

REVUE DE STATISTIQUE APPLIQUÉE

J. FLESSELLES

**Un mode de présentation des résultats d'essais comparatifs
entre laboratoires portant sur l'évaluation de la
qualité d'une méthode de mesure**

Revue de statistique appliquée, tome 28, n° 3 (1980), p. 29-43

http://www.numdam.org/item?id=RSA_1980__28_3_29_0

© Société française de statistique, 1980, 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/>

UN MODE DE PRÉSENTATION DES RÉSULTATS D'ESSAIS COMPARATIFS ENTRE LABORATOIRES PORTANT SUR L'ÉVALUATION DE LA QUALITÉ D'UNE MÉTHODE DE MESURE

J. FLESSELLES

Ingénieur en Chef, (SEITA) Direction du Développement

SOMMAIRE

Pour émettre un jugement précis sur la validité d'une méthode de mesure on est fréquemment conduit à organiser des expériences comparatives entre laboratoires, lesquelles permettent de déterminer la *répétabilité* et la *reproductibilité* de la méthode testée, paramètres qui caractérisent sa qualité.

A partir de développements théoriques simples, illustrés par un exemple concret correspondant à une méthode de détermination de taux d'humidité dans le tabac, on montre l'intérêt d'une présentation des résultats d'analyse statistique, spécifique de ce type d'expérience comparative, qui permet notamment de définir l'*ajustement* et la *cohérence* de chaque laboratoire par rapport à l'ensemble de ceux qui ont pris part à l'expérience : Un diagramme "en cible" est proposé qui permet d'illustrer de façon parlante et synthétique la situation relative de chacun des laboratoires par rapport à l'optimum.

I. ASPECTS GENERAUX DE LA METHODE

I.1. POSITION DU PROBLEME

Il est fréquent que des expériences collectives soient organisées entre divers laboratoires afin d'estimer la *répétabilité* et la *reproductibilité* d'une méthode de mesure.

Un document très complet a été publié en 1970 par l'AFNOR sous la référence NF X 06-041, exposant les théories statistiques auxquelles il est fait appel en la matière et décrivant en détail les techniques d'analyse et d'interprétation à mettre en oeuvre dans les divers cas qui peuvent se présenter.

Nous renvoyons le lecteur intéressé à cette norme fondamentale qui constitue un outil de travail de qualité pour résoudre ce type de problème.

Nous extrayons simplement de cette norme, afin de fixer les idées, les quelques définitions suivantes :

— La *fidélité* d'une méthode de mesure caractérise "l'étroitesse de l'accord entre les valeurs expérimentales obtenues au cours d'un ensemble d'expériences faites dans des conditions déterminées.

— La *reproductibilité* est la “fidélité déterminée dans le cas d’opérateurs travaillant dans des laboratoires différents ou dans le même laboratoire à des époques différentes, chacun d’eux obtenant avec la même méthode des résultats individuels sur un produit identique soumis à l’essai”.

Ces définitions appellent les commentaires suivants :

a) Les notions de *répétabilité* et de *reproductibilité* définies ci-dessus, sont en pratique concrétisées par la détermination de deux paramètres numériques, ayant les dimensions d’une variance (ou de l’écart-type correspondant qui est sa racine carrée), dénommés : variance de répétabilité et variance de reproductibilité.

Ce sont ces deux paramètres numériques qui *caractérisent* en quelque sorte la *qualité de la méthode*, c’est-à-dire son *adéquation* au but recherché, compte tenu des niveaux de précision recherchés.

Ces deux variances se déterminent à partir des résultats d’analyse de variances classiques déduits du plan d’expérience monté à cette fin.

La nature des calculs à effectuer dépend de la structure de ce plan et en particulier du nombre de niveaux (un ou plusieurs) considérés au cours de l’expérience collective.

b) Les définitions mêmes de la répétabilité et de la reproductibilité impliquent, en théorie du moins, que les essais soient conduits sur des *produits identiques* et cela est d’ailleurs fort compréhensible :

La qualité de la mesure elle-même doit s’exprimer de façon absolue, c’est-à-dire en éliminant les fluctuations dues à la variabilité de l’échantillon soumis au test.

Il existe des cas où il est effectivement possible de conduire les expériences avec des échantillons réellement identiques. Ces cas sont cependant assez rares, soit que la mesure soit destructive, soit qu’il faille, pour des raisons pratiques, disposer simultanément de plusieurs échantillons.

Or on ne peut jamais être certain que des échantillons rendus aussi identiques que possible les uns des autres le soient absolument. On peut même être certain du contraire.

La fluctuation entre échantillons est quelquefois importante, selon la nature du produit dont ils sont extraits.

C’est précisément l’un des rôles du plan d’expérience, généralement construit avec des *répétitions*, c’est-à-dire la conduite d’une série de mesures portant sur une même modalité (même appareil, même opérateur, même moment ou moment consécutif), de permettre de saisir et de déterminer la variabilité due à l’échantillonnage.

Nous n’insisterons pas davantage sur ces techniques maintenant assez largement répandues et utilisées, et nous nous tournerons vers un des aspects connexes de ce type d’étude.

1.2. COMPARAISONS ENTRE LABORATOIRES

Au cours des analyses de résultats mentionnées ci-dessus, on est conduit à comparer les laboratoires entre eux, notamment pour éliminer ceux dont les résultats sont aberrants, c’est-à-dire s’écartent des résultats moyens d’une valeur

telle qu'avec un haut niveau de probabilité on doit conclure qu'ils ne sont pas alignés avec l'ensemble du groupe participant à l'expérience.

De même, certains laboratoires alignés correctement peuvent néanmoins être aussi considérés comme aberrants, dans la mesure où la dispersion de leurs résultats est significativement trop élevée, comparée à celle des autres.

Enfin, dans les expériences à plusieurs niveaux, c'est-à-dire portant sur plusieurs valeurs possibles du paramètre, on peut être également conduit, par des méthodes d'analyses adéquates, à examiner la qualité de la mesure soit niveau par niveau, soit globalement pour l'ensemble des niveaux.

Parallèlement aux objectifs initiaux, qui consistent, nous le rappelons, à préciser la qualité de la méthode, on peut se proposer également de voir comment se répartissent les différents laboratoires les uns par rapport aux autres. Nous suggérons à cette fin un mode de présentation des résultats qui permet de visualiser de façon très concrète les positions des laboratoires respectifs, par l'introduction des notions d'*ajustement* et de *cohérence* que nous précisons ci-après.

Afin de rendre notre exposé plus parlant, nous allons nous placer dans un cas concret. Nous développerons les aspects théoriques de nos propositions. Nous illustrerons ensuite notre propos par un exemple tiré d'une expérience réellement conduite.

I.3. CADRE DE L'EXPERIENCE

Soient k laboratoires participant à une expérience collective destinée à tester une méthode de détermination de la teneur en eau du tabac.

Chaque résultat, exprimé en %, est ce qu'on désigne habituellement par "taux d'humidité".

Nous supposerons en outre que l'expérience comporte n niveaux, c'est-à-dire qu'à chaque laboratoire est fourni une série de n échantillons différents les uns des autres par leur teneur en eau, mais "identiques" chacun à chacun pour un même niveau dans les différents laboratoires.

Sur chacun des n échantillons, chaque laboratoire procède à r mesures ou répétitions.

Nous avons ainsi une expérience factorielle classique

$$k \times n \times r .$$

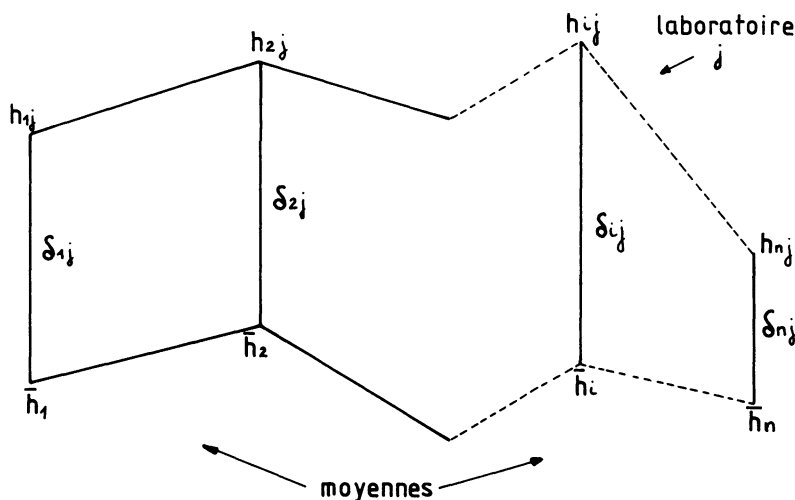
I.4. Définition de l'ajustement et de la cohérence

Désignons par $\bar{h}_1, \bar{h}_2, \dots, \bar{h}_i, \dots, \bar{h}_n$, les moyennes respectives de k laboratoires pour chacun des n niveaux (ou variantes).

Soit $\bar{h}_{ij}$ la valeur correspondant à la moyenne des r répétitions du laboratoire j pour le niveau i .

Posons
$$\delta_{ij} = \bar{h}_{ij} - \bar{h}_i \quad (i = 1, n)$$

qui représente l'écart du laboratoire j à la moyenne du groupe pour le niveau i



et

$$\delta_j = \frac{1}{n} \sum_i \delta_{ij} \quad (1)$$

δ_j est l'écart arithmétique moyen du laboratoire par rapport à la moyenne, sur l'ensemble des niveaux ou variantes.

Posons d'autre part

$$E_j^2 = \frac{1}{n} \sum_{i=1}^n \delta_{ij}^2 \quad (2)$$

E est une sorte d'écart quadratique moyen, chaque écart étant rapporté non pas à une même base, mais à la valeur du niveau correspondant.

Il résulte des formules classiques d'analyse de la variance et des formules de définition (1) et (2) ci-dessus que E_j^2 se décompose en la somme de deux carrés.

En effet, soit $C_j^2 = \frac{1}{n} \sum (\delta_{ij} - \delta_j)^2$

C_j^2 représente la variance des écarts δ_{ij} pour le laboratoire j .

Développant cette expression, il vient

$$C_j^2 = \frac{1}{n} (\sum \delta_{ij}^2 + n \delta_j^2 - 2 \delta_j \sum \delta_{ij})$$

Comme

$$\sum \delta_{ij} = n \bar{\delta}_j \text{ selon (1)}$$

il vient

$$C_j^2 = \frac{1}{n} (\sum \delta_{ij}^2 + n \bar{\delta}_j^2 - 2 n \bar{\delta}_j^2) = \frac{1}{n} (\sum \delta_{ij}^2 - n \bar{\delta}_j^2)$$

ou encore

$$C_j^2 = E_j^2 - \bar{\delta}_j^2$$

Soit encore de façon générale, quel que soit le laboratoire j

$$E^2 = \bar{\delta}^2 + C^2$$

Nous désignerons δ par *écart d'ajustement*

et C par *écart de cohérence*.

Ces désignations se justifient par les considérations suivantes :

δ caractérise l'écart moyen du laboratoire considéré par rapport à la moyenne du groupe.

Si $\delta = 0$, le laboratoire est, en moyenne sur les n niveaux, ajusté à la moyenne du groupe. Concrètement, il n'est pas décalé en moyenne. C est l'écart-type des écarts de ce laboratoire considérés sur l'ensemble des niveaux.

Plus C est faible, plus l'écart moyen δ précédent est significatif, autrement dit, plus le décalage observé est cohérent d'un niveau à un autre.

A la limite, si $C = 0$, cela signifie que le laboratoire reproduit les résultats moyens du groupe pour chacun des niveaux avec un décalage fixe.

Cette notion de cohérence est très importante en pratique.

Il est en effet relativement fréquent que des laboratoires travaillent soigneusement, mais qu'une erreur de manipulation ou de lecture entraîne un décalage systématique : il peut être utile d'apprécier ce décalage et d'estimer sa signification. Un cas concret, typique de l'exemple que nous présentons plus loin, porte sur la lecture d'un volume d'eau dans une recette graduée. Si les règles de repérage mutuel du ménisque et de la graduation ne sont pas ou suffisamment précisées ou observées, de tels décalages peuvent se manifester de manière systématique.

I.5. Représentation graphique des résultats

Dans un plan orthonormé gradué respectivement en δ de $-\infty$ à $+\infty$ et en C de 0 à $+\infty$ (C étant strictement > 0), on porte l'ensemble des points correspondant à chaque laboratoire repéré par ses deux coordonnées δ_j et C_j .

L'origine 0 des coordonnées correspondant à la *moyenne* des laboratoires participants, on complète le diagramme par un faisceau de cercles concentriques de centre 0, correspondant à des valeurs de E en progression arithmétique.

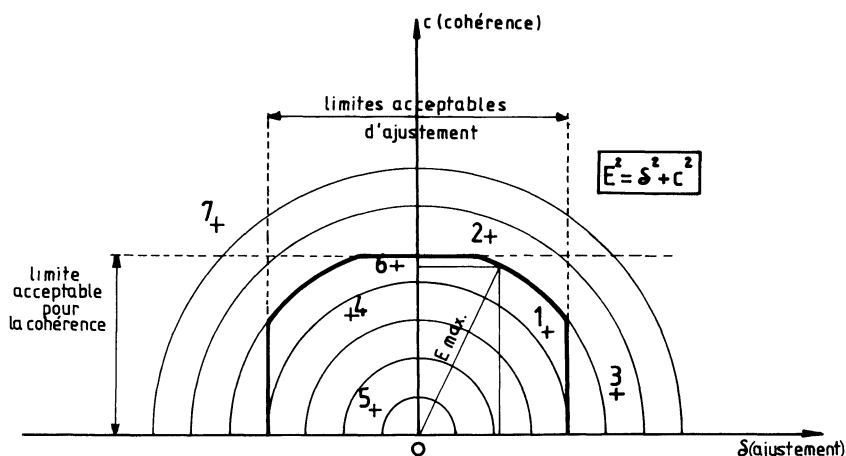
Ce schéma-cible visualise de manière assez parlante l'ensemble des résultats.

L'écart quadratique E_j , caractéristique de chaque laboratoire, est ipso facto décomposé en ses composants δ et C du fait de la relation

$$E_j^2 = \delta_j^2 + C_j^2.$$

Sur ce diagramme on peut également porter les limites acceptables pour $|\delta|$ et C , ainsi que le cercle de rayon E_M correspondant à la valeur maximale admissible pour l'écart quadratique total.

Ainsi se trouve présentée globalement la situation de l'ensemble des laboratoires quant à leur ajustement par rapport à la moyenne et la cohérence de cet ajustement selon les niveaux.



Exemple de diagramme cible

A titre d'exemple on peut déduire du diagramme présenté les conclusions suivantes :

- le laboratoire 5 est celui qui est le plus proche de la moyenne du groupe (E minimal),
- le laboratoire 2 est bien ajusté mais présente un écart de cohérence trop important (la fluctuation de ses écarts à la moyenne dépend fortement des niveaux considérés),
- le laboratoire 3, au contraire, est cohérent, mais son décalage est excessif,
- le laboratoire 7 est à la fois trop décalé et peu cohérent.

Les autres laboratoires constituent le groupe acceptable, leurs points figuratifs se trouvant à l'intérieur de la zone admissible.

II. CAS CONCRET, APPLICATION PRATIQUE

II.1. Données de l'expérience collective

12 laboratoires ont pris part à une expérience collective (*) portant sur l'évaluation d'une méthode de détermination du *taux d'humidité du tabac*. Il s'agit de la méthode par distillation azéotropique au cyclohexane proposée par le projet de norme ISO n° 6488 .

Chacun d'eux a travaillé sur *trois types de tabac* donnant 3 niveaux d'humidité différents. Pour chaque combinaison laboratoire x traitement, chacun d'eux a procédé à 5 répétitions.

(*) organisée par le CORESTA (Centre de Coopération pour les Recherches Scientifiques relatives au Tabac).

TABLEAU I

Expérience collective 1979. Détermination du taux d'humidité par distillation au cyclohexane

N° lab.	Lot A					Lot B					Lot C				
	Résultats individuels					Résultats individuels					Résultats individuels				
	X moy.	σ				X moy.	σ				X moy.	σ			
1	14,9	14,6	14,6	14,6	14,6	14,66	0,134	12,3	12,5	12,5	12,5	12,4	0,089	15,1	14,8
2	14,7	14,7	14,8	14,7	14,8	14,74	0,055	12,6	12,6	12,7	12,6	12,6	0,045	15,0	14,9
3	14,2	14,3	14,4	14,4	14,2	14,30	0,100	12,2	12,2	12,2	12,2	12,2	0	14,6	14,5
4	14,6	14,2	14,1	13,85	14,15	14,18	0,271	11,3	11,95	11,65	12,0	11,55	0,290	13,75	14,35
5	14,0	13,8	13,9	14,0	14,1	13,96	0,114	11,8	11,9	11,5	11,6	11,6	0,164	14,2	14,1
6	14,75	14,75	14,8	14,75	14,75	14,76	0,022	12,5	12,5	12,45	12,5	12,5	0,022	15,0	15,0
7	15,0	14,8	15,0	14,6	14,4	14,76	0,261	12,0	12,0	11,8	12,0	11,8	0,110	15,0	15,0
8	14,6	14,5	14,5	14,4	14,7	14,54	0,114	12,6	12,2	12,3	12,2	12,3	0,164	15,0	14,7
9	14,7	14,6	14,6	14,6	14,6	14,62	0,045	12,5	12,5	12,6	12,5	12,6	0,055	14,7	14,7
10	14,4	14,2	—	—	—	14,30	0,141	12,2	12,3	—	—	—	0,071	14,6	14,4
11	14,7	14,6	14,5	14,7	14,7	14,64	0,089	12,3	12,3	12,3	12,3	12,4	0,045	14,9	14,7
12	14,8	14,8	14,6	14,6	14,6	14,68	0,110	12,9	12,4	12,8	12,6	12,4	0,228	15,4	15,0
Moyennes pondérées															
						14,52								12,26	14,72

L'expérience est donc du type factoriel $12 \times 3 \times 5$.

Remarque : Cette situation ne correspond pas totalement au cas théorique que nous avons défini précédemment, en ce sens que les trois niveaux d'humidité, ou "variantes de traitement", ne résultent pas uniquement d'une variation d'humidité provoquée sur des échantillons identiques, mais du choix délibéré d'échantillons de tabac de variétés différentes, dont les humidités se trouvent différentes. Dans le cas de l'expérience, deux des moyennes sont d'ailleurs voisines.

Cette circonstance n'empêche cependant pas d'appliquer la méthode de décomposition des écarts en ajustement et cohérence telle que nous l'avons décrite.

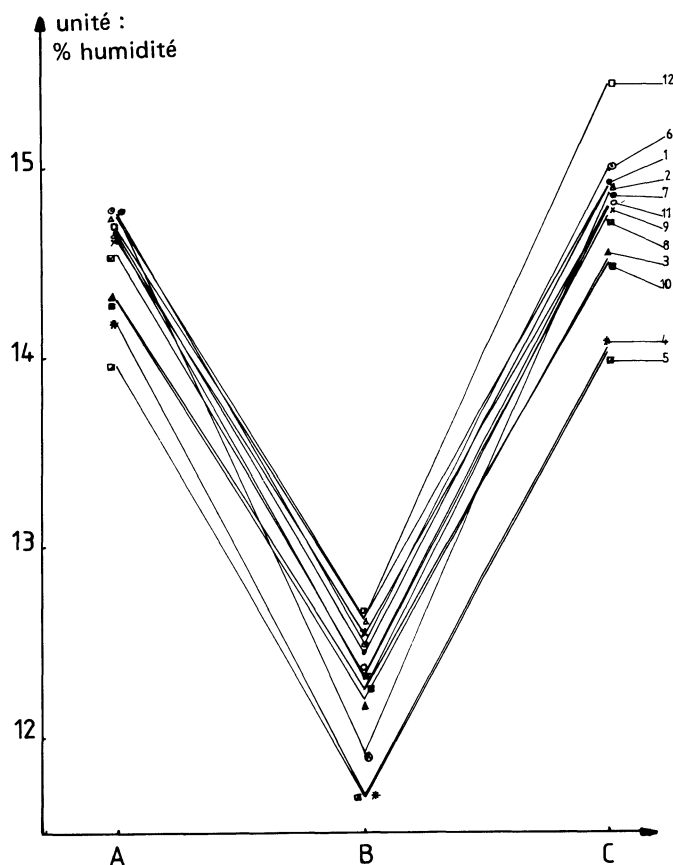
II.2. Résultats

Le tableau I donne les résultats individuels et les moyennes trouvées pour chaque variante "traitement" ou "lot" et chaque combinaison "laboratoire $\times$ traitement".

Ces résultats sont interprétés par les graphiques ① et ②.

Expérience collective CORESTA 1979

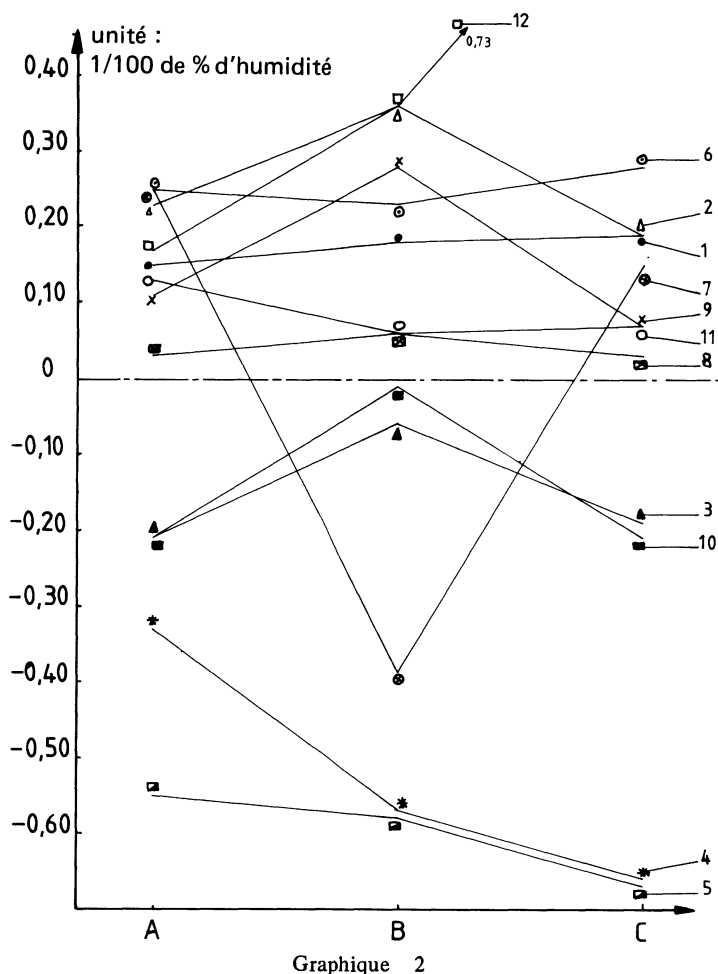
Taux d'humidité



Graphique 1

Expérience collective CORESTA 1979

Taux d'humidité



Le graphique ② n'est qu'une transposition du graphique ① qui prépare à la décomposition des écarts.

Au lieu de porter les valeurs de l'humidité elles-mêmes, on porte les écarts à la moyenne du lot correspondant. Ceci permet en outre de rendre le graphique plus clair et de dégager visuellement certains faits marquants.

Ainsi, la ligne 0 correspond, *pour chaque lot*, à la valeur de sa moyenne.

Les observations les plus directes que l'on puisse faire sont les suivantes :

a) deux laboratoires apparaissent nettement détachés du groupe : les laboratoires 4 et 5,

b) le faisceau des autres courbes est à la fois groupé et présente des sous-ensembles d'allures parallèles,

c) le laboratoire n° 7, quoique ajusté en moyenne par rapport à la ligne moyenne, présente pour le lot B un résultat apparemment "peu cohérent" avec

l'ensemble ; il en est de même du laboratoire n° 12, dont le point représentatif du lot C est hors épure.

II.3. Analyse des résultats

Ces observations demandent naturellement à être précisées et fondées sur une analyse rigoureuse.

L'analyse de variance classique, effectuée lot par lot, donne les résultats suivants :

	Source	Somme des carrés	D.D.L.	Carré moyen	Rapport F
Lot A	Q_L : entre laborat.	3,65	11	0,3318	16,43
	Q_r : répétitions	0,91	45	0,0202	
	Q_T : total	4,56	56		
Lot B	Q_L : entre laborat.	6,05	11	0,5502	13,32
	Q_r : répétitions	1,86	45	0,0413	
	Q_T : total	7,93	56		
Lot C	Q_L : entre laborat.	8,26	11	0,7509	28,88
	Q_r : répétitions	1,17	45	0,0260	
	Q_T : total	9,43	56		

Degrés de liberté $p = 12$ (laborat.) $\rightarrow p - 1 = 11$
 $N = 57$ (résultats) $\rightarrow N - 1 = 56$

Elle met principalement en évidence l'effet laboratoire très significatif qui est assez bien illustré par le graphique ① mentionné précédemment.

Par ailleurs, l'observation a) ci-dessus peut être testée ce qui permet de considérer du point de vue des moyennes, les laboratoires 4 et 5 comme effectivement excentrés.

Le tableau II montre le calcul correspondant.

Analyse sur les moyennes : Elimination des laboratoires excentrés

Considérons le tableau des moyennes par lot et par laboratoire et calculons les moyennes par laboratoire qui figurent en dernière colonne.

Les valeurs moyennes qui correspondent aux laboratoires 4 et 5 sont anormalement basses, respectivement 13,31 et 13,22.

Les observations b) et c) émises au paragraphe 2° ci-dessus sont alors testées par la décomposition en ajustement et cohérence. (Les moyennes prises en compte sont celles qui résultent de l'élimination des laboratoires 4 et 5).

Le tableau III développe ce calcul. Il donne pour les 10 laboratoires retenus, c'est-à-dire non éliminés pour défaut d'ajustement global, les écarts respectifs exprimés en 1/100 de point, entre les valeurs de chaque laboratoire et la moyenne de celles-ci pour chaque lot A, B, C.

On observe alors que si les ajustements globaux sont satisfaisants (0,26 point pour le laboratoire le plus décentré et 0,02 point pour le mieux ajusté), il n'en est pas de même de la cohérence. Pour ce paramètre, les deux laboratoires n° 7 et n° 12 ont des écarts de cohérence C^2 anormalement élevés.

TABLEAU II

Lab.	$\bar{A}$	$\bar{B}$	$\bar{C}$	$\bar{M}$
1	14,66	12,44	14,90	14,00
2	14,74	12,62	14,90	14,09
3	14,30	12,20	14,52	13,67
4	14,18	11,69	14,05	13,31
5	13,96	11,68	14,04	13,22
6	14,76	12,49	14,99	14,08
7	14,76	11,92	14,86	13,84
8	14,54	12,32	14,74	13,87
9	14,62	12,54	14,78	13,98
10	14,30	12,25	14,50	13,68
11	14,64	12,32	14,78	13,91
12	14,68	12,62	15,44	14,25
Moy.	14,51	12,26	14,71	13,83
Moy. sans 4 et 5	14,60	12,37	14,84	13,94

A l'ensemble des laboratoires restant (10 laboratoires) correspond une moyenne de 13,94 et un écart-type de 0,1824

$$t_{0,95} (9 \text{ dd}\ell) = 2,26$$

Intervalle de confiance pour une valeur:

$$2,26 \times 0,1824 = 0,412$$

Tout laboratoire dont la moyenne est extérieure à l'intervalle

$$13,94 \pm 0,412$$

doit être considéré comme n'appartenant pas à cette population.

Intervalle de confiance à 95 % de probabilité:

$$13,53 \% - 14,35 \%$$

Les laboratoires 4 et 5 ont des résultats *extérieurs* à cet intervalle. On les élimine pour l'analyse ultérieure.

TABLEAU III

N°	δa	δb	δc	δ	C^2	C
1	6	7	6	6,333	0,22	0,47
2	14	25	6	15	60,7	7,8
3	- 30	- 17	- 32	- 26,333	44,2	6,6
6	16	12	15	14,333	2,9	1,7
7	16	- 45	2	- 9	680,7	26,0
8	- 6	- 5	- 10	- 7	4,7	2,2
9	2	17	- 6	4,333	90,9	9,5
10	- 30	- 12	- 26	- 22,66	59,8	7,7
11	4	- 5	- 6	- 2,3	20,2	4,5
12	8	25	60	31	468,7	21,6
Moyenne de référence pour le calcul de δ	14,60	12,37	14,84	13,94	35,4	

Unité : 1/100 de pt

C^2 moyen hors (7)
et (12)

Ceci s'explique d'ailleurs parfaitement par l'examen attentif des résultats, la valeur trouvée par le laboratoire n° 7 pour le lot B étant anormalement basse (non cohérente avec le niveau moyen du laboratoire) et la valeur trouvée par le laboratoire n° 12 pour le lot C étant au contraire anormalement élevée.

On peut d'ailleurs tester ces variances en formant le rapport F entre celles-ci et la valeur moyenne de C^2 qui vaut : $C^2 \text{ moyen} = 35,4$.

Les rapports F des variances estimées (à 16 ddl) valent respectivement 19,2 et 13,2 et dépassent largement la valeur théorique au seuil de probabilité à 95 % de F à 2/16 ddl, soit $F = 3,63$

$$\begin{array}{l} \text{[ddl du numérateur} \quad 3 - 1 = 2 \\ \text{'' du dénominateur } (3 - 1) \times 8 = 16] \end{array}$$

On constate donc effectivement que les laboratoires 7 et 12 sont caractérisés par des valeurs de C^2 très significativement supérieures à la valeur moyenne de C^2 des autres laboratoires.

II.4. Illustration des résultats

Si l'on porte alors le point représentatif de chaque laboratoire dans le plan δ , C, on obtient le *diagramme cible* n° 3.

Cette représentation permet de visualiser de façon synthétique la situation de l'ensemble des laboratoires dans le cadre de l'expérience.

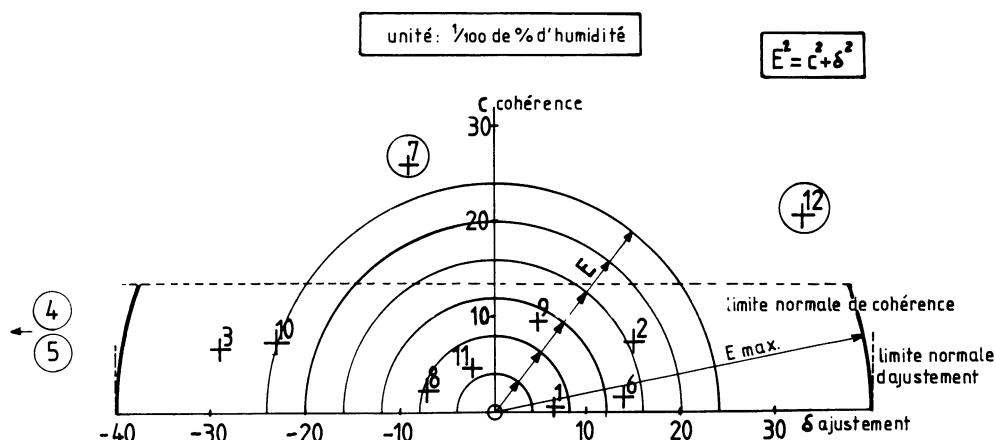
Plus le point représentatif est proche du centre de la cible, meilleure est sa position en ajustement et cohérence.

Tout écart parallèle à l'axe des abscisses s'interprète comme un défaut d'ajustement positif ou négatif.

Tout écart positif parallèle à l'axe des ordonnées s'interprète comme un défaut de cohérence.

Expérience collective CORESTA 1979-Détermination du taux d'humidité -

Ajustement et cohérence interne des laboratoires



Tout point situé à l'intérieur du cercle limite de rayon $E_{\max}$ doit être considéré comme acceptable. Au contraire, les points extérieurs à ce cercle correspondent à des laboratoires dont les résultats sont globalement sujets à caution.

II.5. Analyse définitive

Les laboratoires 7 et 12 étant non retenus dans l'analyse définitive pour défaut de cohérence, il subsiste huit laboratoires.

Nous ne conservons pas également le laboratoire n° 10, ce pour de simples raisons techniques : ce laboratoire n'a fourni que deux valeurs (au lieu de cinq) par combinaison labo x lot et cette circonstance rend difficilement praticable son inclusion dans le plan d'analyse de variance à deux facteurs contrôlés avec répétitions (on peut en ce cas procéder à une analyse lot par lot, comme il l'a été fait précédemment, mais non à une analyse globale).

L'analyse de variance effectuée sur les laboratoires restants

1, 2, 3, 6, 8, 9, 11

donne les résultats suivants :

Source	Somme des carrés	Degrés de liberté	Carré moyen	Rapport F
Lot	122.662	2	61.330	7.479
Laboratoires	1.909	6	0.318	38.8
Intéraction	0.201	12	0.017	2.05
Résiduelle	0.688	84	0.0082	—
Total	125.460	104	1.206	—

Commentaires

L'effet lot apparaît très hautement significatif : cela est logique puisque la différence entre lots existe par construction.

On note encore une différence significative entre laboratoires. Elle est d'amplitude faible, mais elle est significative.

Les moyennes générales des laboratoires sont les suivantes :

Labo. n°	Moyenne
1	14,00
2	14,09
3	13,67
6	14,08
8	13,87
9	13,98
11	13,91

Les écarts maximaux imputables *au hasard* entre deux laboratoires sont facilement déduits de la variance résiduelle $S_E^2 = 0.0082$.

Variance d'une moyenne de $3 \times 5 = 15$ mesures (3 lots $\times$ 5 répétitions)

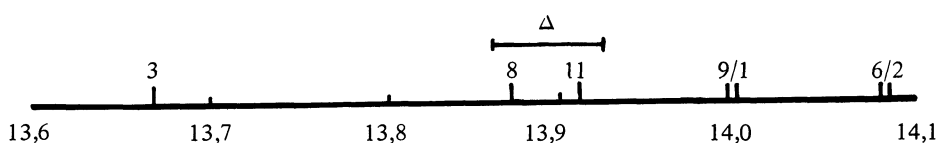
$$S_M^2 = \frac{0.0082}{15}$$

Variance de la différence :

$$2 \times \frac{0.0082}{15}$$

Valeur de $t_{0.95}$ à 84 ddl = 1.99

$$\Delta = 1.99 \times \sqrt{\frac{0.0082 \times 2}{15}} = 0.066$$



Malgré la disjonction significative entre laboratoires, six d'entre eux sont très groupés.

Enfin, on observe que l'interaction lot $\times$ labo. apparaît comme significative, mais assez faiblement.

$$F_{0.95}(12/84) = 1.89 \text{ et } F_{0.99}(12/84) = 2.44.$$

Estimations définitives des variances de répétabilité et de reproductibilité

A) La variance de *répétabilité* est donnée par

$$V(r) = \frac{Q_r}{pq(n-1)} (*)$$

où Q_r est la somme des carrés résiduelle $Q_r = 0.688$

p le nombre de laboratoires

$p = 7$

q le nombre de niveaux, c'est-à-dire de lots

$q = 3$

$pq(n-1) = 84$

n le nombre de répétitions

$n = 5$

$$V(r) = \frac{Q_r}{84}$$

Ecart-type de répétabilité

$$\sigma(r) = \sqrt{V(r)} = 0.090$$

B) La variance de *reproductibilité* est donnée par

$$V(R) = \frac{Q_L + Q_I}{nq(p-1)} + \frac{Q_r}{npq} (*)$$

(*) Voir NF X 06-041. Note technique n° 2.

où Q_L est la somme des carrés "laboratoires"
 Q_I la somme des carrés "d'interaction"
 Q_r , n , p , q ayant les significations précédentes.

$$nq(p-1) = 15 \times 6 = 90 \quad npq = 105$$

d'où

$$V(R) = \frac{1.909 + 0.201}{90} + \frac{0.688}{105} = 0.029996$$

Ecart-type de reproductibilité

$$\sigma(R) = \sqrt{V(R)} = 0.17$$

L'analyse est ainsi poussée à son terme.

III. CONCLUSION

Nous avons montré que, dans une expérience factorielle entre plusieurs laboratoires destinée à déterminer les variances de répétabilité et de reproductibilité, on pouvait, parallèlement aux calculs menés de façon classique pour aboutir à ces valeurs, effectuer une décomposition de la variance des résultats propre à éclairer le jugement global à émettre sur la qualité des résultats fournis par chaque laboratoire, lorsque la grandeur mesurée a été traitée à plusieurs niveaux.

Introduisant les notions d'ajustement et de cohérence, nous avons ainsi proposé une représentation graphique assez parlante de la situation déduite des résultats, laquelle permet de juger d'un coup d'œil de la position de chacun des laboratoires quant à ces deux critères.

Cette méthode est suffisamment générale pour s'appliquer dans la quasi totalité des expériences collectives entre laboratoires portant sur l'agrément d'un procédé de mesure, si la grandeur mesurée fait l'objet d'essais à plusieurs niveaux.