REVUE DE STATISTIQUE APPLIQUÉE

F. BASTENAIRE

Estimation d'une relation structurale et comparaison de deux méthodes de mesure d'une même grandeur

Revue de statistique appliquée, tome 3, n° 2 (1955), p. 83-99

<http://www.numdam.org/item?id=RSA_1955__3_2_83_0>

ESTIMATION D'UNE RELATION STRUCTURALE ET COMPARAISON DE DEUX MÉTHODES DE MESURE D'UNE MÊME GRANDEUR

par

F. BASTENAIRE

Ingénieur de l'École de Physique et Chimie Industrielles
Statisticien à l'Institut de Recherches de la Sidérurgie

Il arrive très souvent que l'on puisse mesurer une même caractéristique (physique ou chimique par exemple) par plusieurs méthodes différentes. Leurs valeurs respectives dépendent évidemment non seulement de leurs reproductibilités mais aussi de leurs sensibilités.

Entre une première méthode, dont les résultats sont très peu dispersés autour d'une valeur moyenne presque indépendante de la caractéristique étudiée, et une méthode pour laquelle des mesures répétées sont entachées d'erreurs importantes, mais dont la moyenne est très sensible aux variations de cette même caractéristique, il se peut qu'il faille préférer la seconde.

Quel que soit le critère adopté pour faire un choix, il est donc nécessaire de tenir compte de la sensibilité en même temps que de la précision.

Lorsque ce choix est d'autre part basé sur une expérience, des erreurs d'échantillonnage interviennent pour une part plus ou moins large et affectent en même temps les mesures des précisions, des sensibilités et de leurs rapports; aussi ne peut-on éviter de rechercher d'abord une estimation de la « relation structurale », de calculer sa variance et de prendre en considération les incertitudes attachées au critère adopté.

Cet article a été divisé en trois parties :

1re *PARTIE :*

ESTIMATION d'une RELATION STRUCTURALE entre DEUX GRANDEURS dont les MESURES sont ENTACHÉES d'ERREURS.

2e *PARTIE :*

COMPARAISON de DEUX MÉTHODES de MESURE d'une MÊME GRANDEUR.

3e *PARTIE :*

APPLICATION à la COMPARAISON de DEUX ESSAIS de COKES MÉTALLURGIQUES.

L'auteur croit devoir avertir les lecteurs qu'il a fait quelques approximations et que, dans certains cas, il a fait appel à des solutions intuitives mais non démontrées pour parvenir à un résultat lui semblant plus proche de la vérité. Ce manque de rigueur mathématique n'affecte que les quotients tandis que les expressions données pour les espérances mathématiques, les variances et les covariances des numérateurs et dénominateurs sont exactes.

Le problème posé présentant un intérêt pratique, une solution même approximative a semblé préférable à l'absence de solution.

D'autre part, l'exposé ici présenté a été limité aux grandes lignes du raisonnement nécessaire pour l'établissement des résultats. L'auteur tient à la disposition du lecteur soucieux de la rigueur un additif mentionné dans le texte à diverses reprises.

PREMIÈRE PARTIE

ESTIMATION D'UNE RELATION STRUCTURALE ENTRE DEUX GRANDEURS DONT LES MESURES SONT ENTACHÉES D'ERREURS

1. - INTRODUCTION

Soient x et y deux grandeurs liées par la relation linéaire :

$$y = Ax + B$$

et soient X et Y leurs mesures.

Nous admettrons que celles-ci sont entachées d'erreurs ε et η accidentelles, en entendant par là que ε et η sont des variables normales de moyenne nulle, d'écart-type σ_ε et σ_η et sont indépendantes l'une de l'autre en probabilité.

Entre les variables :

$$X = x + \varepsilon \; ; \qquad Y = y + \eta \; ,$$

il n'y a plus une relation fonctionnelle comme entre x et y mais une relation statistique qui peut être caractérisée par son coefficient de corrélation et ses deux droites de régression par exemple. Cependant, le plus souvent, ces deux droites sont distinctes et aucune d'entre elles ne correspond à la relation "vraie" entre x et y.

Lorsque le choix des valeurs de x définissant un ensemble de points expérimentaux a lieu au hasard, selon un processus indépendant de l'observateur, les fluctuations des coefficients de corrélation et de régression dues au hasard de l'échantillonnage suivent des lois de probabilité connues et ces derniers peuvent être estimés et accompagnés d'intervalles de confiance. S'ils n'indiquent rien sur la relation fonctionnelle entre x et y, ce sont, du moins, des caractéristiques stables de la liaison.

Il n'en va plus de même lorsque le choix des valeurs de x est plus ou moins systématique et dépend en partie de l'expérimentateur. Dans ce cas, il n'y a plus de permanences statistiques et les variations des caractéristiques de liaison peuvent largement dépasser celles que l'on peut attendre des effets du hasard.

Nous montrerons que, si l'on possède des estimations de σ_ε et de σ_η, on peut retrouver des estimations de A et de B, et que lorsque X et Y sont deux mesures différentes d'une même grandeur, la connaissance de A et des écarts-types σ_ε et σ_η permet de déterminer objectivement quelle est la meilleure des deux méthodes de mesure.

2. - RÉSUMÉ

Il est toujours possible d'écrire la relation linéaire entre x et y sous la forme paramétrique :

$$(1) \qquad \begin{cases} x = at + x_0 \\ y = bt + y_0 \end{cases}$$

et nous verrons d'ailleurs que l'estimation de la pente $\frac{b}{a}$ de la droite ne dépend pas de la représentation paramétrique choisie.

Nous supposerons possible de réaliser un échantillonnage en grappes, c'est-à-dire de répéter un certain nombre k de mesures de x et y pour une même valeur de t afin de pouvoir estimer les écarts-types σ_ε et σ_η et pour simplifier les calculs, nous supposerons k constant.

Soient alors t_i ($i = 1, 2, \ldots, n$) les valeurs du paramètre t au cours de n expériences différentes et soient X_{ij} et Y_{ij} les résultats des lectures de la $j^{ème}$ observation ($j = 1, 2, \ldots, k$) effectuée pour $t = t_i$.

Nous écrirons :

$$X_{ij} = a\,t_i + x_0 + \varepsilon_{ij} \tag{2}$$

$$Y_{ij} = b\,t_i + y_0 + \eta_{ij} \tag{2'}$$

où ε_{ij} et η_{ij} sont les erreurs commises dans l'observation de :

$$x_{ij} = a\,t_i + x_0 \tag{3}$$

et de

$$y_{ij} = b\,t_i + y_0 \tag{3'}$$

Soient $X_{i.}$ et $Y_{i.}$ les moyennes arithmétiques :

$$X_{i.} = \frac{X_{i1} + X_{i2} + \ldots + X_{ik}}{k}$$

$$Y_{i.} = \frac{Y_{i1} + Y_{i2} + \ldots + Y_{ik}}{k}$$

et $X_{..}$ et $Y_{..}$ les moyennes générales :

$$X_{..} = \frac{1}{nk} \sum_{i,j} X_{ij}$$

$$Y_{..} = \frac{1}{nk} \sum_{i,j} Y_{ij}$$

On peut, quelle que soit l'hypothèse relative aux t_i utiliser comme estimations des variances σ_ε^2 et σ_η^2 les expressions :

$$s_\varepsilon^2 = \frac{1}{n(k-1)} \sum_{i,j} (X_{ij} - X_{i.})^2$$

$$s_\eta^2 = \frac{1}{n(k-1)} \sum_{i,j} (Y_{ij} - Y_{i.})^2$$

qui suivent des lois de probabilités de la forme $\frac{\sigma_\varepsilon^2\,\chi^2_{n(k-1)}}{n(k-1)}$ et $\frac{\sigma_\eta^2\,\chi^2_{n(k-1)}}{n(k-1)}$ et ont par suite toutes les propriétés habituelles des variances.

Pour estimer le carré $\frac{b^2}{a^2}$ de la pente de la droite représentative de la liaison structurale, on prendra le rapport $\frac{S_b^2}{S_a^2}$ où :

$$S_a^2 = \sum_i (X_{i.} - X_{..})^2 - \frac{(n-1)}{nk(k-1)} \sum_{i,j} (X_{ij} - X_{i.})^2$$

$$S_b^2 = \sum_i (Y_{i.} - Y_{..})^2 - \frac{(n-1)}{nk(k-1)} \sum_{i,j} (Y_{ij} - Y_{i.})^2$$

Le calcul pratique des quantités $\sum_i (X_{i.} - X_{..})^2$ et $\sum_{i,j} (X_{ij} - X_{i.})^2$ est très simple et s'effectue exactement comme dans une analyse de variance à une seule sorte de catégories.

La loi de probabilité de $\frac{S_b^2}{S_a^2}$ varie selon que l'on suppose les quantités t_i aléatoires ou certaines. Cette dernière hypothèse revenant à étudier les variations de $\frac{S_b^2}{S_a^2}$ en supposant qu'on repète des ensembles d'observations pour un système fixe des t_i.

Dans les deux cas, le quotient $\frac{S_b^2}{S_a^2}$ est une estimation de $\frac{b^2}{a^2}$ qui possède un biais tendant vers zéro quand n tend vers l'infini, biais que l'on peut estimer au moyen de l'expression :

$$\frac{S_b^2\, s_\varepsilon^2}{S_a^4} \left[\frac{4}{k} + \frac{2(n-1)}{k(k-1)} \frac{s_\varepsilon^2}{S_a^2} \right]$$

La variance de $\frac{S_b^2}{S_a^2}$ est approchée par l'expression :

$$\frac{S_b^4}{S_a^4}\left[\frac{4}{k}\left(\frac{s_\varepsilon^2}{S_a^2}+\frac{s_\eta^2}{S_b^2}\right)+\frac{2(n-1)}{k(k-1)}\left(\frac{s_\varepsilon^4}{S_a^4}+\frac{s_\eta^4}{S_b^4}\right)\right]$$

Cette variance tend, comme le biais, vers zéro lorsque n augmente indéfiniment et est de l'ordre de $\frac{1}{n}$ (car S_a^2 et S_b^2 sont proportionnels à n).

La distribution du rapport $\frac{S_b^2}{S_a^2}$ est asymptotiquement normale lorsque les distributions des différentes variables le sont et les limites de confiance du rapport $\frac{b}{a}$ sont obtenues en extrayant les racines carrées des limites de confiance du rapport $\frac{S_b^2}{S_a^2}$ (1).

3. - PROPRIÉTÉS DE L'ESTIMATION $\frac{S_b^2}{S_a^2}$

On peut écrire les identités classiques de la théorie de l'analyse de la variance(2) :

$$\sum_{i,j}(X_{ij}-X_{..})^2=\sum_{i,j}(X_{ij}-X_{i.})^2+k\sum_{i,j}(X_{i.}-X_{..})^2 \qquad (4)$$

$$\sum_{i,j}(Y_{ij}-Y_{..})^2=\sum_{i,j}(Y_{ij}-Y_{i.})^2+k\sum_{i,j}(Y_{i.}-Y_{..})^2 \qquad (4')$$

où

$$X_{i.}=\frac{\sum_j X_{ij}}{k} \qquad \text{et} \qquad X_{..}=\frac{\sum_{i,j} X_{ij}}{nk}$$

et de même pour Y.

D'après (2) et (2') on a :

$$\sum_{i,j}(X_{ij}-X_{i.})^2=\sum_{i,j}(\varepsilon_{i,j}-\varepsilon_{i.})^2 \qquad (5)$$

$$\sum_{i,j}(Y_{ij}-Y_{i.})^2=\sum_{i,j}(\eta_{i,j}-\eta_{i.})^2 \qquad (5')$$

et

$$\sum_i(X_{i.}-X_{..})^2=\sum_i\left[a(t_i-t_.)+(\varepsilon_{i.}-\varepsilon_{..})\right]^2 \qquad (6)$$

$$\sum_i(Y_{i.}-Y_{..})^2=\sum_i\left[b(t_i-t_.)+(\eta_{i.}-\eta_{..})\right]^2 \qquad (6')$$

En s'appuyant sur les mêmes considérations qu'en analyse de variance, on trouve comme pour la variance intraclass dans une analyse à une sorte de catégories :

$$E\left[\sum_{i,j}(\varepsilon_{ij}-\varepsilon_{i.})^2\right]=n(k-1)\,\sigma_\varepsilon^2 \qquad (7)$$

et

$$E\left[\sum_{i,j}(\eta_{ij}-\eta_{i.})^2\right]=n(k-1)\,\sigma_\eta^2 \qquad (7')$$

D'autre part, le développement du second membre de (6) donne :

$$\sum_i(X_{i.}-X_{..})^2=\sum_i a^2(t_i-t_.)^2+\sum_i a(t_i-t_.)(\varepsilon_{i.}-\varepsilon_{..})+\sum_i(\varepsilon_{i.}-\varepsilon_{..})^2 \qquad (8)$$

(1) La démonstration de la convergence en probabilité, lorsque $n\to\infty$, de la variable $\frac{S_b^2}{S_a^2}$ vers une variable aléatoire distribuée normalement est donnée dans l'additif (appendice 2).

(2) Voir par exemple M.G. KENDALL - The advanced theory of Statistics.

et comme en analyse de variance :

$$E\left[\sum_i k(\varepsilon_{i.} - \varepsilon_{..})^2\right] = (n-1)\,\sigma_\varepsilon^2$$

On peut alors faire deux hypothèses quant aux t_i :

1° - On peut considérer les t_i comme formant un échantillon de n individus tirés au hasard dans la population.

2° - On peut considérer les valeurs des t_i comme constantes.

3. A. - Estimation de $\frac{b}{a}$ lorsque les t_i sont aléatoires

Lorsque les observations sont effectuées pour des valeurs aléatoires de t, nous considérerons les t_i comme un échantillon de n valeurs extraites d'une population normale donnée, de variance σ_t^2 ; Nous supposerons d'autre part que les variables aléatoires t_i sont indépendantes en probabilité des erreurs ε_{ij} et η_{ij}.

Avec ces hypothèses, on a :

$$E\left[\sum a(t_i - t_.)(\varepsilon_{i.} - \varepsilon_{..})\right] = a\sum\left[E(t_i - t_.)\,E(\varepsilon_{i.} - \varepsilon_{..})\right] = 0$$

et par suite, d'après (8)

$$E\left[\sum_i (X_{i.} - X_{..})^2\right] = a^2(n-1)\,\sigma_t^2 + \frac{(n-1)}{k}\,\sigma_\varepsilon^2$$

On en déduit d'après (7) et (5) que l'expression :

$$S_a^2 = \sum(X_{i.} - X_{..})^2 - \frac{(n-1)}{nk(k-1)}\sum_{i,j}(X_{ij} - X_{i.})^2$$

a pour espérance mathématique : $a^2(n-1)\,\sigma_t^2$.

De même :

$$S_b^2 = \sum(Y_{i.} - Y_{..})^2 - \frac{(n-1)}{nk(k-1)}\sum_{i,j}(Y_{ij} - Y_{i.})^2$$

a pour espérance mathématique : $b^2(n-1)\,\sigma_t^2$.

Nous proposons donc de prendre pour estimation de $\frac{b^2}{a^2}$ la quantité $\frac{S_b^2}{S_a^2}$

3.A.1 - ESPÉRANCE MATHÉMATIQUE ET VARIANCE DE $\frac{S_b^2}{S_a^2}$

D'après des résultats connus en analyse de variance, les formes quadratiques $\sum_i (X_{i.} - X_{..})^2$ et $\sum_i (X_{ij} - X_{i.})^2$ sont indépendantes et distribuées comme les variables

$$\left(a^2\sigma_t^2 + \frac{\sigma_\varepsilon^2}{k}\right)\chi_{n-1}^2 \qquad \text{et} \qquad \sigma_\varepsilon^2\,\chi_{n(k-1)}^2 .$$

Comme la variance d'une variable χ_p^2 est égale à 2p, on trouve que la variance $V(S_a^2)$ est :

$$V(S_a^2) = 2\left(a^2\sigma_t^2 + \frac{\sigma_\varepsilon^2}{k}\right)^2(n-1) + \frac{2\,(n-1)^2}{nk^2(k-1)}\,\sigma_\varepsilon^4 \tag{9}$$

et de même :

$$V(S_b^2) = 2\left(b^2\sigma_t^2 + \frac{\sigma_\eta^2}{k}\right)^2(n-1) + \frac{2\,(n-1)^2}{nk^2(k-1)}\,\sigma_\eta^4 \tag{9'}$$

Les variables aléatoires S_a^2 et S_b^2 ne sont pas indépendantes car les t_i figurent à la fois dans ces deux quantités. Il faut donc calculer la covariance $\operatorname{cov}(S_a^2\, S_b^2)$. Nous passerons ce calcul assez long et essentiellement basé sur le fait que l'espérance mathématique d'un produit de variables indépendantes est égal au produit des espérances mathématiques de ces variables.

On trouve :

$$E\left[S_a^2\ S_b^2\right] = a^2 b^2 \sigma_t^4 (n^2-1)$$

$$\mathrm{cov}\left[S_a^2\ S_b^2\right] = E\left[S_a^2\ S_b^2\right] - E\left[S_a^2\right]E\left[S_b^2\right] = 2a^2 b^2 \sigma_t^4 (n-1) \qquad (10)$$

En appliquant les formules d'approximation donnant l'espérance mathématique d'un quotient et sa variance (par développement limité) (1), on a :

$$E\left[\frac{S_b^2}{S_a^2}\right] = \frac{b^2}{a^2}\left[1+\frac{4}{k(n-1)}\ \frac{\sigma_\varepsilon^2}{a^2\sigma_t^2} + \frac{2(1-1/nk)}{k(k-1)(n-1)}\ \frac{\sigma_\varepsilon^4}{a^4\sigma_t^4}\right] \qquad (11)$$

et :

$$V\left[\frac{S_b^2}{S_a^2}\right] = \left(\frac{b}{a}\right)^4\left[\frac{4\left(\frac{\sigma_\eta^2}{b^2\sigma_t^2}+\frac{\sigma_\varepsilon^2}{a^2\sigma_t^2}\right)}{k(n-1)} + \frac{2\left(1-\frac{1}{nk}\right)\left(\frac{\sigma_\eta^4}{b^4\sigma_t^4}+\frac{\sigma_\varepsilon^4}{a^4\sigma_t^4}\right)}{k(n-1)(k-1)}\right] \qquad (12)$$

On constate que le biais de l'estimation de $\frac{b^2}{a^2}$ tend vers zéro comme $\frac{1}{n-1}$ et qu'il en est de même de la variance.

En pratique, pour calculer le biais et la variance de $\frac{S_b^2}{S_a^2}$ on remplaçera les quantités σ_ε^2, σ_η^2, $a^2\sigma_t^2$ et $b^2\sigma_t^2$ intervenant dans ces formules par leurs estimations s_ε^2, s_η^2, $\frac{S_a^2}{n-1}$, $\frac{S_b^2}{n-1}$.

3.A.2 - RELATION ENTRE LA LIAISON FONCTIONNELLE ET L'ELLIPSE DE DISPERSION -

Il est intéressant de déterminer les caractéristiques statistiques de la population constituée par l'ensemble des couples :

$$X_i = a t_i + x_0 + \varepsilon_i$$

$$Y_i = b t_i + y_0 + \eta_i$$

et d'étudier la position de la droite :

$$x = at + x_0$$

$$y = bt + y_0$$

par rapport à la répartition correspondante à deux variables.

On calcule facilement les variances et covariances :

$$V(X) = a^2\sigma_t^2 + \sigma_\varepsilon^2$$

$$V(Y) = b^2\sigma_t^2 + \sigma_\eta^2$$

$$\mathrm{cov}(XY) = a\,b\,\sigma_t^2$$

d'où l'on déduit le coefficient de corrélation :

$$\rho_{XY} = \frac{1}{\sqrt{1+\frac{\sigma_\varepsilon^2}{a^2\sigma_t^2}}\sqrt{1+\frac{\sigma_\eta^2}{b^2\sigma_t^2}}}$$

(1) Ces formules sont : $E\left[\frac{X}{Y}\right] = \frac{E[X]}{E[Y]}(1+\gamma_Y^2 - \rho\gamma_X\gamma_Y)$

$$V\left[\frac{X}{Y}\right] = \left(\frac{E[X]}{E[Y]}\right)^2(\gamma_X^2 - 2\rho\gamma_X\gamma_Y + \gamma_Y^2)$$

où γ_X et γ_Y sont les coefficients de variation de X et Y.

et les coefficients de régression :

$$b_{Y/X} = \frac{b}{a}\left(\frac{1}{1+\dfrac{\sigma_\varepsilon^2}{a^2\,\sigma_t^2}}\right),$$

$$b_{X/Y} = \frac{a}{b}\left(\frac{1}{1+\dfrac{\sigma_\eta^2}{b^2\,\sigma_t^2}}\right).$$

On voit immédiatement que la condition nécessaire et suffisante pour que :

$$b_{Y/X} = \frac{b}{a}$$

est que : $\sigma_\varepsilon = 0$,

et que de même, pour que : $b_{X/Y} = \frac{a}{b}$

il faut et il suffit que : $\sigma_\eta = 0$.

On voit de plus que l'estimation fournie par $b_{Y/X}$ diffère de la pente $\frac{b}{a}$ de la liaison fonctionnelle de x et y par une erreur systématique de l'ordre de : $\frac{\sigma_\varepsilon^2}{a^2\,\sigma_t^2}$ en valeur relative.

Ces faits sont bien connus, mais on est parfois tenté, lorsque l'on recherche les caractéristiques d'une liaison structurale dont on postule l'existence, d'identifier cette dernière avec le grand axe de l'ellipse de dispersion. Nous allons donc chercher la condition nécessaire et suffisante pour que le grand axe de l'ellipse soit représentatif de la liaison entre x et y.

La distribution du couple (X, Y) est une distribution normale à deux variables entièrement définie par sa matrice symétrique de variance-covariance :

$$\Lambda = \begin{pmatrix} V(X) & \operatorname{cov}(XY) \\ \operatorname{cov}(XY) & V(Y) \end{pmatrix} = \begin{pmatrix} a^2\,\sigma_t^2 + \sigma_\varepsilon^2 & a\;b\;\sigma_t^2 \\ a^2\;b\;\sigma_t^2 & b^2\,\sigma_t^2 + \sigma_\eta^2 \end{pmatrix}$$

qui est toujours régulière sauf si $\sigma_\varepsilon = \sigma_\eta = 0$ (condition pour que son déterminant s'annule).

La densité de probabilité dans le plan (X, Y) définie par cette distribution est

$$\frac{1}{2\pi}\; e^{-\frac{1}{2}Q(X,Y)}\; dX\,dY\ ,$$

où Q (X,Y) est une forme quadratique que l'on peut représenter matriciellement par

$$Q(X,Y) = \left(\frac{X}{Y}\right)\Lambda^{-1}\,(X,Y)$$

ou en écrivant le vecteur (X, Y) sous forme matricielle :

$$(X,Y) = Z\ ,$$

par

$$Q(X,Y) = Z'\,\Lambda^{-1}\,Z$$

La condition nécessaire et suffisante pour qu'un vecteur P soit l'un des axes principaux de l'ellipse Q (X, Y) = Cte est que P soit vecteur propre de la matrice régulière Λ^{-1} donc qu'il existe un nombre λ non nul tel que :

$$\Lambda^{-1}\,P = \lambda P\ .$$

ce qui est équivalent à :

$$\frac{1}{\lambda}\,P = \Lambda P$$

donc que P soit vecteur propre pour la matrice des moments. Ecrivons donc que la direction (a,b) définissant la liaison fonctionnelle est vecteur propre :

$$\begin{pmatrix} a^2\,\sigma_t^2 + \sigma_\varepsilon^2 & a\,b\,\sigma_t^2 \\ a\,b\,\sigma_t^2 & b^2\,\sigma_t^2 + \sigma_\eta^2 \end{pmatrix} \cdot (a\,,b) = \lambda(a\,,b)$$

soit, après avoir effectué le produit de matrices :

$$\frac{a^3\,\sigma_t^2 + a_\varepsilon^2 + a\,b^2\,\sigma_t^2}{a} = \frac{a^2\,b\,\sigma_t^2 + b^3\,\sigma_t^2 + b\,\sigma_\eta^2}{b} = \lambda$$

et après simplification : $\sigma_\varepsilon^2 = \sigma_\eta^2$

Ainsi, la condition nécessaire et suffisante pour que (a, b) soit l'un des axes principaux de l'ellipse est que les écarts-types des erreurs commises sur X et sur Y soient égaux.

Enfin, on peut déterminer directement les valeurs propres de la matrice Λ en formant le polynôme caractéristique qui est ici une équation du second degré. On constate que la valeur commune λ des rapports précédents est toujours égale à la plus grande des deux racines du polynôme caractéristique (sauf si $\sigma_t^2 = 0$) d'où il résulte que si (a,b) est vecteur propre, il correspond nécessairement au grand axe.

En conclusion :

La condition nécessaire et suffisante pour que le grand axe de l'ellipse de dispersion coincide avec la relation fonctionnelle entre x et y est que les erreurs absolues commises sur x et y aient même écart-type.

3. B. - Estimation de $\frac{b}{a}$ lorsque les t_i sont considérés comme certains

Posons :

$$\sum (t_i - t_.)^2 = (n-1)\,\sigma_t^2$$

et les espérances mathématiques $E(S_a^2)$ et $E(S_b^2)$ garderont la forme :

$$E\left[S_a^2\right] = a^2(n-1)\,\sigma_t^2$$
$$E\left[S_b^2\right] = b^2(n-1)\,\sigma_t^2$$

Par développement de S_a^2 et de S_b^2 et en utilisant les équations (7) et (8), on peut calculer comme précédemment la covariance et l'on trouve cette fois :

$$\operatorname{cov}(S_a^2\;S_b^2) = 0\,.$$

Il reste à calculer $V(S_a^2)$ et $V(S_b^2)$. On constate que S_a^2 et S_b^2 sont la somme de deux quantités indépendantes, les variances des termes de la forme $\sum (X_{ij} - \bar{X}_{i.})^2$ étant déjà connues.

Pour le terme :

$$\sum_i (X_{i.} - X_{..})^2 = \sum_i \left[a\,(t_i - t_.) + (\varepsilon_{i.} - \varepsilon_{..})\right]$$

on peut procéder comme suit : Comme en analyse de variance, on sait qu'il existe un changement de coordonnées qui permet de substituer aux n quantités $\varepsilon_{i.} - \varepsilon_{..}$ liées par la relation $\sum(\varepsilon_{i.} - \varepsilon_{..}) = 0$ des quantités v_1, v_2, ...v_{n-1} en nombre n - 1 seulement, distribuées indépendamment les unes des autres avec pour variance commune $V(\varepsilon_i)$ c'est-à-dire $\frac{\sigma_\varepsilon^2}{k}$.

Le même changement de coordonnées est applicable aux termes $a\,(t_i - t_.)$ et puisque la transformation se réduit à une rotation d'axes rectangulaires, toute somme de carrés est invariante.

Après avoir effectué une homothétie $u_i = \frac{k}{\sigma_\varepsilon} v_i$ (= 1,2,.. n-1) on peut donc poser :

$$\sum (\varepsilon_{i.} - \varepsilon_{..})^2 = \frac{\sigma_\varepsilon^2}{k} \sum_{i=1}^{n-1} u_i^2 \qquad\qquad \sum_i a^2(t_i - t_.)^2 = \frac{\sigma_\varepsilon^2}{k} \sum_{i=1}^{n-1} C_i^2$$

et

$$\sum_i \left[a(t_{i.} - t_{.}) + (\varepsilon_{i.} - \varepsilon_{..})\right]^2 = \frac{\sigma_\varepsilon^2}{k} \sum_{i=1}^{n-1} (C_i + u_i)^2 ,$$

les u_i étant des variables aléatoires normales centrées de variance unité.

Or, on peut facilement déterminer l'espérance mathématique et la variance de $(C_i + u_i)^2$. On a :

$$E\left[(C_i + u_i)^2\right] = C_i^2 + 1$$

et

$$V\left[(C_i + u_i)^2\right] = 4\,C_i^2 + 2$$

Notre somme $\sum_i (X_{i.} - X_{..})^2$ ayant été mise sous la forme d'une somme de variables indépendantes, on en tire :

$$V\left[\sum_i (X_{i.} - X_{..})^2\right] = \frac{\sigma_\varepsilon^4}{k^2}\left[\sum (4\,C_i^2) + 2\,(n-1)\right]$$

Soit :

$$V\left[\sum_i (X_{i.} - X_{..})^2\right] = 2\,(n-1)\,\frac{\sigma_\varepsilon^4}{k^2}\left(1 + \frac{2k\,a^2\,\sigma_t^2}{\sigma_\varepsilon^2}\right)$$

Rappelons que lorsque nous avions supposé les t_i aléatoires, nous avions trouvé :

$$V\left[\left(\sum (X_{i.} - X_{..})\right)^2\right] = 2\,(n-1)\,\frac{\sigma_\varepsilon^4}{k^2}\left(1 + \frac{k\,a^2\,\sigma_t^2}{\sigma_\varepsilon^2}\right)^2$$

La quantité $\frac{a^2\,\sigma_t^2}{\sigma_\varepsilon^2/k}$ est une mesure de l'importance relative des variations réelles du point X, Y observé par rapport aux erreurs commises. En posant $U = \frac{a^2\,\sigma_t^2}{\sigma_\varepsilon^2/k}$ il vient :

$$\rho = \frac{\text{Variance de } \sum_i \left[(X_{i.} - X_{..})^2\right] \text{ pour } t_i \text{ certain}}{\text{Variance de } \sum_i \ (X_{i.} - X_{..})^2 \text{ pour } t_i \text{ aléatoire}} = \frac{1 + 2u}{(1 + U)^2}$$

On constate que ρ est toujours inférieur à 1, ce qui est naturel sauf si $\sigma_t^2 = 0$ auquel cas les deux hypothèses sont équivalentes.

Pour les variances de S_a^2 et S_b^2 on obtient en tenant compte de l'indépendance des deux termes qui y figurent :

$$V\left[S_a^2\right] = 2\,(n-1)\,\frac{\sigma_\varepsilon^4}{k^2}\left(1 + \frac{2\,k\,a^2\,\sigma_t^2}{\sigma_\varepsilon^2}\right) + \frac{2\,(n-1)^2}{n\,k^2(k-1)}\,\sigma_\varepsilon^4 \, . \qquad (13)$$

Par application de la formule donnant l'espérance mathématique d'un quotient on trouve :

$$E\left[\frac{S_b^2}{S_a^2}\right] = \frac{b^2}{a_2}\left[1 + \frac{4}{k(n-1)}\,\frac{\sigma_\varepsilon^2}{a^2\,\sigma_t^2} + \frac{2\,\sigma_\varepsilon^4}{a^4\,\sigma_t^4}\,\frac{(1 - 1/nk)}{k(k-1)(n-1)}\right] . (14)$$

Il ne faut pas s'étonner de retrouver un résultat identique à celui que nous avons obtenu en supposant les t_i aléatoires car si la variance $V\,(S_a^2)$ a changé, la covariance $\mathrm{cov}\,(S_a^2\ S_b^2)$ a également changé.

Pour la variance du rapport $\frac{S_b^2}{S_a^2}$ on retrouve :

$$V\left[\frac{S_b^2}{S_a^2}\right] \frac{b^4}{a^4}\,\frac{4}{k(n-1)}\left(\frac{\sigma_\varepsilon^2}{a^2\,\sigma_t^2} + \frac{\sigma_\eta^2}{b^2\,\sigma_t^2}\right) + \frac{2\,(1 - 1/nk)}{k(k-1)(n-1)}\left(\frac{\sigma_\varepsilon^4}{a^4\,\sigma_t^4} + \frac{\sigma_\eta^4}{b^4\,\sigma_t^4}\right) . \quad (15)$$

DEUXIÈME PARTIE

COMPARAISON DE DEUX MÉTHODES DE MESURE D'UNE MÊME GRANDEUR

1. - INTRODUCTION

Lorsque l'on mesure une grandeur t au moyen de deux appareils ou de deux méthodes de sensibilités différentes, on recueille ainsi deux mesures que, pour de faibles variations de t, nous supposerons fonctions linéaires de t : $at + x_0$, $bt + y_0$ Les coefficients "d'amplification" a et b représentent les sensibilités de ces méthodes, définies comme les rapports des accroissements des mesures correspondant à un accroissement donné de la grandeur t, à savoir $\frac{dx}{dt}$, $\frac{dy}{dt}$.

La valeur d'une méthode de mesure ne dépend pas seulement de sa sensibilité mais aussi de sa fidélité définie comme son aptitude à fournir des indications cohérentes pour des observations répétées d'une même grandeur et elle peut aisément être repérée au moyen de l'écart-type de ces dernières.

Pour choisir entre diverses méthodes, il est évidemment nécessaire de tenir compte à la fois des qualités de sensibilité et de fidélité de celles-ci, mais de quelle façon ?

Si nous supposons connues d'une manière générale les "courbes d'étalonnage" des deux méthodes à étudier, cela signifie que sous l'hypothèse de linéarité nous connaissons a, b, x_0 et y_0. De nos mesures X et Y nous pouvons déduire des estimations de t :

$$\hat{t}_1 = \frac{X - x_0}{a} \; ; \qquad \hat{t}_2 = \frac{Y - y_0}{b} \; ,$$

et puisque X et Y ont respectivement pour variances celles des erreurs commises ε et η dans les mesures de x et y, on a :

$$\sigma^2_{\hat{t}_1} = \frac{\sigma^2_\varepsilon}{a^2} \; ; \qquad \sigma^2_{\hat{t}_2} = \frac{\sigma^2_\eta}{b^2}$$

Nous proposons de définir la meilleure méthode comme étant celle qui fournit une variance $\sigma^2_{\hat{t}}$ minimum.

Ainsi, quand il arrive que diverses méthodes de mesures soient destinées dans l'esprit de leurs auteurs à caractériser une même propriété physique (ou autre) sans que l'on ait nécessairement une idée exacte sur la nature de cette grandeur et sans que l'on connaisse de relation précise entre cette dernière et les résultats obtenus par les diverses méthodes, mais si l'on pense que cette propriété est définie par une caractéristique unique, on peut encore distinguer entre deux méthodes en déterminant la valeur du rapport :

$$\frac{\sigma^2_{\hat{t}_1}}{\sigma^2_{\hat{t}_2}} = \frac{\sigma^2_{\varepsilon/a^2}}{\sigma^2_{\eta/b^2}} \; ,$$

ce qui est possible sans qu'il soit nécessaire de passer par l'intermédiaire des estimations $\hat{t}_1$ et $\hat{t}_2$ mais en se basant seulement sur les estimations de σ^2_ε , σ^2_η et du rapport $\frac{b^2}{a^2}$.

2. - RÉSUMÉ

On suppose l'expérience conduite comme précédemment c'est-à-dire comme pour estimer le rapport $\frac{b^2}{a^2}$. On effectue les mesures par chaque méthode pour n valeurs différentes de la caractéristique t à repérer.

Posons :

$$S_a^2 = \sum_i (X_{i.} - X_{..})^2 - \frac{n-1}{nk(k-1)} \sum_{i,j} (X_{ij} - X_{i.})^2 \ ;$$

$$S_b^2 = \sum_i (Y_{i.} - Y_{..})^2 - \frac{n-1}{nk(k-1)} \sum_{i,j} (Y_{ij} - Y_{i.})^2 \ ;$$

$$s_\varepsilon^2 = \frac{\sum_{i,j} (X_{ij} - X_{i.})^2}{n(k-1)} \ ;$$

$$s_\eta^2 = \frac{\sum_{i,j} (Y_{ij} - Y_{i.})^2}{n(k-1)} \ ;$$

et

$$T = \frac{s_\varepsilon^2 \ S_b^2}{s_\eta^2 \ S_a^2} \ .$$

La loi de probabilité exacte de T est complexe mais le rapport $\frac{s_\varepsilon^2}{s_\eta^2}$ joue pratiquement un rôle prépondérant.

En effet, si $\frac{\sigma_\varepsilon^2}{a^2 \ \sigma_t^2}$ et $\frac{\sigma_\eta^2}{b^2 \ \sigma_t^2}$ sont par exemple tous deux égaux à 0,2 (ce qui représente déjà des mesures assez grossières) et en prenant k = 3, n = 11, on trouve (d'après la première partie, formule 12) :

$$V\left(\frac{S_{b^2}/b^2}{S_{a^2}/a^2}\right) = 0{,}056 \ .$$

Sous l'hypothèse "nulle" où $\sigma_\varepsilon^2 \ b^2 = \sigma_\eta^2 \ a^2$, le rapport $\frac{s_\varepsilon^2 \ b^2}{s_\eta^2 \ a^2}$ suit une loi de F et comme on peut écrire :

$$T = \frac{s_\varepsilon^2 \ b^2}{s_\eta^2 \ a^2} \times \frac{S_{b^2}/b^2}{S_{a^2}/a^2} \ ,$$

il est clair que la loi de probabilité de T est fortement influencée par le premier rapport dont on sait que sa distribution est en même temps très dispersée et très dissymétrique.

Comme nous n'avons pu que calculer une expression approchée de la variance de T en appliquant la formule donnant la variance d'un rapport, nous proposons de prendre pour limites de l'intervalle de confiance de T en vue de tester l'hypothèse $\sigma_\varepsilon^2 \ b^2 = \sigma_\eta^2 \ a^2$ les quantités :

$$L_1 = F\sqrt{\frac{V(T)}{V(F)}} \ ; \qquad L_2 = \frac{1}{L_1}$$

où F désigne la limite de confiance lue dans la table de Snédécor [1] avec n(k-1) = d degrés de liberté au numérateur et au dénominateur.

---

(1) Si la limite F choisie correspond à la probabilité $1-\alpha$, les limites L_1, L_2 définissent un intervalle de confiance auquel est associé une probabilité $1-2\alpha$. Les limites L_1 et L_2 sont inverses l'une de l'autre parce que le nombre de degrés de liberté est le même au numérateur et au dénominateur.

On prendra :

$$V(T)=\frac{4d(d-1)}{(d-2)^2(d-4)}+\frac{4(nk+1)}{n(n-1)(k-1)k}(\alpha+\beta)+\frac{2(nk+1)}{n(n-1)k^2(k-1)}(\alpha^2+\beta^2)+$$

$$+\frac{4(\alpha-\beta)^2}{n^2k^2(k-1)^2},$$

où

$$\alpha=\frac{(n-1)\,s_\varepsilon^2}{S_a^2}\;;\qquad \beta=\frac{(n-1)\,s_\eta^2}{S_b^2}\;;\qquad d=n(k-1)\,.$$

On peut en outre considérer T comme une estimation de $\frac{\sigma_\varepsilon^2/a^2}{\sigma_\eta^2/b^2}$. Cette dernière est légèrement biaisée puisqu'en 1ère approximation :

$$E\left[T\right]=\frac{\sigma_\varepsilon^2\,b^2}{\sigma_\eta^2\,a^2}\left[1+\frac{2}{n(k-1)}+\frac{2\beta}{nk(k-1)}+\frac{(4k-2)\alpha}{k(k-1)(n-1)}+\frac{2\alpha^2}{k(k-1)(n-1)}\right]$$

Enfin la variance de T lorsque l'hypothèse nulle n'est pas satisfaite est en première approximation égale à l'expression précédemment écrite de V(T) multipliée par :

$$\frac{s_\varepsilon^4\;S_b^4}{s_\eta^4\;S_a^4}\,.$$

3. - ÉTUDE MATHÉMATIQUE DU RAPPORT T

3. A. - Espérance mathématique du numérateur et dénominateur

On a :

$$E\left[s_\varepsilon^2\;S_b^2\right]=E\left[\left(\frac{\sum(X_{ij}-X_{i.})^2}{n(k-1)}\right)\left(\sum(Y_{i.}-Y_{..})^2-\frac{n-1}{n(k-1)}\sum(Y_{ij}-Y_{i.})^2\right)\right]$$

Le premier terme n'étant fonction que des ε_{ij} et le second des Y_{ij} il y a indépendance et :

$$E\left[s_\varepsilon^2\;S_b^2\right]=\sigma_\varepsilon^2\,b^2(n-1)\,\sigma_t^2\,.$$

de même :

$$E\left[s_\eta^2\;S_a^2\right]=\sigma_\eta^2\,a^2(n-1)\,\sigma_t^2\,.$$

3. B. - Variances et covariance

On a :

$$V(s_\varepsilon^2\;S_b^2)=E\left[s_\varepsilon^4\;S_b^4\right]-\left(E\left[s_\varepsilon^2\;S_b^2\right]\right)^2$$

et du fait de l'indépendance entre s_ε^2 et S_b^2 :

$$E\left[s_\varepsilon^4\;S_b^4\right]=E\left[s_\varepsilon^4\right]\cdot E\left[S_b^4\right]\,.$$

or :

$$E\left[s_\varepsilon^4\right]=\frac{\sigma_\varepsilon^4}{n(k-1)^2}E\left[\chi^4_{n(k-1)}\right]$$

$$=\sigma_\varepsilon^4\left(1+\frac{2}{n(k-1)}\right),$$

et (sous l'hypothèse où les t_i sont aléatoires) :

$$E\left[S_b^4\right] = V\left[S_b^2\right] + \left(E\left[S_b^2\right]\right)^2 =$$

$$= 2\left(b^2\,\sigma_t^2 + \frac{\sigma_\eta^2}{k}\right)^2 (n-1) + \frac{2(n-1)^2}{nk^2(k-1)}\,\sigma_\eta^4 + b^4 (n-1)^2\,\sigma_t^4\ ;$$

d'où : $$V(s_\varepsilon^2\,S_b^2) = \sigma_\varepsilon^4\left(1+\frac{2}{n(k-1)}\right)\left[2\left(b^2\,\sigma_t^2 + \frac{\sigma_\eta^2}{k}\right)^2 (n-1) + \right.$$

$$\left. + \frac{2(n-1)^2}{nk^2(k-1)}\,\sigma_\eta^4 + b^4 (n-1)^2\,\sigma_t^4\right] - \sigma_\varepsilon^4\,b^4 (n-1)^2\,\sigma_t^4\ .$$

Pour le calcul de la covariance, on détermine d'abord E $(s_\varepsilon^2\ S_b^2\ s_\eta^2\ S_a^2)$ qui se met sous la forme d'une somme de produits de termes indépendants sauf pour le premier produit pour lequel on a besoin de :

$$E\left[\sum (X_{i.} - X_{..})^2 \sum (Y_{i.} - Y_{..})^2\right]$$

qui a déjà été calculé pour obtenir E $(S_a^2\ S_b^2)$.

Après simplifications, il vient :

$$\operatorname{cov}\left[(s_\varepsilon^2\,S_b^2)(s_\eta^2\,S_b^2)\right] = E\left[s_\varepsilon^2\,S_b^2\,s_\eta^2\,S_a^2\right] - E\left[s_\varepsilon^2\,S_b^2\right] E\left[s_\eta^2\,S_a^2\right]$$

$$= 2\,a^2\,b^2\,\sigma_t^4 (n-1)\,\sigma_\varepsilon^2\,\sigma_\eta^2 - \frac{2b^2\,\sigma_t^2\,\sigma_\eta^2\,\sigma_\varepsilon^4 (n-1)^2}{kn(k-1)} -$$

$$- \frac{2\,a^2\,\sigma_t^2\,\sigma_\varepsilon^2\,\sigma_\eta^4 (n-1)^2}{kn(k-1)} + \frac{4(n-1)^2\,\sigma_\varepsilon^4\,\sigma_\eta^4}{k^2\,n^2 (k-1)^2}\ .$$

3. C. - Variance du rapport T

Nous possédons maintenant tous les éléments pour calculer la variance et l'espérance mathématique approchées du rapport au moyen de la formule déjà utilisée pour le rapport S_b^2/S_a^2

Tous calculs faits, après avoir posé :

$$\alpha = \frac{\sigma_\varepsilon^2}{a^2\,\sigma_t^2}\ ; \qquad \beta = \frac{\sigma_\eta^2}{b^2\,\sigma_t^2}\ ,$$

il vient en négligeant pour E (T) les termes d'ordre égal ou supérieur à $\frac{1}{n^2}$:

$$E(T) = \frac{b^2\,\sigma_\varepsilon^2}{a^2\,\sigma_\eta^2}\left[1 + \frac{2}{n(k-1)} + \frac{2\beta}{nk(k-1)} + \frac{(4k-2)\alpha}{k(k-1)(n-1)} + \frac{2\alpha^2}{k(k-1)(n-1)}\right]$$

$$V(T) = \frac{\sigma_\varepsilon^4\,b^4}{\sigma_\eta^4\,a^4}\left[\frac{4(n+1)}{n(n-1)(k-1)} + \frac{4(nk+1)}{n(n-1)\,k\,(k-1)}(\alpha+\beta) + \right.$$

$$\left. + \frac{2(nk+1)}{n(n-1)(k-1)\,k^2}(\alpha^2+\beta^2) + \frac{4(\alpha-\beta)^2}{n^2\,k^2 (k-1)^2}\right].$$

Il est évident que si dans le quotient T on suppose que l'écart-type σ_t devient infini, le nombre n restant fixe, le rapport $\frac{S_b^2}{S_a^2}$ converge en probabilité vers $\frac{b^2}{a^2}$ (1) Cela se manifeste d'ailleurs dans les formules donnant l'espérance mathématique et la variance de $\frac{S_b^2}{S_a^2}$ qui ont été établies dans la première partie. Sous l'hypothèse

(1) La démonstration rigoureuse de la convergence en probabilité est donnée dans l'additif. (Appendice 1).

d'équivalence des deux procédés de mesure $\frac{b^2\ \sigma_\varepsilon^2}{a^2\ \sigma_\varepsilon^2} = 1$, T est alors distribué comme le rapport de deux variances ayant toutes deux $n(k - 1)$ degrés de liberté.

Si l'on fait $\sigma_t = \infty$ dans la formule donnant V (T), on trouve alors $\alpha = \beta = 0$ et

$$V(T) = \frac{4(n+1)}{n(n-1)(k-1)} ,$$

alors que la variance du rapport de deux variances est en posant $d = n(k - 1)$:

$$V(F) = \frac{4d(d-1)}{(d-2)^2(d-4)} .$$

Ces deux expressions ne coïncident qu'asymptotiquement. C'est à cause de cela que nous préférons substituer au terme :

$$\frac{4(n-1)}{n(n-1)(k-1)} , \qquad \frac{4d(d-1)}{(d-2)^2(d-4)}$$

puisqu'en agissant ainsi nous obtenons une expression exacte de V (T) lorsque $\alpha = \beta = 0$. (Formule donnée à la fin du paragraphe 2 - 2ème partie).

Nous pensons d'autre part que les termes correctifs en α, β, α^2, β^2 et $(\alpha - \beta)^2$ permettent de représenter en première approximation l'augmentation de la variance qui résulte de la multiplication de :

$$\frac{s_\varepsilon^2}{s_\eta^2} \qquad \text{par} \qquad \frac{S_b^2/b^2}{S_a^2/a^2}$$

Enfin, pour tester l'hypothèse $\frac{b\,\sigma_\varepsilon}{a\,\sigma_\eta} = 1$, nous admettrons que T suit une loi voisine de la loi de F mais de variance plus grande, c'est ainsi que nous remplaçons la limite de confiance F par $F\sqrt{\frac{V(T)}{V(F)}}$, ce qui n'apporte qu'une correction assez minime à condition que n soit assez grand et α et β assez petits.

REMARQUE IMPORTANTE

Il importe donc de noter que la comparaison de deux méthodes de mesure n'est possible par la méthode que nous venons d'indiquer que lorsque $\frac{\sigma_\varepsilon^2}{k a^2 \sigma_t^2}$ et $\frac{\sigma_{\bar\eta}}{k\, b^2\, \sigma_t^2}$ sont petits (par exemple inférieurs à 0,1) et n assez grand ($\geqslant$ 15 par exemple) de façon que la variance du rapport $\frac{S_b^2}{S_a^2}$ soit assez petite et que l'on puisse rendre compte de la loi du rapport T par une simple correction à la loi du rapport des variances $\frac{s_\varepsilon^2}{s_\eta^2}$.

TROISIÈME PARTIE

APPLICATION A LA COMPARAISON DE DEUX ESSAIS DES COKES MÉTALLURGIQUES

INTRODUCTION

Un bon coke métallurgique doit, tout en étant léger afin d'être facilement pénétré par les gaz qui circulent dans le haut-fourneau, résister à la compression et à l'abrasion à laquelle il est soumis lors de sa descente dans cet appareil.

Pour mesurer la résistance du coke, on utilise divers procédés empiriques, dont l'essai MICUM qui est probablement le plus répandu en Europe.

Il consiste à séparer une fraction granulométrique donnée du coke à essayer et à en introduire 50 kg dans un tambour tournant où les morceaux s'usent les uns contre les autres et se brisent en retombant. Le coke est ensuite extrait de l'appareil, passé sur des tamis vibrants, et les différentes fractions granulométriques pesées. Ces poids ramenés en pourcentages du poids du coke introduit dans l'appareil constituent les différents indices MICUM :

M_{40} : % de coke ne passant pas au travers du tamis de 40 mm

M_{20} : % de coke passant au travers du tamis de 20 mm

M_{10} : % de coke passant au travers du tamis de 10 mm

On peut évidemment faire de cet essai de nombreuses variantes en augmentant par exemple le nombre des révolutions du tambour ou en changeant les grosseurs de mailles des tamis.

Comme l'essai MICUM normal est effectué sur la fraction du coke prélevé au moyen d'une fourche dont les dents sont espacées de 40mm, l'étude de variantes à cet essai devient particulièrement intéressante si l'on est amené, dans certaines circonstances, à vouloir évaluer la qualité d'un coke dont la fraction supérieure à 40 mm est faible ou inexistante.

C'est dans ce but que l'IRSID (*) a étudié un essai spécial que nous qualifierons ici d'essai IRSID.

La comparaison des deux essais du point de vue de la reproductibilité et de la sensibilité a évidemment une grosse importance puisque ces deux qualités déterminent, dans une certaine mesure, l'intérêt économique qui s'attache à chacune des deux méthodes. Utilisant des résultats expérimentaux obtenus par le Département Coke et Fonte de l'IRSID, nous avons, en conséquence, procédé à la comparaison des deux modes d'essai du coke par la méthode qui a été exposée ici sur le plan théorique.

COMPARAISON DE L'ESSAI MICUM A L'ESSAI IRSID

L'indice MICUM à retenir dans des circonstances données dépend de diverses considérations et principalement de l'utilisation du coke.

Puisqu'il ne s'agit ici que de donner un exemple d'application de la méthode théorique que nous avons exposée, nous nous limiterons à la comparaison des indices appelés M_{10} (Essai MICUM) et I_{10} (Essai IRSID).

Chaque coke était, au cours des essais dont nous allons parler, représenté par un wagon de 20 tonnes qui fut échantillonné de façon continue au cours du

(*) Institut de Recherches de la Sidérurgie.

déchargement. Avec l'échantillon ainsi constitué, on a procédé à six essais MICUM et six essais IRSID. On peut considérer le wagon comme constituant l'unité primaire du sondage et il n'y a, par ailleurs, aucune raison d'apparier les mesures IRSID aux mesures MICUM d'une manière plutôt que d'une autre du moins pour les besoins de la comparaison. Cependant, les mesures s'étant, au cours des essais, succédées alternativement et l'intérêt de pouvoir représenter les résultats par un diagramme étant évident, nous les avons associées deux à deux dans l'ordre où elles se sont succédées.

Les 39 wagons de coke échantillonnés provenaient de cinq origines désignées C, F, L, R, T (Fig.1). Les résultats des essais sont représentés par des signes indiquant la provenance, accompagnés du numéro d'ordre du wagon de cette provenance. Pour chaque signe le même numéro est donc répété six fois.

Les différents groupes de six points sont plus ou moins voisins d'une ligne donnant la tendance générale et fournissant la correspondance entre les deux essais.

Les données relatives aux essais MICUM d'une part et IRSID d'autre part sont semblables à celles de deux analyses de variance distinctes à une seule sorte de catégories. On peut calculer très simplement, selon les méthodes connues, les variances interclass et intraclass et le rapport $\frac{S_I^2}{S_M^2}$ dont nous avons parlé dans l'étude théorique :

$$\frac{S_I^2}{S_M^2} = \frac{\text{Variance interclass IRSID - Variance intraclass IRSID}}{\text{Variance interclass MICUM - Variance intraclass MICUM}}$$

Il constitue une estimation du carré $\frac{i^2}{m^2}$ de la pente de la droite représentative de la liaison existant entre I_{10} et M_{10}. Le biais et la variance de cette estimation étant connus, on peut en déduire aisément les limites de confiance à 95 % de la pente elle-même que l'on trouve être pour cette série de résultats :

$$2,88 < \frac{i}{m} < 3,21$$

mais nous avons vu que la comparaison de deux méthodes de mesure d'une même grandeur ne doit pas s'arrêter là. Il faut également tenir compte des dispersions propres aux deux méthodes.

La supériorité constatée de l'essai IRSID est en effet compensée par une moins bonne reproductibilité et l'on trouve pour le rapport T :

$$T = \frac{S_I^2 \; s_M^2}{S_M^2 \; s_I^2} = 1,96 \; .$$

On peut calculer d'une part des limites L_1 L_2 entre lesquelles T aurait une probabilité 0,95 ou 0,99 de se trouver si les deux méthodes étaient équivalentes c'est-à-dire si l'on avait :

$$\frac{\sigma_m^2 \; i^2}{\sigma_i^2 \; m^2} = 1 \; .$$

Ce test montre que la valeur de T obtenue est significative (*). (Tableau I).

TABLEAU I : Test de l'hypothèse $\frac{\sigma_m^2 \; i^2}{\sigma_i^2 \; m^2} = 1$

	Limites à 95 %	Limites à 99 %
T	0,75 1,32	0,70 1,42

(*) Le nombre des degrés de liberté étant ici égal à 195, il dépasse les possibilités des tables usuelles de Snédécor. Il est inutile de tenter d'appliquer la méthode que nous avons indiquée en vue de modifier les limites de F. On peut se contenter de former ici : $1 + 1,96 \sqrt{V(T)}$ pour $p = 0,95$ et $1 + 2,57 \sqrt{V(T)}$ pour $p = 0,99$. Les limites supérieures de F peuvent en effet être calculées au moyen de l'approximation normale, mais pas les limites inférieures. Celles-ci doivent être prises égales à $\frac{1}{1 + 1,96 \sqrt{V(T)}}$ et $\frac{1}{1 + 2,57 \sqrt{V(T)}}$

On peut d'autre part utiliser T comme une estimation de $\frac{\sigma_m^2 \, i^2}{\sigma_i^2 \, m^2}$ dont on connait la variance et le biais et extraire les racines carrées de ses limites de confiance pour obtenir celles de $\frac{\sigma_m \, i}{\sigma_i \, m}$. (Tableau II).

TABLEAU II

	Limites à 0,95		Limites à 0,99	
T	1,22	1,61	1,17	1,67

D'après les résultats de ce tableau, il apparait que la méthode d'essai IRSID du coke est, du seul point de vue de son pouvoir de caractérisation du coke, supérieure à la méthode MICUM en ce qui concerne les indices M_{10} et I_{10}.