REVUE DE STATISTIQUE APPLIQUÉE

FRANCIS MAURIN

R. SCHOLLER

La régression linéaire dans l'hypothèse où l'écart-type de la variable aléatoire $Y(x)$ mesurée dépend linéairement de la variable x

Revue de statistique appliquée, tome 29, n° 3 (1981), p. 5-17

<http://www.numdam.org/item?id=RSA_1981__29_3_5_0>

LA REGRESSION LINEAIRE DANS L'HYPOTHESE OU L'ECART-TYPE DE LA VARIABLE ALEATOIRE Y(x) MESUREE DEPEND LINEAIREMENT DE LA VARIABLE x

Francis MAURIN

Professeur à l'Université Paris V

R. SCHOLLER

Directeur de la Fondation de Recherche en Hormonologie

I. INTRODUCTION

Pour chaque valeur x_i d'une variable certaine x, on dispose d'une valeur y_i (ou de plusieurs valeurs y_{ij}) d'une variable aléatoire Y. On suppose que Y et x sont liées par une relation linéaire à une variable aléatoire près ϵ additive. On est alors en présence du modèle de la régression linéaire classique

$$Y = a + bx + \epsilon$$

où a, b, x sont des quantités certaines, ϵ est une variable aléatoire normale centrée d'écart-type constant.

Le modèle que nous allons examiner diffère du précédent par la dernière hypothèse qui devient : ϵ est une variable aléatoire normale centrée d'écart-type $\sigma = \sigma_0 + \lambda x$ ou σ_0 et λ sont des quantités certaines positives.

Le problème précédent est traité par la méthode du maximum de vraisemblance qui permet d'estimer a, b, et éventuellement σ_0 et λ.

Nous devons remarquer qu'il est différent du problème dit "des variances d'erreur non constantes" (traité, par exemple par John A. JACQUEZ, F. J. MATHER et Charles R. CRAWFORD).

En effet, les paramètres qui vont intervenir ici sont a, b, σ_0 et λ alors que dans le problème où on suppose les variances d'erreur non constantes mais sans liaison mathématique connue à l'avance, les paramètres qui interviennent sont $a, b, \sigma_1, \sigma_2, \ldots, \sigma_n$.

Mots-Clès : **Régression, Maximum de Vraisemblance, Pente, ordonnée à l'origine.**

II. REMARQUE PRELIMINAIRE

A partir de la relation $Y = a + bx + \epsilon$ (II.1)
où ϵ est une variable aléatoire normale centrée d'écart-type $\sigma = \sigma_0 + \lambda\, x$, on peut se ramener, en supposant $x + \frac{\sigma_0}{\lambda} > 0$ (hypothèse non restrictive car s'il existait des $x_i < 0$, il suffirait de prendre comme nouvelle variable $x' = x + A$ avec $A > \sup_{x_i<0} |x_i|$), à la relation :

$$\frac{Y}{x + \frac{\sigma_0}{\lambda}} = b + \frac{a - b\,\frac{\sigma_0}{\lambda}}{x + \frac{\sigma_0}{\lambda}} + \epsilon' \tag{II.2}$$

où ϵ' est une variable aléatoire normale centrée d'écart-type constant λ.

Le changement de variables $u = \frac{1}{x + \frac{\sigma_0}{\lambda}}$, $Z = \frac{Y}{x + \frac{\sigma_0}{\lambda}}$ nous donne la relation :

$$Z = b + \left(a - b\,\frac{\sigma_0}{\lambda}\right) u + \epsilon' \tag{II.3}$$

ce qui signifie qu'il existe une relation linéaire entre Z et u à une variable aléatoire additive près ϵ' normale centrée d'écart-type constant λ.

La vraisemblance relative aux observations (x_i, y_i) s'écrit

$$\mathcal{L} = \sum_{i=1}^{n} \left[-\operatorname{Log}\sqrt{2\pi} - \operatorname{Log}\lambda - \operatorname{Log}\left(x_i + \frac{\sigma_0}{\lambda}\right) - \frac{\left(\frac{y_i}{x_i + \frac{\sigma_0}{\lambda}} - b - \left(a - b\,\frac{\sigma_0}{\lambda}\right)\frac{1}{x_i + \frac{\sigma_0}{\lambda}}\right)^2}{2\,\lambda^2} \right]$$

Nous examinons d'abord le cas où $\frac{\sigma_0}{\lambda}$ est connu, puis celui ou $\frac{\sigma_0}{\lambda}$ est inconnu.

III. LA SOLUTION DU MAXIMUM DE VRAISEMBLANCE DANS LE CAS OU LA QUANTITE $\frac{\sigma_0}{\lambda}$ EST CONNUE : ESTIMATIONS DE a, b ET λ

La vraisemblance est alors de la forme :

$$\mathcal{L} = C^{te} + \sum_{i=1}^{n} \left[- \operatorname{Log} \lambda + \operatorname{Log} u_i - \frac{\left(z_i - b - \left(a - b \frac{\sigma_0}{\lambda}\right) u_i\right)^2}{2\lambda^2} \right]$$

Les estimations $\hat{a}$, $\hat{b}$ et $\hat{\lambda}$ de a, b et λ s'obtiennent en annulant les dérivées partielles de $\mathcal{L}$ par rapport à a, b et λ $\left(\frac{\sigma_0}{\lambda}\right.$ étant considéré comme connu, on ne doit pas considérer, dans le calcul de $\frac{\partial \mathcal{L}}{\partial \lambda}$, ce terme comme dépendant de $\left.\lambda\right)$.

On aboutit aux 3 équations suivantes :

$$\sum_{i=1}^{n} \left[z_i - \hat{b} - \left(\hat{a} - \hat{b} \frac{\sigma_0}{\lambda}\right) u_i \right] u_i = 0$$

$$\sum_{i=1}^{n} \left[z_i - \hat{b} - \left(\hat{a} - \hat{b} \frac{\sigma_0}{\lambda}\right) u_i \right] \left(1 - \frac{\sigma_0}{\lambda} u_i\right) = 0$$

$$-\frac{n}{\hat{\lambda}} + \frac{\sum_{i=1}^{n} \left[z_i - \hat{b} - \left(\hat{a} - \hat{b} \frac{\sigma_0}{\lambda}\right) u_i \right]^2}{\hat{\lambda}^3} = 0$$

Les 3 équations précédentes peuvent s'écrire :

$$\sum_{i=1}^{n} \left[z_i - \hat{b} - \left(\hat{a} - \hat{b} \frac{\sigma_0}{\lambda}\right) u_i \right] = 0 \qquad \text{(III.1)}$$

$$\sum_{i=1}^{n} \left[z_i - \hat{b} - \left(\hat{a} - \hat{b} \frac{\sigma_0}{\lambda}\right) u_i \right] u_i = 0 \qquad \text{(III.2)}$$

$$\hat{\lambda}^2 = \frac{\sum_{i=1}^{n} \left[z_i - \hat{b} - \left(\hat{a} - \hat{b} \frac{\sigma_0}{\lambda}\right) u_i \right]^2}{n} \qquad \text{(III.3)}$$

La résolution du système constitué par (III.1) et (III.2) donne :

$$\hat{a} = \frac{n \sum_{i=1}^{n} z_i u_i - \sum_{i=1}^{n} z_i \sum_{i=1}^{n} u_i + \frac{\sigma_0}{\lambda}\left(\sum_{i=1}^{n} u_i^2 \sum_{i=1}^{n} z_i - \sum_{i=1}^{n} u_i \sum_{i=1}^{n} z_i u_i\right)}{n \sum_{i=1}^{n} u_i^2 - \left(\sum_{i=1}^{n} u_i\right)^2} \qquad \text{(III.4)}$$

$$\hat{b} = \frac{\sum_{i=1}^{n} u_i^2 \sum_{i=1}^{n} z_i - \sum_{i=1}^{n} u_i \sum_{i=1}^{n} z_i u_i}{n \sum_{i=1}^{n} u_i^2 - \left(\sum_{i=1}^{n} u_i\right)^2} \qquad \text{(III.5)}$$

La substitution de $\hat{a}$ et $\hat{b}$ par leurs expressions dans l'équation (III.3) donne :

$$\hat{\lambda}^2 = \frac{\left[n \sum_{i=1}^{n} u_i^2 - \left(\sum_{i=1}^{n} u_i\right)^2\right]\left[n \sum_{i=1}^{n} z_i^2 - \left(\sum_{i=1}^{n} z_i\right)^2\right] - \left[n \sum_{i=1}^{n} z_i u_i - \sum_{i=1}^{n} z_i \sum_{i=1}^{n} u_i\right]^2}{n^2 \left[n \sum_{i=1}^{n} u_i^2 - \left(\sum_{i=1}^{n} u_i\right)^2\right]} \qquad \text{(III.6)}$$

Soient $\hat{A}$ et $\hat{B}$ les estimateurs dont $\hat{a}$ et $\hat{b}$ sont les estimations. On obtient pour les variances $V(\hat{A})$ et $V(\hat{B})$ les résultats suivants :

$$V(\hat{A}) = \frac{n - 2\frac{\sigma_0}{\lambda} \sum_{i=1}^{n} u_i + \left(\frac{\sigma_0}{\lambda}\right)^2 \sum_{i=1}^{n} u_i^2}{n \sum_{i=1}^{n} u_i^2 - \left(\sum_{i=1}^{n} u_i\right)^2} \lambda^2$$

$$V(\hat{B}) = \frac{\sum_{i=1}^{n} u_i^2}{n \sum_{i=1}^{n} u_i^2 - \left(\sum_{i=1}^{n} u_i\right)^2} \lambda^2$$

Soient S_a et S_b des estimations respectives de l'écart-type de a et de b :

$$S_a = \left(\frac{n}{n-2}\right)^{1/2} \left[\frac{n - 2\frac{\sigma_0}{\lambda} \sum_{i=1}^{n} u_i + \left(\frac{\sigma_0}{\lambda}\right)^2 \sum_{i=1}^{n} u_i^2}{n \sum_{i=1}^{n} u_i^2 - \left(\sum_{i=1}^{n} u_i\right)^2}\right]^{1/2} \hat{\lambda}$$

$$S_b = \left(\frac{n}{n-2}\right)^{1/2} \left[\frac{\sum_{i=1}^{2} u_i^2}{n \sum_{i=1}^{n} u_i^2 - \left(\sum_{i=1}^{n} u_i\right)^2} \right]^{1/2} \hat{\lambda}$$

On en déduit des intervalles de confiance pour a et b :

pour a : $\hat{a} \pm t_\alpha \quad S_a$
pour b : $b \pm t_\alpha \quad S_b$

où t_α est pris dans une table de Student à $(n-2)$ degrés de liberté au risque α.

On en déduit également une région de confiance pour le couple (a, b), constituée par l'intérieur de l'ellipse dont l'équation est :

$$\sum_{i=1}^{n} u_i^2 (a-\hat{a})^2 + 2\left[\sum_{i=1}^{n} u_i - \frac{\sigma_0}{\lambda}\sum_{i=1}^{n} u_i^2\right](a-\hat{a})(b-\hat{b})$$
$$+ \left[\left(\frac{\sigma_0}{\lambda}\right)^2 \sum_{i=1}^{n} u_i^2 - 2\frac{\sigma_0}{\lambda}\sum_{i=1}^{n} u_i + n\right](b-\hat{b})^2$$
$$- 2F_\alpha \frac{n}{n-2}\hat{\lambda}^2 = 0$$

où F_α est pris dans une table de Snedecor pour $\nu_1 = 2, \nu_2 = n-2$ au risque α.

Cas particulier où $\sigma_0 = 0$:

On doit alors noter que les formules donnant $\hat{a}$, $\hat{b}$, $\hat{\lambda}$, S_a, S_b et l'ellipse de confiance deviennent beaucoup plus simples. On obtient, en reprenant les notations initiales :

$$\hat{a} = \frac{n\sum_{i=1}^{n}\frac{y_i}{x_i^2} - \sum_{i=1}^{n}\frac{y_i}{x_i}\sum_{i=1}^{n}\frac{1}{x_i}}{n\sum_{i=1}^{n}\frac{1}{x_i^2} - \left(\sum_{i=1}^{n}\frac{1}{x_i}\right)^2}$$

$$\hat{b} = \frac{\sum_{i=1}^{n}\frac{1}{x_i^2}\sum^{n}\frac{y_i}{x_i} - \sum_{i=1}^{n}\frac{1}{x_i}\sum_{i=1}^{n}\frac{y_i}{x_i^2}}{n\sum_{i=1}^{n}\frac{1}{x_i^2} - \left(\sum_{i=1}^{n}\frac{1}{x_i}\right)^2}$$

$$\lambda^2 = \frac{\left[n\sum_{i=1}^{n}\frac{1}{x_i^2} - \left(\sum_{i=1}^{n}\frac{1}{x_i}\right)^2\right]\left[n\sum_{i=1}^{n}\frac{y_i^2}{x_i^2} - \left(\sum_{i=1}^{n}\frac{y_i}{x_i}\right)^2\right] - \left[n\sum_{i=1}^{n}\frac{y_i}{x_i^2} - \sum_{i=1}^{n}\frac{1}{x_i}\sum_{i=1}^{n}\frac{y_i}{x_i}\right]^2}{n^2\left[n\sum_{i=1}^{n}\frac{1}{x_i^2} - \left(\sum_{i=1}^{n}\frac{1}{x_i}\right)^2\right]}$$

$$S_a = \left(\frac{n}{n-2}\right)^{1/2} \left[\frac{n}{n \sum_{i=1}^{n} \frac{1}{x_i^2} - \left(\sum_{i=1}^{n} \frac{1}{x_i}\right)^2} \right]^{1/2} \hat{\lambda}$$

$$S_b = \left(\frac{n}{n-2}\right)^{1/2} \left[\frac{\sum_{i=1}^{n} \frac{1}{x_i^2}}{n \sum_{i=1}^{n} \frac{1}{x_i^2} - \left(\sum_{i=1}^{n} \frac{1}{x_i}\right)^2} \right] \hat{\lambda}$$

$$\sum_{i=1}^{n} \frac{1}{x_i^2} (a - \hat{a})^2 + 2 \sum_{i=1}^{n} \frac{1}{x_i} (a - \hat{a})(b - \hat{b}) + n(b - \hat{b})^2 - 2 F_\alpha \frac{n}{n-2} \hat{\lambda}^2 = 0$$

IV. LA SOLUTION DU MAXIMUM DE VRAISEMBLANCE LORSQUE $\frac{\sigma_0}{\lambda}$ EST INCONNU : ESTIMATIONS DE a, b, σ_0 ET λ :

Nous reprenons alors l'expression de la vraisemblance

$$\mathcal{L} = \sum_{i=1}^{n} \left[- \operatorname{Log} \sqrt{2\pi} - \operatorname{Log} \lambda - \operatorname{Log} \left(x_i + \frac{\sigma_0}{\lambda}\right) - \frac{\left(\frac{y_i}{x_i + \frac{\sigma_0}{\lambda}} - b - \left(a - b \frac{\sigma_0}{\lambda}\right) \frac{1}{x_i + \frac{\sigma_0}{\lambda}}\right)^2}{2 \lambda^2} \right]$$

L'annulation des dérivées partielles de $\mathcal{L}$, d'une part par rapport à a, d'autre part par rapport à b donne des équations analogues à (III.1) et (III.2) mais dans lesquelles σ_0 et λ sont remplacés par leurs estimations respectives $\hat{\sigma}_0$ et $\hat{\lambda}$:

$$\sum_{i=1}^{n} \left[\frac{y_i}{x_i + \frac{\hat{\sigma}_0}{\hat{\lambda}}} - \hat{b} - \left(\hat{a} - \hat{b} \frac{\sigma_0}{\hat{\lambda}}\right) \frac{1}{x_i + \frac{\hat{\sigma}_0}{\hat{\lambda}}} \right] = 0 \qquad \text{(IV.1)}$$

$$\sum_{i=1}^{n} \left[\frac{y_i}{x_i + \frac{\hat{\sigma}_0}{\hat{\lambda}}} - \hat{b} - \left(\hat{a} - \hat{b} \frac{\sigma_0}{\hat{\lambda}}\right) \frac{1}{x_i + \frac{\hat{\sigma}_0}{\hat{\lambda}}} \right] \frac{1}{x_i + \frac{\hat{\sigma}_0}{\hat{\lambda}}} = 0 \qquad \text{(IV.2)}$$

L'annulation des dérivées partielles de $\mathcal{L}$, d'une part par rapport à σ_0, d'autre part par rapport à λ, donne les 2 équations suivantes :

$$\frac{1}{\hat{\lambda}^2} \sum_{i=1}^{n} \frac{(y_i - \hat{a} - \hat{b} x_i)^2}{\left(x_i + \frac{\sigma_0}{\hat{\lambda}}\right)^3} - \sum_{i=1}^{n} \frac{1}{x_i + \frac{\hat{\sigma}_0}{\hat{\lambda}}} = 0 \qquad \text{(IV.3)}$$

$$-\frac{n}{\hat\lambda}+\frac{1}{\hat\lambda^3}\sum_{i=1}^{n}\frac{(y_i-\hat a-\hat b\,x_i)^2}{\left(x_i+\frac{\hat\sigma_0}{\hat\lambda}\right)^2}-\frac{\hat\sigma_0}{\hat\lambda^4}\sum_{i=1}^{n}\frac{(y_i-\hat a-\hat b\,x_i)^2}{\left(x_i+\frac{\hat\sigma_0}{\hat\lambda}\right)^3}+\frac{\hat\sigma_0}{\hat\lambda^2}\sum_{i=1}^{n}\frac{1}{x_i+\frac{\hat\sigma_0}{\hat\lambda}}=0 \quad \text{(IV.4)}$$

En tenant compte de (IV. 3), l'équation (IV.4) donne :

$$n\,\hat\lambda^2=\sum_{i=1}^{n}\frac{(y_i-\hat a-\hat b\,x_i)^2}{\left(x_i+\frac{\hat\sigma_0}{\hat\lambda}\right)^2} \quad \text{(IV.4)}'$$

La combinaison de (IV. 3) et (IV. 4)' nous donne enfin :

$$n\sum_{i=1}^{n}\frac{(y_i-\hat a-\hat b\,x_i)^2}{\left(x_i+\frac{\hat\sigma_0}{\hat\lambda}\right)^2}=\sum_{i=1}^{n}\frac{(y_i-\hat a-\hat b\,x_i)^2}{\left(x_i+\frac{\hat\sigma_0}{\hat\lambda}\right)^2}\sum_{i=1}^{n}\frac{1}{x_i+\frac{\hat\sigma_0}{\hat\lambda}} \quad \text{(IV.5)}$$

La résolution du système constitué par (IV. 1) et (IV. 2) donne $\hat a$ et $\hat b$ en fonction de $\frac{\hat\sigma_0}{\hat\lambda}$ par des formules analogues à (III.4) et (III.5) mais dans lesquelles $\frac{\sigma_0}{\lambda}$ est remplacé par $\frac{\hat\sigma_0}{\hat\lambda}$. $\hat\lambda$ peut être calculé en fonction de $\frac{\hat\sigma_0}{\hat\lambda}$ indifféremment par (IV. 3) ou (IV. 4)' dans lesquelles on aura remplacé a et b par leurs expressions en fonction de $\frac{\hat\sigma_0}{\hat\lambda}$.

Il reste alors à résoudre (IV. 5) après y avoir remplacé $\hat a$ et $\hat b$ par leurs expressions en fonction de $\frac{\hat\sigma_0}{\hat\lambda}$. La résolution de cette dernière équation donnera $\frac{\hat\sigma_0}{\hat\lambda}$ et le problème du calcul de $\hat a$, $\hat b$, $\hat\sigma_0$ et $\hat\lambda$ sera résolu.

En revenant aux notations u_i et z_i, l'équation (IV. 5) s'écrit :

$$\sum_{i=1}^{n}\left[z_i\left[n\sum_{i=1}^{n}u_i^2-\left(\sum_{i=1}^{n}u_i\right)^2\right]-u_i\left[n\sum_{i=1}^{n}z_i\,u_i-\sum_{i=1}^{n}z_i\sum_{i=1}^{n}u_i\right]\right.$$

$$\left.-\left[\sum_{i=1}^{n}u_i^2\sum_{i=1}^{n}z_i-\sum_{i=1}^{n}u_i\sum_{i=1}^{n}z_i\,u_i\right]\right]^2\left(u_i-\frac{\sum_{i=1}^{n}u_i}{n}\right)=0 \quad \text{(IV.6)}$$

Il ne paraît pas possible de résoudre littéralement cette équation par rapport à $\frac{\hat\sigma_0}{\hat\lambda}$ qui figure en dénominateur à la fois dans u_i et dans z_i. On doit utiliser dans chaque cas une méthode numérique. On peut procéder par tatonnements en se donnant une première valeur $\left(\frac{\hat\sigma_0}{\hat\lambda}\right)_1$ de $\frac{\hat\sigma_0}{\hat\lambda}$ et en calculant le 1[er] membre de l'équation.

Si ce dernier est positif, on cherche une $2^{\text{ème}}$ valeur $\left(\frac{\hat{\sigma}_0}{\hat{\lambda}}\right)_2$ de $\frac{\hat{\sigma}_0}{\hat{\lambda}}$ qui rende le 1^{er} membre négatif.

Puis on essaie avec $\left(\frac{\hat{\sigma}_0}{\hat{\lambda}}\right)_3 = \frac{1}{2}\left[\left(\frac{\hat{\sigma}_0}{\hat{\lambda}}\right)_1 + \left(\frac{\hat{\sigma}_0}{\hat{\lambda}}\right)_2\right]$.

Si $\left(\frac{\hat{\sigma}_0}{\hat{\lambda}}\right)_3$ rend le 1^{er} membre positif, on recommence avec

$$\left(\frac{\hat{\sigma}_0}{\hat{\lambda}}\right)_4 = \frac{1}{2}\left[\left(\frac{\hat{\sigma}_0}{\hat{\lambda}}\right)_2 + \left(\frac{\hat{\sigma}_0}{\hat{\lambda}}\right)_3\right]$$

Si, au contraire, $\left(\frac{\hat{\sigma}_0}{\hat{\lambda}}\right)_3$ rend le 1^{er} membre négatif, on recommence avec $\left(\frac{\hat{\sigma}_0}{\hat{\lambda}}\right)_4 = \frac{1}{2}\left[\left(\frac{\hat{\sigma}_0}{\hat{\lambda}}\right)_1 + \left(\frac{\hat{\sigma}_0}{\hat{\lambda}}\right)_3\right]$ et ainsi de suite jusqu'à ce que l'on juge que l'écart entre $\left(\frac{\hat{\sigma}_0}{\hat{\lambda}}\right)_k$ et $\left(\frac{\hat{\sigma}_0}{\hat{\lambda}}\right)_{k+1}$ est assez faible.

V. UNE SOLUTION PRATIQUE DE PREMIERE APPROXIMATION LORSQU'IL Y A REPETITION DES EXPERIENCES A CHAQUE NIVEAU DE LA VARIABLE x

On suppose alors que, pour chaque valeur x_i $(i = 1, 2, \ldots, n)$, on a effectué l_i expériences qui ont donné les résultats y_{ij} $(j = 1, 2, \ldots, l_i)$ avec $N = \sum_{i=1}^{n} l_i$.

Les formules (III.4), (III.5) et (III.6) du paragraphe III s'écrivent alors, lorsque $\frac{\sigma_0}{\lambda}$ est connu, avec ces nouvelles notations :

$$\hat{a} = \frac{N \sum_{i=1}^{n} l_i u_i \sum_{j=1}^{li} z_{ij} - \sum_{i=1}^{n} l_i u_i \sum_{i=1}^{n} \sum_{j=1}^{li} z_{ij} + \frac{\sigma_0}{\lambda}\left(\sum_{i=1}^{n} l_i u_i^2 \sum_{i=1}^{n} \sum_{j=1}^{li} z_{ij} - \sum_{i=1}^{n} l_i u_i \sum_{i=1}^{n} \sum_{j=1}^{li} l_i u_i z_{ij}\right)}{N \sum_{i=1}^{n} l_i u_i^2 - \left(\sum_{i=1}^{n} l_i u_i\right)^2}$$

$$\hat{b} = \frac{\sum_{i=1}^{n} l_i u_i^2 \sum_{i=1}^{n} \sum_{j=1}^{l_i} z_{ij} - \sum_{i=1}^{n} l_i u_i \sum_{i=1}^{n} \sum_{j=1}^{l_i} l_i u_i z_{ij}}{N \sum_{i=1}^{n} l_i u_i^2 - \left(\sum_{i=1}^{n} l_i u_i\right)^2}$$

$$\hat{\lambda}^2 = \frac{\left[N \sum_{i=1}^{n} l_i u_i^2 - \left(\sum_{i=1}^{n} l_i u_i\right)^2\right]\left[N \sum_{i=1}^{n} \sum_{j=1}^{l_i} z_{ij}^2 - \left(\sum_{i=1}^{n} \sum_{j=1}^{l_i} z_{ij}\right)^2\right] - \left[N \sum_{i=1}^{n} l_i u_i \sum_{j=1}^{l_i} z_{ij} - \sum_{i=1}^{n} l_i u_i \sum_{i=1}^{n} \sum_{j=1}^{l_i} z_{ij}\right]^2}{N^2\left[\sum_{i=1}^{n} l_i u_i^2 - \left(\sum_{i=1}^{n} l_i u_i\right)^2\right]}$$

Lorsque $\frac{\sigma_0}{\lambda}$ n'est pas connu, on peut être tenté de le remplacer dans les

3 formules précédentes par une estimation $\frac{\hat{\sigma}_0}{\hat{\lambda}}$. Il se trouve, puisque nous disposons de l_i observations lorsque $x = x_i$, que nous pouvons estimer σ_i (valeur de σ lorsque $x = x_i$) par :

$$s_i = \left[\frac{l_i \sum_{j=1}^{l_i} y_{ij}^2 - \left(\sum_{j=1}^{l_i} y_{ij} \right)^2}{l_i (l_i - 1)} \right]^{1/2}$$

Nous nous trouvons alors devant le problème de la régression linéaire classique : il s'agit d'ajuster par une droite le nuage de points (x_i, s_i). On utilise l'ajustement des moindres carrés qui nous donne :

$$\frac{\hat{\sigma}_0}{\hat{\lambda}} = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} s_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i s_i}{n \sum_{i=1}^{n} x_i s_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} s_i}$$

Si cette dernière quantité était négative, on prendrait alors $\hat{\sigma}_0 = 0$ et on tomberait dans le cas particulier du paragraphe III.

Exemple

Le dosage de l'estriol dans l'urine de femme enceinte – Chromatographie en phase gazeuse (R. SCHOLLER).

La variable x représente des quantités d'estriol, pesées avec la plus grande précision, ajoutées à l'urine. La variable Y dont les valeurs sont notées y_{ij} représente le résultat du dosage de l'estriol correspondant. La méthode employée pour effectuer ce dosage est celle de la Chromatographie en phase gazeuse. Cet exemple a déjà été traité par la régression linéaire classique (σ constant) par R. SCHOLLER. Nous supposons ici que $\sigma = \sigma_0 + \lambda x$.

1) Estimation de $\frac{\sigma_0}{\lambda}$

x_i	$\sum_{j=1}^{5} y_{ij}$	$\left(\sum_{j=1}^{5} y_{ij}\right)^2$	$5 \sum_{j=1}^{5} y_{ij}$	$5 \sum_{j=1}^{5} y_{ij} - \left(\sum_{j=1}^{5} y_{ij}\right)^2$	s_i
2,5	12,330	152,0289	152,0650	0,0361	0,0420
5	24,760	613,0576	613,6680	0,6104	0,1747
10	47,400	2246,7600	2247,1750	0,4150	0,1440
15	71,925	5173,2056	5174,5215	1,3159	0,2565
20	92,100	8482,4100	8487,4500	5,0400	0,5019
25	113,55	12893,6020	12901,970	8,368	0,6468
50	232,25	53940,0620	53950,9375	10,875	0,7374

Nous avons $s_i = \left[\frac{l_i \sum_{j=1}^{l_i} y_{ij}^2 - \left(\sum_{j=1}^{l_i} y_{ij} \right)^2}{l_i (l_i - 1)} \right]^{1/2}$ qui nous donne les valeurs de s_i inscrites dans le tableau ci-dessus. Puis $\frac{\hat{\sigma}_0}{\hat{\lambda}}$ qui nous est donné par

$$\frac{\hat{\sigma}_0}{\hat{\lambda}} = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} s_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i s_i}{n \sum_{i=1}^{n} x_i s_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} s_i}$$

et qui nous conduit aux calculs suivants :

$$\frac{\hat{\sigma}_0}{\hat{\lambda}} = \frac{3881{,}25 \times 2{,}5033 - 127{,}5 \times 69{,}3440}{7 \times 69{,}3440 - 127{,}5 \times 2{,}5033}$$

On trouve :

$$\frac{\hat{\sigma}_0}{\hat{\lambda}} = 5{,}2610$$

s_i	x_i^2	$x_i\, s_i$
0,0420	6,25	0,1050
0,1747	25	0,8735
0,1440	100	1,4400
0,2565	225	3,8475
0,5019	400	10,0380
0,6468	625	16,1700
0,7374	2500	36,8700

2) Calculs de $\hat{a}$, $\hat{b}$ et $\hat{\lambda}$

x_i	y_{ij}	$x_i + \frac{\hat{\sigma}_0}{\hat{\lambda}}$	u_i	z_{ij}	u_i^2	$z_{ij}\, u_i$	z_{ij}^2
2,5	2,51 2,435 2,475 2,5 2,41	7,761	0,12885	0,32341 0,31374 0,31890 0,32212 0,31052	0,01660	0,041671 0,040425 0,041090 0,041505 0,040010	0,10459 0,09843 0,10169 0,10376 0,09642
5	4,95 4,9 4,85 5,25 4,81	10,261	0,09745	0,48237 0,47750 0,47263 0,51161 0,46873	0,00949	0,047006 0,046532 0,046057 0,049856 0,045677	0,23268 0,22800 0,22338 0,26174 0,21970

x_i	y_{ij}	$x_i + \frac{\hat{\sigma}_0}{\hat{\lambda}}$	u_i	z_{ij}	u_i^2	$z_{ij}\, u_i$	z_{ij}^2
	9,55			0,62571		0,040996	0,39151
	9,40			0,61588		0,040352	0,37931
10	9,40	15,261	0,06552	0,61588	0,00429	0,040352	0,37931
	9,7			0,63554		0,041640	0,40391
	9,35			0,61261		0,040128	0,37529
	14,475			0,71434		0,035252	0,51028
	14,625			0,72174		0,035617	0,52091
15	14,475	20,261	0,04935	0,71434	0,00243	0,035252	0,51028
	14,4			0,71064		0,035070	0,50500
	13,95			0,68843		0,033974	0,47394
	17,7			0,70056		0,027728	0,49078
	18,7			0,74014		0,029294	0,54781
20	18,7	25,261	0,03958	0,74014	0,00156	0,029294	0,54781
	18,1			0,71639		0,028354	0,51321
	18,9			0,74806		0,029608	0,55959
	22			0,72688		0,024016	0,52835
	22,875			0,75579		0,024971	0,57122
25	22,25	30,261	0,03304	0,73514	0,00109	0,024289	0,54043
	22,75			0,75166		0,024834	0,56499
	23,675			0,78222		0,025844	0,61187
	46,25			0,83666		0,015135	0,70000
	45,5			0,82309		0,014889	0,67748
50	46,75	55,261	0,01809	0,84570	0,00032	0,015298	0,73834
	46,25			0,83666		0,015135	0,70000
Σ			2,1584	22,355	0,1789	1,1627	15,1272

On trouve par l'utilisation du tableau précédent :

$$\hat{b} = 0{,}918$$
$$\hat{a} = 0{,}250$$
$$\hat{\lambda} \quad 0{,}02117$$
$$\hat{\sigma}_0 = 0{,}11137$$

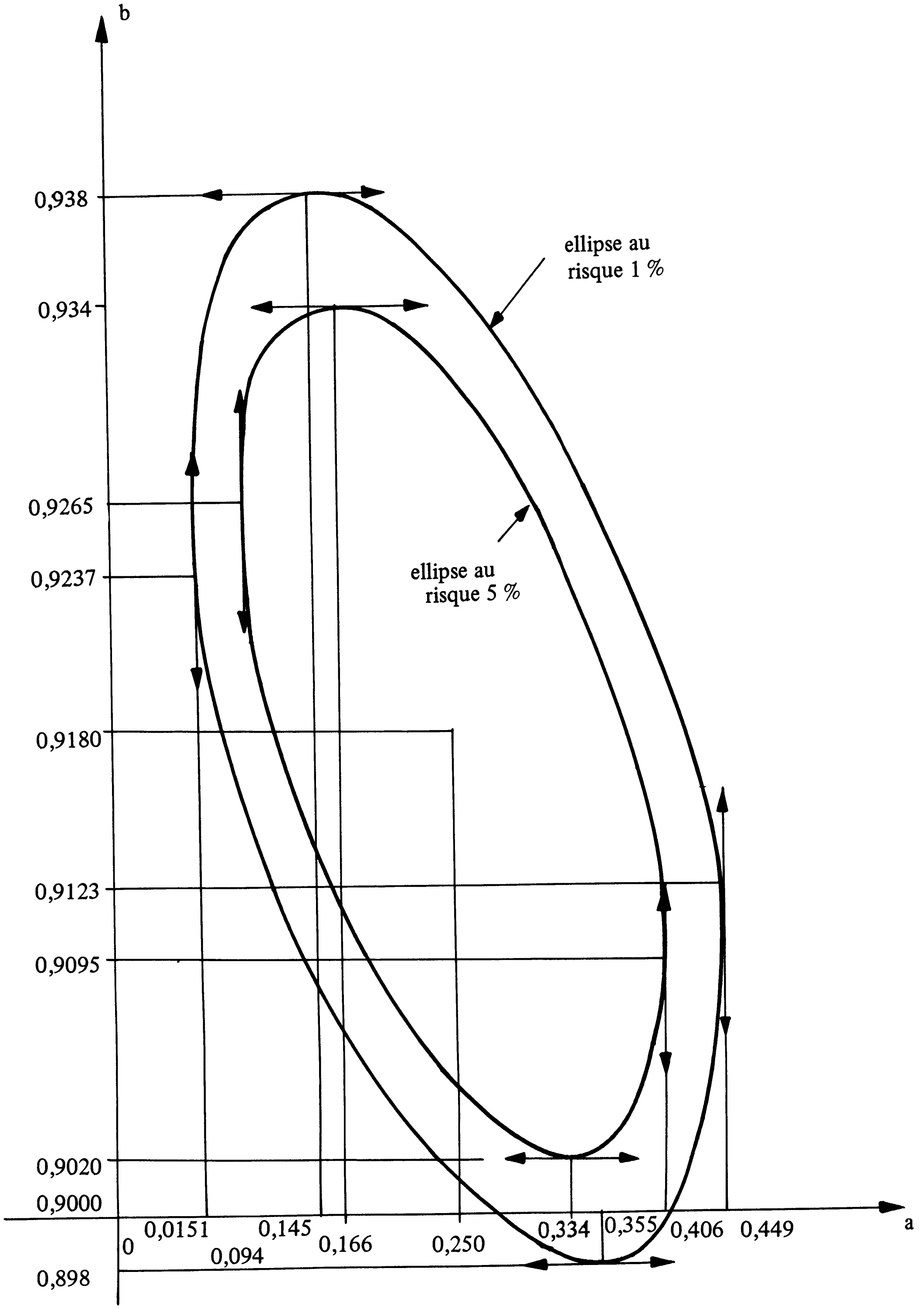
3) Estimation par intervalle de a et b

On trouve $s_a = 0{,}0715$ et $s_b = 0{,}0073$ si bien qu'au risque $\alpha = 5\ \%$, ($t_{0,05} = 2{,}04$ pour 33 degrés de liberté), on a :

$$0{,}104 < a < 0{,}396$$
$$0{,}903 < b < 0{,}933$$

4) Détermination de l'ellipse de confiance

Compte tenu de ce que $F_{0,05} = 3{,}29$ et $F_{0,01} = 5{,}23$ $\left(\begin{array}{l}\text{pour } \nu_1 = 2 \\ \text{et } \nu_2 = 33\end{array}\right)$


b
0,938
0,934
0,9265
0,9237
0,9180
0,9123
0,9095
0,9020
0,9000
0,898
ellipse au
risque 1 %
ellipse au
risque 5 %
0
0,0151
0,094
0,145
0,166
0,250
0,334
0,355
0,406
0,449
a

on obtient les équations des ellipses de confiance

– *au risque 5 %*

$$0{,}1789\,(a - 0{,}25)^2 + 1{,}8823\,(a - 0{,}25)\,(b - 0{,}918) + 17{,}2415\,(b - 0{,}918)^2 = 0{,}00313$$

– *au risque 1 %* :

$$0{,}1789\,(a - 0{,}25)^2 + 1{,}8823\,(a - 0{,}25)\,(b - 0{,}918) + 17{,}2415\,(b - 0{,}918)^2 = 0{,}00507$$

Nous représentons ces 2 ellipses à la page suivante.

Nous avons noté sur le graphique toutes les valeurs remarquables.

VI. SOLUTION PRATIQUE LORSQU'IL N'Y A PAS REPETITION DES EXPERIENCES A CHAQUE NIVEAU x_i

On est alors obligé de resoudre numériquement l'équation (IV.6) qui nous donné $\frac{\hat{\sigma}_0}{\hat{\lambda}}$. Une fois ce résultat obtenu, on obtient $\hat{a}$ et $\hat{b}$ par les formules (IV.1) et (IV.2), puis $\hat{\lambda}$ par la formule (IV.4)'. En première approximation, on peut donner des intervalles de confiance de la même manière que dans l'exemple ci-dessus et construire des ellipses de confiance.

BIBLIOGRAPHIE

R. SCHOLLER. – Contrôle de qualité en hormonologie. *Juin 1976,* p. 31-73.

J.A. JACQUEZ, F. J. MATHER et C. R. CRAWFORD. – Linear regression with non constant, unknown error variances : sampling experiments with least squares, weighted least squares and maximum likehood estimators. *Septembre 1978,* Biometrici p. 607-626.