

REVUE DE STATISTIQUE APPLIQUÉE

FRANCIS MAURIN

R. SCHOLLER

Ajustement linéaire lorsque les deux variables sont soumises à des erreurs de variances hétérogènes

Revue de statistique appliquée, tome 30, n° 3 (1982), p. 27-38

http://www.numdam.org/item?id=RSA_1982__30_3_27_0

© Société française de statistique, 1982, 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/>

AJUSTEMENT LINEAIRE LORSQUE LES DEUX VARIABLES SONT SOUMISES A DES ERREURS DE VARIANCES HETEROGENES

Francis MAURIN* et R. SCHOLLER**

*Professeur à l'Université Paris V

**Directeur de la Fondation de Recherche en hormonologie

1. INTRODUCTION

Le problème traité se pose quotidiennement en Sciences Physiques et en Biologie. Supposons, en effet, que nous devions doser une substance qui se trouve en solution dans divers échantillons que nous noterons j ($j = 1, 2, \dots, N$), qu'il existe 2 méthodes, X et Y, pour doser cette substance et que nous désirions comparer ces 2 méthodes. La première méthode est soumise à une erreur normale ϵ_X , la deuxième est soumise à une erreur normale ϵ_Y . Les écart-types de ϵ_X et de ϵ_Y varient en fonction de la valeur véritable du poids de la substance à doser; ils sont inconnus ainsi que la valeur véritable. Nous faisons l'hypothèse de l'existence d'une relation linéaire entre la mesure ξ faite par la méthode X et la mesure η faite par la méthode Y: $\eta = \beta \xi + \alpha$. Il s'agit d'estimer β et α .

La méthode du maximum de vraisemblable mise en oeuvre par LINDLEY, puis par d'autres auteurs est inutilisable car la surface de vraisemblance présente un col et il est nécessaire, pour éviter cette difficulté, de faire une hypothèse supplémentaire, par exemple de supposer que le rapport σ_Y/σ_X est indépendant du poids de la substance à doser.

La méthode des moindres carrés pondérés, mise en oeuvre par YORK, qui conduit à des équations qu'on ne peut résoudre que par des méthodes numériques, est aussi inutilisable car les poids utilisés sont inversement proportionnels aux variances qui, dans le cas que nous étudions ici, sont inconnues.

La méthode que nous proposons ici est inspirée de la méthode de WALD. Nous supposons que nous avons à notre disposition le même nombre m de répétitions des mesures faites par X et Y pour chaque échantillon analysé. Nous étudions l'ajustement linéaire lorsque m tend vers l'infini, le nombre d'échantillons restant fini. Cette hypothèse est d'ailleurs facilement utilisable dans la pratique: il est plus facile d'effectuer un très grand nombre de mesures sur quelques échantillons que d'utiliser un très grand nombre d'échantillons.

2. NOTATIONS

Nous supposons que nous disposons d'un nombre pair, $N = 2k$, d'échantillons à analyser.

Nous considérons les variables aléatoires

$$\begin{cases} X_{ji} = \xi_j + \epsilon_{ji} \\ Y_{ji} = \eta_j + \epsilon'_{ji} \end{cases} \quad (j = 1, 2, \dots, 2k \text{ et } i = 1, 2, \dots, m)$$

Les variables aléatoires X_{ji} , pour j fixé et i variant de 1 à m sont indépendantes entre elles et de même loi. Elles représentent les m répétitions des mesures par la méthode X. Les variables aléatoires Y_{ji} , pour j fixé et i variant de 1 à m sont indépendantes entre elles et de même loi. Elles représentent les m répétitions des mesures par la méthode Y.

ξ_j est l'espérance mathématique de X_{ji} .

η_j est l'espérance mathématique de Y_{ji} .

ϵ_{ji} est une variable aléatoire normale, centrée d'écart-type σ_j , $\forall i = 1, \dots, m$.

ϵ'_{ji} est une variable aléatoire normale, centrée d'écart-type σ'_j , $\forall i = 1, \dots, m$.

On suppose que les variables aléatoires ϵ_{ji} et ϵ'_{ip} sont indépendantes entre elles, $\forall i = 1, 2, \dots, m$; $\forall p = 1, 2, \dots, m$; $\forall j = 1, 2, \dots, 2k$; $\forall l = 1, 2, \dots, 2k$

Les nombres ξ_j , η_j , σ_j et σ'_j sont inconnus.

On suppose que :

$$\eta_j = \beta \xi_j + \alpha \quad \forall j = 1, 2, \dots, 2k$$

On suppose, de plus, que nous sommes assurés que :

$$\xi_j < \xi_1 \quad \forall j = 1, 2, \dots, k$$

$$\eta_j < \eta_1 \quad \forall l = k + 1, k + 2, \dots, 2k$$

Il s'agit d'estimer α et β à partir des réalisations x_{ji} et y_{li} respectives de X_{ji} et Y_{li} .

On pose :

$$B_1 = \left(\sum_{j=1}^k \sum_{i=1}^m X_{ji} - \sum_{l=k+1}^{2k} \sum_{i=1}^m X_{li} \right) / m$$

$$B_2 = \left(\sum_{j=1}^k \sum_{i=1}^m Y_{ji} - \sum_{l=k+1}^{2k} \sum_{i=1}^m Y_{li} \right) / m$$

et :

$$B = B_2 / B_1$$

On pose :

$$\bar{X} = \left(\sum_{j=1}^{2k} \sum_{i=1}^m X_{ji} \right) / 2mk$$

$$\bar{Y} = \left(\sum_{j=1}^{2k} \sum_{i=1}^m Y_{ji} \right) / 2mk$$

et :

$$A = \bar{Y} - B\bar{X}$$

On suppose que :

$$\inf_m \left| \sum_{j=1}^k \sum_{i=1}^m x_{ji} - \sum_{j=k+1}^{2k} \sum_{i=1}^m x_{ji} \right| > 0$$

3. ESTIMATION PONCTUELLE DE β

a) B_1 est un estimateur convergent, lorsque $m \rightarrow +\infty$, de $\sum_{j=1}^k \xi_j - \sum_{j=k+1}^{2k} \xi_j$.

En effet,

d'une part :

$$E(B_1) = \left[\sum_{j=1}^k \sum_{i=1}^m E(X_{ji}) - \sum_{j=k+1}^{2k} \sum_{i=1}^m E(X_{ji}) \right] / m$$

$$E(B_1) = \left(m \sum_{j=1}^k \xi_j - m \sum_{j=k+1}^{2k} \xi_j \right) / m$$

$$E(B_1) = \sum_{j=1}^k \xi_j - \sum_{j=k+1}^{2k} \xi_j$$

d'autre part :

$$V(B_1) = \left[\sum_{j=1}^k \sum_{i=1}^m V(X_{ji}) + \sum_{j=k+1}^{2k} \sum_{i=1}^m V(X_{ji}) \right] / m^2$$

$$V(B_1) = \frac{1}{m} \left(\sum_{j=1}^k \sigma_j^2 + \sum_{j=k+1}^{2k} \sigma_j^2 \right) \xrightarrow{m \rightarrow +\infty} 0$$

b) B_2 est un estimateur convergent, lorsque $m \rightarrow +\infty$, de $\sum_{j=1}^k \eta_j - \sum_{j=k+1}^{2k} \eta_j$.

La démonstration est analogue à la précédente.

c) $B = B_2/B_1$ est un estimateur convergent, lorsque $m \rightarrow +\infty$, de β :

En effet, à cause du a) et du b), B est un estimateur convergent de :

$$\begin{aligned} & \left(\sum_{j=1}^k \eta_j - \sum_{i=k+1}^{2k} \eta_i \right) / \left(\sum_{j=1}^k \xi_j - \sum_{i=k+1}^{2k} \xi_i \right) \\ &= \left(\sum_{j=1}^k (\alpha + \beta \xi_j) - \sum_{i=k+1}^{2k} (\alpha + \beta \xi_i) \right) / \left(\sum_{j=1}^k \xi_j - \sum_{i=k+1}^{2k} \xi_i \right) \\ &= \beta \left(\sum_{j=1}^k \xi_j - \sum_{i=k+1}^{2k} \xi_i \right) / \left(\sum_{j=1}^k \xi_j - \sum_{i=k+1}^{2k} \xi_i \right) = \beta \end{aligned}$$

4. ESTIMATION PONCTUELLE DE α

a) Y est un estimateur convergent, lorsque $m \rightarrow +\infty$, de $\left(\sum_{j=1}^k \eta_j + \sum_{i=k+1}^{2k} \eta_i \right) / 2k$.

En effet,

d'une part :

$$E(Y) = \left(\sum_{j=1}^k \sum_{i=1}^m E(Y_{ji}) + \sum_{i=k+1}^{2k} \sum_{i=1}^m E(Y_{li}) \right) / 2mk$$

$$E(Y) = \left(\sum_{j=1}^k \eta_j + \sum_{i=k+1}^{2k} \eta_i \right) / 2k$$

d'autre part :

$$V(Y) = \left(\sum_{j=1}^k \sum_{i=1}^m V(Y_{ji}) + \sum_{i=k+1}^{2k} \sum_{i=1}^m V(Y_{li}) \right) / 4k^2 m^2$$

$$V(Y) = \frac{1}{4k^2 m} \left(\sum_{j=1}^k \sigma_j'^2 + \sum_{i=k+1}^{2k} \sigma_i'^2 \right) \xrightarrow{m \rightarrow +\infty} 0$$

b) X est un estimateur convergent, lorsque $m \rightarrow +\infty$, de $\left(\sum_{j=1}^k \xi_j + \sum_{i=k+1}^{2k} \xi_i \right) / 2k$.

La démonstration est analogue à la précédente.

c) $A = \bar{Y} - B\bar{X}$ est un estimateur convergent, lorsque $m \rightarrow +\infty$, de α :

D'après 3; 4 a) et 4 b), A est un estimateur convergent de :

$$\begin{aligned} & \left(\sum_{j=1}^k \eta_j + \sum_{1=k+1}^{2k} \eta_1 \right) - \beta \left(\sum_{j=1}^k \xi_j + \sum_{1=k+1}^{2k} \xi_1 \right) / 2k \\ &= \left[\sum_{j=1}^k (\eta_j - \beta \xi_j) + \sum_{1=k+1}^{2k} (\eta_1 - \beta \xi_1) \right] / 2k \\ &= (k\alpha + k\alpha) / 2k = \alpha \end{aligned}$$

5. ESTIMATION DES VARIANCES

Cas particulier où $\sigma_j = \lambda \xi_j + \mu$ et $\sigma'_j = \lambda' \eta_j + \mu'$.

Soit :
$$\bar{X}_j = \frac{1}{m} \sum_{i=1}^m X_{ji} \quad \text{et} \quad \bar{Y}_j = \frac{1}{m} \sum_{i=1}^m Y_{ji},$$

il est classique, dans le cas général où la façon dont varient les écart-types est inconnue, que :

$$\frac{1}{m-1} \sum_{i=1}^m (X_{ji} - \bar{X}_j)^2 \quad \text{est un estimateur de } \sigma_j^2, \quad \forall j = 1, 2, \dots, 2k$$

$$\frac{1}{m-1} \sum_{i=1}^m (Y_{ji} - \bar{Y}_j)^2 \quad \text{est un estimateur de } \sigma_j'^2, \quad \forall j = 1, 2, \dots, 2k$$

Dans l'hypothèse particulière où $\sigma_j = \lambda \xi_j + \mu$ et $\sigma'_j = \lambda' \eta_j + \mu'$, il s'agit d'estimer les paramètres λ, μ, λ' et μ' .

a) Estimation de λ et λ' :

$$\sum_{j=1}^k \left[\frac{1}{m-1} \sum_{i=1}^m (X_{ji} - \bar{X}_j)^2 \right]^{\frac{1}{2}} - \sum_{1=k+1}^{2k} \left[\frac{1}{m-1} \sum_{i=1}^m (X_{1i} - \bar{X}_1)^2 \right]^{\frac{1}{2}}$$

est un estimateur convergent de :

$$\sum_{j=1}^k \sigma_j - \sum_{1=k+1}^{2k} \sigma_1 = \sum_{j=1}^k (\lambda \xi_j + \mu) - \sum_{1=k+1}^{2k} (\lambda \xi_1 + \mu) = \lambda \left(\sum_{j=1}^k \xi_j - \sum_{1=k+1}^{2k} \xi_1 \right)$$

Nous savons déjà que B_1 est un estimateur convergent de la quantité

$$\sum_{j=1}^k \xi_j - \sum_{i=k+1}^{2k} \xi_i$$

Il en résulte que :

$$m \cdot \frac{\sum_{j=1}^k \left[\frac{1}{m-1} \sum_{i=1}^m (X_{ji} - X_j)^2 \right]^{\frac{1}{2}} - \sum_{i=k+1}^{2k} \left[\frac{1}{m-1} \sum_{i=1}^m (X_{li} - X_l)^2 \right]^{\frac{1}{2}}}{\sum_{j=1}^k \sum_{i=1}^m X_{ji} - \sum_{i=k+1}^{2k} \sum_{i=1}^m X_{li}}$$

est un estimateur convergent de λ . On démontre, de la même façon, que :

$$m \cdot \frac{\sum_{j=1}^k \left[\frac{1}{m-1} \sum_{i=1}^m (Y_{ji} - \bar{Y}_j)^2 \right]^{\frac{1}{2}} - \sum_{i=k+1}^{2k} \left[\frac{1}{m-1} \sum_{i=1}^m (Y_{li} - \bar{Y}_l)^2 \right]^{\frac{1}{2}}}{\sum_{j=1}^k \sum_{i=1}^m Y_{ji} - \sum_{i=k+1}^{2k} \sum_{i=1}^m Y_{li}}$$

est un estimateur convergent de λ' , à condition de supposer que :

$$\inf_m \left| \sum_{j=1}^k \sum_{i=1}^m y_{ji} - \sum_{i=k+1}^{2k} \sum_{i=1}^m y_{li} \right| > 0$$

b) Estimation de μ et μ' :

$$\sum_{j=1}^{2k} \left[\frac{1}{m-1} \sum_{i=1}^m (X_{ji} - X_j)^2 \right]^{\frac{1}{2}}$$

est un estimateur convergent de :

$$\sum_{j=1}^{2k} \sigma_j$$

c'est-à-dire de :

$$\lambda \sum_{j=1}^{2k} \xi_j + 2k\mu$$

Nous venons de voir que :

$$m \cdot \frac{\sum_{j=1}^k \left[\frac{1}{m-1} \sum_{i=1}^m (X_{ji} - \bar{X}_j)^2 \right]^{\frac{1}{2}} - \sum_{i=k+1}^{2k} \left[\frac{1}{m-1} \sum_{i=1}^m (X_{li} - X_l)^2 \right]^{\frac{1}{2}}}{\sum_{j=1}^k \sum_{i=1}^m X_{ji} - \sum_{i=k+1}^{2k} \sum_{i=1}^m X_{li}}$$

est un estimateur convergent de λ .

Nous savons enfin (4 . b)) que $2k \bar{X}$ est un estimateur convergent de :

$$\sum_{j=1}^{2k} \xi_j$$

Il en résulte que :

$$\frac{1}{2k} \sum_{j=1}^{2k} \left[\frac{1}{m-1} \sum_{i=1}^m (X_{ji} - \bar{X}_j)^2 \right]^{\frac{1}{2}}$$

$$- m \bar{X} \frac{\sum_{j=1}^k \left[\frac{1}{m-1} \sum_{i=1}^m (X_{ji} - \bar{X}_j)^2 \right]^{\frac{1}{2}} - \sum_{i=k+1}^{2k} \left[\frac{1}{m-1} \sum_{i=1}^m (X_{li} - \bar{X}_l)^2 \right]^{\frac{1}{2}}}{\sum_{j=1}^k \sum_{i=1}^m X_{ji} - \sum_{i=k+1}^{2k} \sum_{i=1}^m X_{li}}$$

est un estimateur convergent de μ .

On démontre, de la même façon, que :

$$\frac{1}{2k} \sum_{j=1}^{2k} \left[\frac{1}{m-1} \sum_{i=1}^m (Y_{ji} - \bar{Y}_j)^2 \right]^{\frac{1}{2}}$$

$$- m \bar{Y} \frac{\sum_{j=1}^k \left[\frac{1}{m-1} \sum_{i=1}^m (Y_{ji} - \bar{Y}_j)^2 \right]^{\frac{1}{2}} - \sum_{i=k+1}^{2k} \left[\frac{1}{m-1} \sum_{i=1}^m (Y_{li} - \bar{Y}_l)^2 \right]^{\frac{1}{2}}}{\sum_{j=1}^k \sum_{i=1}^m Y_{ji} - \sum_{i=k+1}^{2k} \sum_{i=1}^m Y_{li}}$$

est un estimateur convergent de μ' .

6. DETERMINATION D'UN INTERVALLE DE CONFIANCE POUR β

Nous posons:
$$\bar{\epsilon}_j = \frac{1}{m} \sum_{i=1}^m \epsilon_{ji}$$

et:
$$\epsilon'_j = \frac{1}{m} \sum_{i=1}^m \epsilon'_{ji}$$

a) Examinons la variable aléatoire $B_1 (B - \beta) = B_2 - \beta B_1$.

$$B_2 - \beta B_1 = \sum_{j=1}^k \bar{Y}_j - \sum_{i=k+1}^{2k} \bar{Y}_i - \beta \left[\sum_{j=1}^k \bar{X}_j - \sum_{i=k+1}^{2k} \bar{X}_i \right]$$

$$\begin{aligned}
&= \sum_{j=1}^k (\eta_j + \bar{\epsilon}'_j) - \sum_{i=k+1}^{2k} (\eta_1 + \bar{\epsilon}'_1) - \beta \left[\sum_{j=1}^k (\xi_j + \bar{\epsilon}_j) - \sum_{i=k+1}^{2k} (\xi_1 + \bar{\epsilon}_1) \right] \\
&= \sum_{j=1}^k [\eta_j - \beta \xi_j + \bar{\epsilon}'_j - \beta \bar{\epsilon}_j] - \sum_{i=k+1}^{2k} [\eta_1 - \beta \xi_1 + \bar{\epsilon}'_1 - \beta \bar{\epsilon}_1] \\
&= \sum_{j=1}^k [\alpha + \bar{\epsilon}'_j - \beta \bar{\epsilon}_j] - \sum_{i=k+1}^{2k} [\alpha + \bar{\epsilon}'_1 - \beta \bar{\epsilon}_1]
\end{aligned}$$

et finalement,

$$B_1 (B - \beta) = \sum_{j=1}^k (\bar{\epsilon}'_j - \beta \bar{\epsilon}_j) - \sum_{i=k+1}^{2k} (\bar{\epsilon}'_1 - \beta \bar{\epsilon}_1)$$

$\bar{\epsilon}'_j$ moyenne de variables aléatoires normales centrées indépendantes est normale centrée, $j = 1, 2, \dots, k$. De la même façon, $\bar{\epsilon}'_j$, $\bar{\epsilon}'_1$ et $\bar{\epsilon}_1$ sont des variables aléatoires normales centrées.

$B_1 (B - \beta)$, combinaison linéaire de variables aléatoires normales centrées indépendantes entre elles, suit une loi normale centrée.

b) Examinons la variable aléatoire $S^2 = S_Y^2 + \beta^2 S_X^2 - 2\beta S_{XY}$, où :

$$\begin{aligned}
S_Y^2 &= \frac{1}{m} \sum_{i=1}^m \left[\sum_{j=1}^k \sum_{p=1}^k (Y_{ji} - \bar{Y}_j)(Y_{pi} - \bar{Y}_p) + \sum_{i=k+1}^{2k} \sum_{q=k+1}^{2k} (Y_{li} - \bar{Y}_l)(Y_{qi} - \bar{Y}_q) \right. \\
&\quad \left. - 2 \sum_{j=1}^k \sum_{i=k+1}^{2k} (Y_{ji} - \bar{Y}_j)(Y_{li} - \bar{Y}_l) \right]
\end{aligned}$$

$$S_Y^2 = \frac{1}{m} \sum_{i=1}^m \left[\sum_{j=1}^k (Y_{ji} - \bar{Y}_j) - \sum_{i=k+1}^{2k} (Y_{li} - \bar{Y}_l) \right]^2$$

$$\begin{aligned}
S_X^2 &= \frac{1}{m} \sum_{i=1}^m \left[\sum_{j=1}^k \sum_{p=1}^k (X_{ji} - \bar{X}_j)(X_{pi} - \bar{X}_p) + \sum_{i=k+1}^{2k} \sum_{q=k+1}^{2k} (X_{li} - \bar{X}_l)(X_{qi} - \bar{X}_q) \right. \\
&\quad \left. - 2 \sum_{j=1}^k \sum_{i=k+1}^{2k} (X_{ji} - \bar{X}_j)(X_{li} - \bar{X}_l) \right]
\end{aligned}$$

$$S_X^2 = \frac{1}{m} \sum_{i=1}^m \left[\sum_{j=1}^k (X_{ji} - \bar{X}_j) - \sum_{i=k+1}^{2k} (X_{li} - \bar{X}_l) \right]^2$$

et :

$$S_{XY} = \frac{1}{m} \sum_{i=1}^m \left[\sum_{j=1}^k \sum_{p=1}^k (X_{ji} - \bar{X}_j)(Y_{pi} - \bar{Y}_p) - \sum_{j=1}^k \sum_{i=k+1}^{2k} (X_{ji} - \bar{X}_j)(Y_{li} - \bar{Y}_l) \right]$$

$$\begin{aligned}
& - \sum_{j=1}^k \sum_{i=k+1}^{2k} (Y_{ji} - \bar{Y}_j)(X_{li} - \bar{X}_l) + \sum_{i=k+1}^{2k} \sum_{q=k+1}^{2k} (X_{li} - \bar{X}_l)(Y_{qi} - \bar{Y}_q) \\
S_{XY} &= \frac{1}{m} \sum_{i=1}^m \left[\sum_{j=1}^k (Y_{ji} - \bar{Y}_j) - \sum_{i=k+1}^{2k} (Y_{li} - \bar{Y}_l) \right] \\
& \qquad \qquad \qquad \left[\sum_{j=1}^k (X_{ji} - \bar{X}_j) - \sum_{i=k+1}^{2k} (X_{li} - \bar{X}_l) \right]
\end{aligned}$$

On peut voir, en développant l'expression ci-dessous, que :

$$\begin{aligned}
S^2 &= \frac{1}{m} \sum_{i=1}^m \left[\sum_{j=1}^k (Y_{ji} - \bar{Y}_j) - \beta \sum_{j=1}^k (X_{ji} - \bar{X}_j) - \sum_{i=k+1}^{2k} (Y_{li} - \bar{Y}_l) \right. \\
& \qquad \qquad \qquad \left. + \beta \sum_{i=k+1}^{2k} (X_{li} - \bar{X}_l) \right]^2
\end{aligned}$$

et donc que :

$$\begin{aligned}
S^2 &= \frac{1}{m} \sum_{i=1}^m \left[\sum_{j=1}^k (\epsilon'_{ji} - \beta \epsilon_{ji}) - \sum_{i=k+1}^{2k} (\epsilon'_{li} - \beta \epsilon_{li}) - \sum_{j=1}^k (\epsilon'_j - \beta \epsilon_j) \right. \\
& \qquad \qquad \qquad \left. + \sum_{i=k+1}^{2k} (\epsilon'_i - \beta \epsilon_i) \right]^2
\end{aligned}$$

Si bien qu'en posant :

$$U_i = \sum_{j=1}^k (\epsilon'_{ji} - \beta \epsilon_{ji}) - \sum_{i=k+1}^{2k} (\epsilon'_{li} - \beta \epsilon_{li})$$

on a :

$$S^2 = \frac{1}{m} \sum_{i=1}^m (U_i - \bar{U})^2$$

Les U_i étant des variables aléatoires normales centrées, de même écart-type et indépendantes entre elles, S^2 suit une loi de χ^2 à $(m - 1)$ degrés de liberté.

c) $T = \sqrt{m - 1} \frac{B_1 (B - \beta)}{S}$ suit une loi de Student à $(m - 1)$ degrés de liberté
($S = \sqrt{S^2}$)

En effet, avec la signification que nous venons de donner à U_i , nous avons $B_1 (B - \beta) = \bar{U}$.

$$S^2 = \frac{1}{m} \sum_{i=1}^m (U_i - \bar{U})^2$$

Le résultat annoncé est un résultat classique de la Statistique.

d) Détermination de l'intervalle de confiance :

Appelons $b_1, b_2, b, s^2, s_x^2, s_y^2$ et s_{xy} des réalisations respectives de $B_1, B_2, B, S^2, S_X^2, S_Y^2$ et S_{XY} . Il est alors possible de déterminer un intervalle de confiance pour β en fixant un risque p et en lisant dans une table de Student à $(m - 1)$ degrés de liberté la valeur t_p . L'intervalle sera donné par la résolution de l'inégalité en β

$$\left| \frac{b_1 (b - \beta)}{s} \sqrt{m - 1} \right| \leq t_p$$

soit encore :

$$b_1^2 (b - \beta)^2 \leq (s_y^2 + \beta^2 s_x^2 - 2\beta s_{xy}) \frac{t_p^2}{m - 1}$$

Nous allons montrer que si $b_1^2 > s_x^2 \frac{t_p^2}{m - 1}$, l'équation en β

$$b_1^2 (b - \beta)^2 = (s_y^2 + \beta^2 s_x^2 - 2\beta s_{xy}) \frac{t_p^2}{m - 1}$$

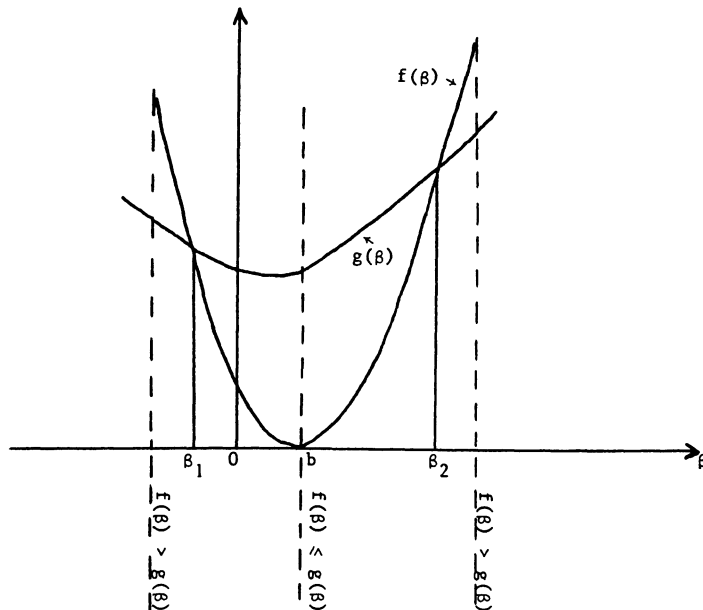
admet 2 racines réelles β_1 et β_2 et que, en supposant $\beta_1 < \beta_2$, avec une probabilité $(1 - p)$, β appartient à l'intervalle $[\beta_1, \beta_2]$.

En effet, désignons par $f(\beta)$ le membre gauche de l'équation précédente et par $g(\beta)$ le membre droit.

Pour $\beta = b$ $f = 0 \leq g$ qui est toujours ≥ 0

pour $\beta = \pm \infty$ $f > g$ puisque $b_1^2 > s_x^2 \frac{t_p^2}{m - 1}$

Les 2 paraboles représentant $f(\beta)$ et $g(\beta)$ se coupent donc bien en 2 points d'abscisses β_1 et β_2 .



7. DETERMINATION D'UN INTERVALLE DE CONFIANCE POUR α LORSQUE β EST CONNU

Comme dans le paragraphe précédent, on montre que, si $A_\beta = \bar{Y} - \beta\bar{X}$,

$$\text{et} \quad 2k(A_\beta - \alpha) = \sum_{j=1}^k (\bar{\epsilon}'_j - \beta\bar{\epsilon}_j) + \sum_{l=k+1}^{2k} (\bar{\epsilon}'_l - \beta\bar{\epsilon}_l)$$

suit donc une loi normale centrée,

que :

$$S'^2 = S_Y'^2 + \beta^2 S_X'^2 - 2\beta S'_{XY}$$

où :

$$S_Y'^2 = \frac{1}{m} \sum_{i=1}^m \left[\sum_{j=1}^k (Y_{ji} - \bar{Y}_j) + \sum_{l=k+1}^{2k} (Y_{li} - \bar{Y}_l) \right]^2$$

$$S_X'^2 = \frac{1}{m} \sum_{i=1}^m \left[\sum_{j=1}^k (X_{ji} - \bar{X}_j) + \sum_{l=k+1}^{2k} (X_{li} - \bar{X}_l) \right]^2$$

$$S'_{XY} = \frac{1}{m} \sum_{i=1}^m \left[\sum_{j=1}^k (Y_{ji} - \bar{Y}_j) + \sum_{l=k+1}^{2k} (Y_{li} - \bar{Y}_l) \right] \left[\sum_{j=1}^k (X_{ji} - \bar{X}_j) + \sum_{l=k+1}^{2k} (X_{li} - \bar{X}_l) \right]$$

suit une loi de χ^2 à $(m - 1)$ degrés de liberté

et que, en posant :

$$V_i = \sum_{j=1}^k (\epsilon'_{ji} - \beta\epsilon_{ji}) + \sum_{l=k+1}^{2k} (\epsilon'_{li} - \beta\epsilon_{li}),$$

$$\text{on a :} \quad \bar{V} = 2k(A_\beta - \alpha)$$

$$\frac{1}{m} \sum_{i=1}^m (V_i - \bar{V})^2 = S'^2$$

où les V_i sont des variables aléatoires normales centrées de même écart-type et indépendantes entre elles et donc :

$$T' = \sqrt{m-1} \frac{2k(A_\beta - \alpha)}{S'} \quad (\text{avec } S' = \sqrt{S'^2})$$

suit une loi de Student à $(m - 1)$ degrés de liberté.

Si a_β et s' sont des réalisations de A_β et S' , l'intervalle de confiance pour α , au risque p , est alors donné par :

$$\left| \frac{a_\beta - \alpha}{s'} \right| \leq \frac{t_p}{2k\sqrt{m-1}}$$

soit :

$$a_\beta - t_p \frac{s'}{2k\sqrt{m-1}} \leq \alpha \leq a_\beta + t_p \frac{s'}{2k\sqrt{m-1}}$$

BIBLIOGRAPHIE

- BARNETT. — Fitting straight lines. The linear relationship with replicated observations. *Appl. Statist.* 1970, 19, p. 135-144.
- BARTLETT. — Fitting a straight line when both variables are subject to error. *Biometrics.* 1949, 5, p. 207-212.
- J.B. COPAS. — The likelihood surface in the linear functional relationship problem. *J.R. Statist. Soc.* 1972, B 34, p. 274-278.
- D. DUGUE. — *Traité de Statistique théorique et appliquée*, 1958. Masson Ed.
- FOURGEAUD et FUCHS. — *Statistique*, 1967, Dunod Ed.
- GEARY. — Determination of linear relations between systematic parts of variables with errors of observation the variances of which are unknown. *Econometrica* 1949, 17, p. 30-59.
- HOUSNER and BRENNAN. — Estimation of linear trends. *A.M.S.*, 1948, 19, p. 380-388.
- KENDALL. — Regression, structure and functional relationships. *Biometrika*, 1951, 38, p. 11-25. *Biometrika*, 1952, 39, p. 96-108.
- LINDLEY. — Estimation of a functional relationship. *Biometrika*, 1953, 40, p. 47-49.
- NAIR and BANERJEE. — A note on fitting of straight lines if both variables are subject to error. *Sankhya.* 1942, 6, p. 331.
- NEYMAN and SCOTT. — On certain methods of estimating the linear structural relation. *A.M.S.* 1951, 22, p. 352-361. *A.M.S.* 1952, 23, p. 115.
- RIERSOL. — Identifiability of a linear relation between two variables which are subject to error. *Econometrica.* 1950, 18, p. 375-389.
- WALD. — The fitting of straight lines if both variables are subject to error. *A.M.S.* 1940, 11, p. 284-300.
- WOLFOWITZ. — The minimum distance method. *A.M.S.* 1957, 28, p. 89-110.
- YORK. — Least — squares fitting of a straight line. *Canad. J. of Physics*, 1966, 44, p. 1079-1086.