetée aux niveaux M_1 et M_2 (comme on pouvait s'y attendre d'après le graphique).

Elle n'est pas non plus rejetée au niveau M_5 ce qui paraît a priori étonnant – en fait cela résulte de ce que les résultats apparamment aberrants (L_6 et L_1) "encadrent" le groupe des 7 autres résultats.

En définitive, on est conduit à rejeter les seuls résultats du laboratoire L_1 aux niveaux 3 et 4.

Reproductibilité en fonction du niveau

Les carrés moyens entre laboratoires se calculent aisément à partir des données du Tableau 6 (en ne tenant pas compte des données encadrées qui sont éliminées). Par exemple, au niveau 1 (9 laboratoires) :

$$M_{L,1} = \frac{1}{8}\left[\frac{[883]^2 + \ldots + [809]^2}{2} - \frac{(7\,188)^2}{18}\right]$$

On trouve ainsi :

$M_{L,1}$	= 936	avec	8	degrés de liberté
$M_{L,2}$	= 6 543	"	8	"
$M_{L,3}$	= 3 045	"	7	"
$M_{L,4}$	= 5 561	"	7	"
$M_{L,5}$	= 59 643	"	8	"

Il apparaît clairement que le carré moyen entre laboratoires dépend du niveau et croît avec celui-ci. En toute rigueur, les tests de Cochran et de Hartley ne sont pas applicables, les nombres de degrés de liberté différant légèrement d'un niveau à l'autre ; on peut cependant remarquer que

$$\frac{M_{L,5}}{M_{L,1}} \# 64$$

alors que la table 1 (test de Hartley) donne

pour $k = 5$ $\nu = 8$ $q_{0,99} = 13{,}2$

pour $k = 5$ $\nu = 7$ $q_{0,99} = 16{,}5$

Estimation des variances intra et interlaboratoires et de la variance de reproductibilité à chaque niveau

Les carrés moyens intralaboratoires $M_{r,1}, \ldots, M_{r,5}$ s'obtiennent à partir des quantités Σw^2 qui figurent à la dernière ligne du Tableau 6, ces quantités étant divisées par 2p, p étant le nombre de laboratoires auxquels elles s'appliquent (p = 9, pour les niveaux 1, 2, 5, p = 8 pour les niveaux 3, 4). Par exemple :

$$M_{r,1} = \frac{1\,384}{18} = 77$$

Les carrés moyens entre laboratoires $M_{L,1}, \ldots, M_{L,5}$ ont été calculés ci-dessus.

Le tableau ci-après donne, pour chaque niveau :

l'estimation $s_r^2 = M_r$ de la variance intralaboratoire σ_r^2 (répétabilité)

l'estimation $s_L^2 = \dfrac{M_L - M_r}{2}$ de la variance entre laboratoires σ_L^2

l'estimation $s_R^2 = \dfrac{M_L + M_r}{2}$ de la variance totale σ_R^2 (reproductibilité)

	M_1	M_2	M_3	M_4	M_5
$M_r = s_r^2$	77	285	161	1 134	3 426
M_L	936	6 543	3 045	5 561	59 643
$\dfrac{M_L - M_r}{2} = s_L^2$	430	3 129	1 442	2 213	28 108
$\dfrac{M_L + M_r}{2} = s_R^2$	507	3 414	1 603	3 348	31 534

Estimation de la répétabilité et de la reproductibilité à chaque niveau

Le tableau ci-après donne les valeurs des écarts-types s_r et s_R *dans l'unité initiale* (c'est-à-dire après division par 100 des quantités s_L et s_R résultant du tableau ci-dessus) ainsi que les valeurs de la répétabilité et de la reproductibilité

$$r = 2{,}8\ s_r \qquad R = 2{,}8\ s_R$$

	M_1	M_2	M_3	M_4	M_5
s_r s_R	0,088 0,225	0,169 0,584	0,127 0,400	0,337 0,579	0,585 1,775
r R	0,25 0,63	0,47 1,63	0,36 1,12	0,94 1,62	1,64 4,97
Moyenne m des résultats(1)	3,99	8,40	14,18	15,59	20,51

(1) déduite de l'avant dernière ligne du Tableau 6.

La variation de r et R en fonction de m est donnée par le graphique ci-après.

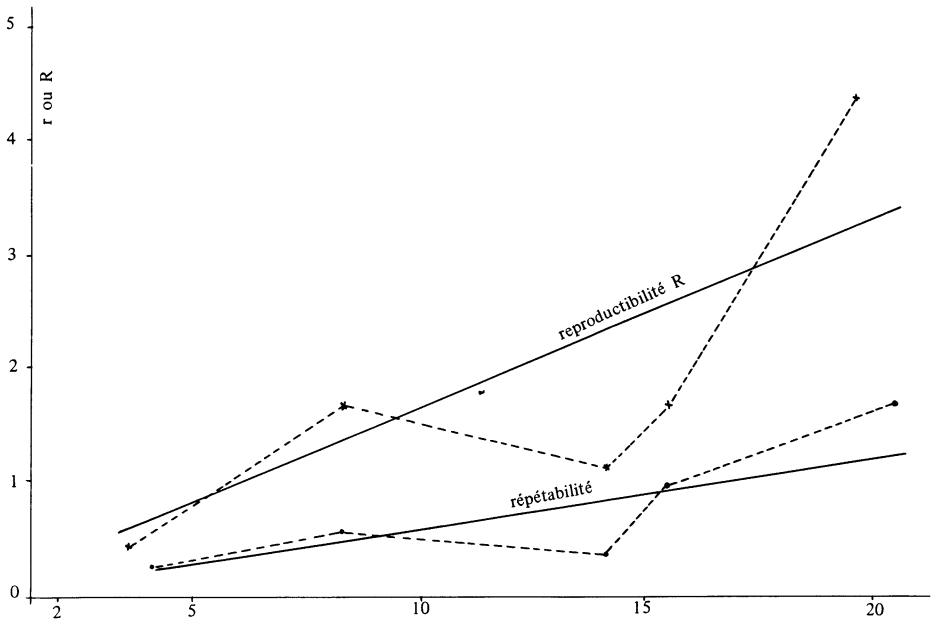
En première approximation –compte tenu de l'erreur qui affecte les estimations de r et R– on peut admettre que la répétabilité et la reproductibilité varient proportionnellement au niveau.

L'ajustement linéaire classique –représenté par les droites en trait plein du graphique– conduit aux relations suivantes :

$$r = 0{,}06 \text{ m}$$

$$R = 0{,}17 \text{ m}$$

En fait, l'ajustement est médiocre –surtout en ce qui concerne la reproductibilité– et il y aurait lieu de s'interroger sur l'origine du "creux" qui se manifeste au niveau M_3.



5 – QUELQUES CAS D'UTILISATION PRATIQUE DE LA REPETABILITE ET DE LA REPRODUCTIBILITE [1], [13]

Soient r et R la répétabilité et la reproductibilité d'une méthode d'essai, valables dans une certaine plage des valeurs de la grandeur étudiée. Les § ci-après ne sont applicables que pour des mesures effectuées à l'intérieur de cette plage, et sous condition que l'erreur intralaboratoire (variance σ_r^2) et l'erreur totale (variance $\sigma_R^2 = \sigma_r^2 + \sigma_L^2$) soient distribuées suivant des lois normales ou approximativement normales.

5.1 – Résultats obtenus dans un seul laboratoire

5.1.1 – *Comparaison de deux ou plusieurs échantillons*

Soit w l'écart (valeur absolue de la différence) entre les résultats obtenus sur deux échantillons.

Les échantillons pourront être considérés comme identiques si $w \leqslant r$.

Si le laboratoire a fait des mesures sur n échantillons, dont les résultats sont $y_1, y_2, \ldots y_n$, les échantillons peuvent être considérés comme identiques si $w = y_{max} - y_{min}$ est tel que :

$$w \leqslant A_n r$$

A_n est un coefficient déduit de la loi de distribution de l'étendue (cas de la loi normale) ; il est donné dans la table ci-après.

n	2	3	4	5	6	7	8	9	10
A_n	1	1,17	1,28	1,36	1,42	1,47	1,52	1,55	1,53

5.1.2 – *Intervalle de confiance attaché à une moyenne de résultats*

Le laboratoire, dont la répétabilité est supposée correcte, a obtenu sur le même produit, à des intervalles de temps rapprochés, n résultats dont la moyenne est $\bar{y}$.

L'intervalle de confiance bilatéral pour la vraie valeur m de ce produit est approximativement, au niveau de confiance 95 % :

$$\bar{y} \pm 2\sqrt{s_L^2 + \frac{s_r^2}{n}}$$

Tenant compte des relations

$$s_R^2 = s_L^2 + s_r^2$$

$$r = 2{,}8\ s_r$$

$$R = 2{,}8\ s_R$$

cette expression s'écrit :

$$\bar{y} \pm 0{,}7\sqrt{R^2 - \left(1 - \frac{1}{n}\right)r^2}$$

Un calcul analogue montre que les intervalles unilatéraux, au même niveau de confiance,sont approximativement :

$$m \leqslant \bar{y} + 0{,}6\sqrt{R^2 - \left(1 - \frac{1}{n}\right)r^2} \qquad m \geqslant \bar{y} - 0{,}6\sqrt{R^2 - \left(1 - \frac{1}{n}\right)r^2}$$

Au niveau de confiance 99 %, les coefficients 0,7 (cas bilatéral) et 0,6 (cas unilatéral) sont à remplacer respectivement par 0,9 et 0,8.

5.2 – Résultats obtenus dans plusieurs laboratoires

5.2.1 – *Accord entre deux ou plusieurs laboratoires*

Soit w l'écart entre les résultats obtenus sur le même échantillon ou deux échantillons aussi identiques que possible par deux laboratoires. On peut considérer qu'il y a accord entre ces résultats si $w \leqslant R$.

Dans le cas de n laboratoires obtenant chacun un résultat sur le même échantillon ou des échantillons aussi identiques que possible, on peut considérer qu'il y a accord entre ces résultats si l'étendue $w = y_{max} - y_{min}$ est telle que :

$$w \leqslant A_n \, R$$

A_n étant le coefficient donné par la table du § 5.1.1.

5.2.2 – *Intervalle de confiance attaché à une moyenne de résultats*

On suppose qu'à partir du même échantillon ou d'échantillons aussi identiques que possible, chaque laboratoire d'un groupe de laboratoires (dont la répétabilité et la reproductibilité sont correctes) a obtenu n résultats. La moyenne générale des np résultats est $\bar{y}$.

L'intervalle de confiance bilatéral pour la vraie valeur m du produit est approximativement, au niveau de confiance 95 % :

$$\bar{y} \pm 2\sqrt{\frac{s_L^2}{p} + \frac{s_r^2}{np}}$$

Tenant compte des relations exprimées en 5.1.2, cette expression s'écrit :

$$\bar{y} \pm 0{,}7\sqrt{\frac{R^2}{p} - \left(1 - \frac{1}{n}\right)\frac{r^2}{p}} \qquad \bar{y} \pm 0{,}7\frac{R}{\sqrt{p}} \quad \text{si } n = 1$$

Les intervalles de confiance unilatéraux, au même niveau de confiance, sont approximativement :

$$m \leqslant \bar{y} + 0{,}6\sqrt{-\frac{R^2}{p} - \left(1 - \frac{1}{n}\right)\frac{r^2}{p}} \qquad m \geqslant \bar{y} - 0{,}6\sqrt{\frac{R^2}{p} - \left(1 - \frac{1}{n}\right)\frac{r^2}{p}}$$

Au niveau de confiance 99 %, les coefficients 0,7 (cas bilatéral) et 0,6 (cas unilatéral) sont à remplacer respectivement par 0,9 et 0,8.

TABLES STATISTIQUES

Table 1. – Table pour le test de Hartley
Table 2. – Table pour le test de Cochran
Table 3. – Table de la varianble χ^2 (test de Bartlett)

Table 1

Table pour le test de Hartley

Valeurs de q au seuil 5 % ($q_{0,95}$)

ν \ k	2	3	4	5	6	7	8	9	10	11	12
2	39,0	87,5	142,0	202,0	266,0	333,0	403,0	475,0	550,0	626,0	704,0
3	15,44	26,6	36,8	46,9	55,1	63,8	72,1	80,5	87,4	93,7	101,0
4	9,60	14,8	19,2	23,1	26,7	29,9	33,2	36,2	38,5	40,9	43,4
5	7,15	10,4	12,9	15,3	17,3	19,1	20,7	22,2	23,6	25,0	26,0
6	5,82	8,14	10,0	11,6	12,9	14,1	15,0	16,1	17,0	17,8	18,5
7	4,99	6,81	8,13	9,34	10,3	11,1	11,9	12,6	13,2	13,7	14,4
8	4,43	5,90	6,99	7,88	8,61	9,32	9,90	10,4	10,8	11,4	11,8
9	4,03	5,22	6,13	6,90	7,55	8,10	8,59	8,95	9,31	9,68	10,1
10	3,72	4,77	5,54	6,18	6,70	7,18	7,55	7,86	8,17	8,50	8,85
12	3,28	4,10	4,71	5,21	5,58	5,93	6,23	6,49	6,69	6,96	7,17
15	2,86	3,49	3,94	4,31	4,57	4,86	5,05	5,21	5,37	5,58	5,76
20	2,46	2,92	3,25	3,49	3,71	3,86	4,02	4,14	4,26	4,35	4,48
30	2,08	2,39	2,59	2,75	2,89	3,00	3,10	3,16	3,22	3,29	3,35
60	1,67	1,84	1,95	2,03	2,10	2,16	2,20	2,25	2,29	2,32	2,34
∞	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

Valeurs de q au seuil 1 % ($q_{0,99}$)

ν \ k	2	3	4	5	6	7	8	9	10	11	12
2	199	448	729	1 036	1 362	1 705	2 063	2 432	2 813	3 204	3 605
3	47,5	85	120	151	184	216	249	281	310	337	361
4	23,2	37	49	59	69	79	89	97	106	113	120
5	14,9	22	28	33	38	42	46	50	54	57	60
6	11,1	15,5	19,1	22	25	27	30	32	34	36	37
7	8,89	12,1	14,5	16,5	18,4	20	22	23	24	26	27
8	7,50	9,9	11,7	13,2	14,5	15,8	16,9	17,9	18,9	19,8	21
9	6,54	8,5	9,9	11,1	12,1	13,1	13,9	14,7	15,3	16,0	16,6
10	5,85	7,4	8,6	9,6	10,4	11,1	11,8	12,4	12,9	13,4	13,9
12	4,91	6,1	6,9	7,6	8,2	8,7	9,1	9,5	9,9	10,2	10,6
15	4,07	4,9	5,5	6,0	6,4	6,7	7,1	7,3	7,5	7,8	8,0
20	3,32	3,8	4,3	4,6	4,9	5,1	5,3	5,5	5,6	5,8	5,9
30	2,63	3,0	3,3	3,4	3,6	3,7	3,8	3,9	4,0	4,1	4,2
60	1,96	2,2	2,3	2,4	2,4	2,5	2,5	2,6	2,6	2,7	2,7
∞	1,00	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0

Table 2

Table pour le test de Cochran

Valeurs de g au seuil 5 % ($g_{0,95}$)

k \ ν	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	0,999	0,975	0,939	0,906	0,877	0,853	0,833	0,816	0,801	0,788	0,734	0,660	0,581	0,500
3	0,967	0,871	0,798	0,746	0,707	0,677	0,653	0,633	0,617	0,603	0,547	0,475	0,403	0,333
4	0,907	0,768	0,684	0,629	0,590	0,560	0,537	0,518	0,502	0,488	0,437	0,372	0,309	0,250
5	0,841	0,684	0,598	0,544	0,507	0,478	0,456	0,439	0,424	0,412	0,365	0,307	0,251	0,200
6	0,781	0,616	0,532	0,480	0,445	0,418	0,398	0,382	0,368	0,357	0,314	0,261	0,212	0,167
7	0,727	0,561	0,480	0,431	0,397	0,373	0,354	0,338	0,326	0,315	0,276	0,228	0,183	0,143
8	0,680	0,516	0,438	0,391	0,360	0,336	0,319	0,304	0,293	0,283	0,246	0,202	0,162	0,125
9	0,639	0,478	0,403	0,358	0,329	0,307	0,290	0,277	0,266	0,257	0,223	0,182	0,145	0,111
10	0,602	0,445	0,373	0,331	0,303	0,282	0,267	0,254	0,244	0,235	0,203	0,166	0,131	0,100
12	0,541	0,392	0,326	0,288	0,262	0,244	0,230	0,219	0,210	0,202	0,174	0,140	0,110	0,083
15	0,471	0,335	0,276	0,242	0,220	0,203	0,191	0,182	0,174	0,167	0,143	0,114	0,089	0,067
20	0,389	0,271	0,221	0,192	0,174	0,160	0,150	0,142	0,136	0,130	0,111	0,088	0,068	0,050
24	0,343	0,235	0,191	0,166	0,149	0,137	0,129	0,122	0,116	0,111	0,094	0,074	0,057	0,042
30	0,293	0,198	0,159	0,138	0,124	0,114	0,106	0,100	0,096	0,092	0,077	0,060	0,046	0,033
40	0,237	0,158	0,126	0,108	0,097	0,089	0,083	0,078	0,075	0,071	0,060	0,046	0,035	0,025
60	0,174	0,113	0,090	0,077	0,068	0,062	0,058	0,055	0,052	0,050	0,041	0,032	0,023	0,017
120	0,100	0,063	0,050	0,042	0,037	0,034	0,031	0,029	0,028	0,027	0,022	0,017	0,012	0,008
∞	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Valeurs de g au seuil 1 % ($g_{0,99}$)

k \ ν	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	0,9999	0,995	0,979	0,959	0,937	0,917	0,899	0,882	0,867	0,854	0,795	0,707	0,606	0,500
3	0,993	0,942	0,883	0,834	0,793	0,761	0,734	0,711	0,691	0,674	0,606	0,515	0,423	0,333
4	0,968	0,864	0,781	0,721	0,676	0,641	0,613	0,590	0,570	0,554	0,488	0,406	0,325	0,250
5	0,928	0,789	0,696	0,633	0,588	0,553	0,526	0,504	0,485	0,470	0,409	0,335	0,264	0,200
6	0,883	0,722	0,626	0,564	0,520	0,487	0,461	0,440	0,423	0,408	0,353	0,286	0,223	0,167
7	0,838	0,664	0,569	0,508	0,466	0,435	0,411	0,391	0,375	0,362	0,311	0,249	0,193	0,143
8	0,795	0,615	0,521	0,463	0,423	0,393	0,370	0,352	0,337	0,325	0,278	0,221	0,170	0,125
9	0,754	0,573	0,481	0,425	0,387	0,359	0,338	0,321	0,307	0,295	0,251	0,199	0,152	0,111
10	0,718	0,536	0,447	0,393	0,357	0,331	0,311	0,295	0,281	0,270	0,230	0,181	0,138	0,100
12	0,653	0,475	0,392	0,343	0,310	0,286	0,268	0,254	0,242	0,232	0,196	0,154	0,116	0,083
15	0,575	0,407	0,332	0,288	0,259	0,239	0,223	0,210	0,200	0,192	0,161	0,125	0,093	0,067
20	0,480	0,330	0,265	0,229	0,205	0,188	0,175	0,165	0,157	0,150	0,125	0,096	0,071	0,050
24	0,425	0,287	0,230	0,197	0,176	0,161	0,150	0,141	0,134	0,128	0,106	0,081	0,060	0,042
30	0,363	0,241	0,191	0,164	0,145	0,133	0,123	0,116	0,110	0,105	0,087	0,066	0,048	0,033
40	0,294	0,192	0,151	0,128	0,114	0,103	0,096	0,090	0,085	0,082	0,067	0,050	0,036	0,025
60	0,215	0,137	0,107	0,090	0,080	0,072	0,067	0,063	0,059	0,057	0,046	0,034	0,025	0,017
120	0,123	0,076	0,059	0,049	0,043	0,039	0,036	0,033	0,032	0,030	0,024	0,018	0,013	0,008
∞	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 3

Table de la variable χ^2 (Test de Bartlett)

ν	$\chi^2_{0,95}$	$\chi^2_{0,99}$
1	3,84	6,63
2	5,99	9,21
3	7,81	11,34
4	9,49	13,28
5	11,07	15,09
6	12,59	16,81
7	14,07	18,47
8	15,51	20,09
9	16,92	21,67
10	18,31	23,21
11	19,67	24,72
12	20,03	26,22
13	22,36	27,69
14	23,68	29,14
15	25,00	30,58
16	26,30	32,00
17	27,59	33,41
18	28,87	34,80
19	30,14	36,19
20	31,41	37,57
21	32,67	38,93
22	33,92	40,29
23	35,17	41,64
24	36,41	42,98
25	37,65	44,31
26	38,88	45,64
27	40,11	46,96
28	41,34	48,28
29	42,56	49,59
30	43,77	50,89

REFERENCES

[1] AFNOR. – NF X 06 041. Fidélité d'une méthode d'essai. Guide pour les essais interlaboratoires (1970)

[2] ASTM. STP 335 – Manual for conducting an Interlaboratory Study of a Test Method (1963)

[3] The Institute of Petroleum. – Recommended practice for determining precision data for methods on petroleum products and lubricants – Londres 1968.

[4] John MANDEL. – Repeatability and Reproductibility – Journal of Quality Technology – Vol. 4 N° 2 (1972)

[5] E. MORICE. – Test de normalité d'une distribution observée – Revue de Statistique Appliquée – Vol. XX N° 2 (1972).

[6] SHAPIRO S.S. et WILK M.B. – An analysis of variance test for normality. Biometrika (1965) – Vol 52, part 3 et 4, p. 591/611.

[7] BIOMETRIKA TABLES FOR STATISTICIANS – Vol. 1 – Table 31

[8] COCHRAN W.G. – The distribution of the largest of a set estimated variances as a function of their total. Annals of Eugenics – Vol. 11 (1941).

[9] BARTLETT. – Jour. Royal Statist. Soc. Suppl 4 (1937)

[10] BIOMETRIKA TABLES FOR STATISTICIANS – Vol. 1, Tables 32 et 33.

[11] HARJAAE E. – Biometrika (1969), Vol 56 p. 276/7.

[12] DIXON et MASSEY. – Introduction to Statistical Analysis.

[13] AFNOR. – NF M07 033.