

Estimateur de Wald

Y. BRUNET-MORET

Ingénieur hydrologue de l'ORSTOM
Bureau Central Hydrologique, Paris

RÉSUMÉ

Calcul des paramètres d'une droite représentant une liaison linéaire entre deux variables dont aucune (contrairement au cas de la régression linéaire simple) n'est connue sans erreur.

ABSTRACT

Estimation of parameters in linear model between two variates of which none (contrary to linear regression case) is known without error.

1. INTRODUCTION

Les régressions linéaires entre variables sont très utilisées, souvent peut-être abusivement. Quand on écrit :

$$Y = a X + b$$

en appelant cette formulation « régression de Y en X »,

Y variable expliquée (ou dépendante)

X variable explicative

on devrait lire que, pour toute valeur x_i de la variable explicative, l'espérance mathématique de la variable Y, liée à la valeur x_i est égale à « a » fois la valeur x_i parfaitement connue sans erreur, plus la valeur de « b ».

En hydrologie, l'utilisation de la formule ci-dessus, considérée comme fonctionnelle (alors qu'elle n'est que structurelle) pour extension, correction... de données d'observation ou calculées d'après observations, peut être souvent discutable car il est rare que les valeurs x_i puissent être considérées comme connues sans erreur, ou avec des erreurs négligeables.

Dans ce qui suit, nous allons étudier les relations linéaires structurelles entre deux variables et l'estimateur de Wald dont l'emploi facilite l'utilisation de ces relations structurelles. La documentation de base se trouve dans « The advanced theory of statistics » KENDALL et STUART, vol. II (1967), chap. 29, pages 375-418. Nous proposons ensuite une extension de ces relations linéaires structurelles dans le cas de deux variables explicatives ou plus.

2. RELATION LINÉAIRE STRUCTURELLE ENTRE DEUX VARIABLES

2.1. DÉFINITIONS ET HYPOTHÈSES DE BASE

2.1.1. Relation linéaire fonctionnelle, relation linéaire structurelle

La relation linéaire entre deux variables est fonctionnelle lorsque, pour tous les couples $\zeta_i \eta_i$ de ces variables, on peut écrire :

$$\eta_i = a \zeta_i + b$$

a et b étant des paramètres.

à une valeur de ζ correspond une seule valeur possible de η et inversement.

La relation devient structurelle entre les variables X et Y lorsque la variable observée x_i diffère de la valeur exacte ζ_i qu'on aurait dû observer d'une erreur δ_i

$$x_i = \zeta_i + \delta_i$$

et que la variable observée y_i diffère de la valeur exacte η_i qu'on aurait dû observer d'une erreur ε_i

$$y_i = \eta_i + \varepsilon_i$$

d'où

$$y_i - \varepsilon_i = a(x_i - \delta_i) + b$$

à une valeur x_i observée correspond un ensemble de valeurs ζ_i probables d'après la distribution de δ , à chaque valeur de ζ dans cet ensemble correspond une seule valeur de η et à chacune des valeurs possibles pour η correspond un ensemble de valeurs observables y_i d'après la distribution de ε .

La régression linéaire de Y en X est un cas très particulier dans lequel on suppose δ_i identiquement nul quel que soit i . De même, la régression linéaire de X en Y suppose ε_i nul quel que soit i .

2.1.2. Hypothèse de base

Les erreurs δ_i et ε_i sont des variables aléatoires indépendantes entre elles, indépendantes des valeurs observées x_i et y_i , de moyennes nulles, de variances indépendantes des valeurs δ_i , ε_i , x_i et y_i (homoscedasticité).

Il s'ensuit, que dans une représentation graphique, la distance du point $x_i y_i$ à la droite représentant la liaison structurelle sera douée de la qualité d'homoscedasticité.

Pour l'utilisation pratique de l'estimateur de Wald il faut ajouter les hypothèses, assez peu restrictives en fait, que les distributions des erreurs δ_i et ε_i puissent être considérées comme à peu près symétriques et de variances nettement inférieures respectivement à celle de X et à celle de Y, que les variables X puissent être considérées comme indépendantes entre elles de même que les variables Y.

2.2. PRINCIPES DU CALCUL DES PARAMÈTRES

En supposant que, dans l'échantillon de n couples $x_i y_i$

$$\sum \delta_i = 0 \quad \sum \varepsilon_i = 0 \quad \text{covar}(\delta, \varepsilon) = 0$$

nous pouvons écrire les treize équations suivantes :

$$\frac{1}{n} \sum x_i = \bar{x} = \bar{\zeta}$$

$$\frac{1}{n} \sum y_i = \bar{y} = \bar{\eta} = a \bar{x} + b$$

$$\frac{1}{n-1} \sum (x_i - \bar{x})^2 = \text{var } x = \text{var } \zeta + \text{var } \delta$$

$$\frac{1}{n-1} \sum (y_i - \bar{y})^2 = \text{var } y = \text{var } \eta + \text{var } \varepsilon = a^2 \text{var } \zeta + \text{var } \varepsilon$$

$$\frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y}) = \text{covar}(x, y) = \text{covar}(\zeta, \eta) = a \text{var } \zeta$$

les inconnues étant

a b

\bar{x} \bar{y} $\bar{\zeta}$ $\bar{\eta}$

var x var y var ζ var η var δ var ε

covar(x, y) covar(ζ , η)

Tel quel, ce système d'équations est indéterminé : il y a une inconnue de plus que d'équations, et il faut faire une hypothèse sur var ε ou var δ , mais on a toujours $b = \bar{y} - a \bar{x}$ et la droite représentative passe par \bar{x} , \bar{y} (fig. 1).

2.2.1. En prenant $\text{var } \delta = 0$, on retombe sur la régression linéaire de Y en X , la pente de la droite de régression étant $a_y = \frac{\text{covar}(x, y)}{\text{var } x}$

la droite s'écrivant $Y = a_y X + b$.

En prenant $\text{var } \varepsilon = 0$, on retombe sur la régression linéaire de X en Y , la pente de la droite de régression étant $a_x = \frac{\text{var } y}{\text{covar}(x, y)}$

la droite s'écrivant $Y = a_x X + b$

2.2.2. Nous pouvons toujours poser

$\text{var } \varepsilon = \lambda^2 \text{var } \delta$: la valeur de la pente de la droite de régression est

$$a = (\text{var } y - \lambda^2 \text{var } x + \sqrt{(\text{var } y - \lambda^2 \text{var } x)^2 + 4 \lambda^2 \text{covar}^2(x, y)}) / 2 \text{covar}(x, y)$$

qui est toujours comprise entre a_y et a_x .

Deux valeurs particulières de λ^2 sont intéressantes à considérer.

2.2.3. Avec $\lambda^2 = \frac{\text{var } y}{\text{var } x}$ c'est-à-dire lorsqu'on admet que les variances des erreurs sur X et sur Y sont proportionnelles aux variances respectives de x et de y

$$a = \sqrt{\text{var } y / \text{var } x} \quad \text{avec le signe de covar}(x, y)$$

la valeur de la pente est la moyenne géométrique de a_y et de a_x .

2.2.4. *Régression orthogonale*

Avec $\lambda^2 = 1$, c'est-à-dire $\text{var } \varepsilon = \text{var } \delta$, les variances des erreurs sur X et sur Y sont les mêmes, on trouve la régression orthogonale qui minimise la somme des carrés des distances des points représentatifs à la droite dont on cherche les paramètres.

$$a = (\text{var } y - \text{var } x + \sqrt{(\text{var } y - \text{var } x)^2 + 4 \text{covar}^2(x, y)}) / 2 \text{covar}(x, y)$$

En prenant pour équation de la droite

$$X \cos \varphi + Y \sin \varphi - p = 0$$

et en minimisant $\Sigma \overline{M H^2}$ (figure 1), on obtient

$$p = \bar{y} \cos \varphi + \bar{x} \sin \varphi$$

$$\text{tg } 2 \varphi = \frac{-2 \text{covar}(x, y)}{\text{var } y - \text{var } x} = \frac{-2 a}{a^2 - 1}$$

2.2.5. Nous avons représenté sur la même figure 1, X et Y étant supposés répartis normalement :

une ellipse d'égale densité de probabilité d'un couple $x_i y_i$

la droite de régression de X en Y D_Y

la droite de régression de Y en X D_X

la droite correspondant à $\lambda^2 = \frac{\text{var } y}{\text{var } x}$ Δ'

et la droite de régression orthogonale Δ

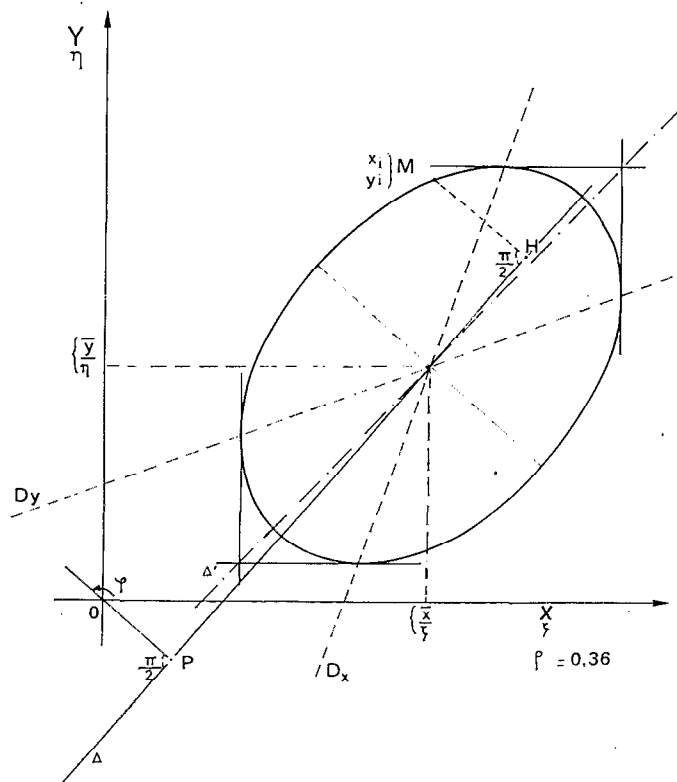


Fig. 1

2.2.6. Quelle que soit la valeur de a (ou de λ) choisie en fonction des connaissances a priori qu'on a des variances $\text{var } \delta$ et $\text{var } \varepsilon$

$$\frac{\text{covar}(x, y)}{\text{var } x} \leq a \leq \frac{\text{var } y}{\text{covar}(x, y)}$$

$$\text{var } \delta = \text{var } x - \text{covar}(x, y)/a$$

$$\text{var } \varepsilon = \text{var } y - a \text{ covar}(x, y)$$

2.2.7. On peut définir une « variance résiduelle apparente » comme étant la variance de la distance des points $M(x_i, y_i)$ à la droite représentative, quelle que soit la façon dont cette droite a été déterminée

$$\text{var}(\text{res}) = \frac{a^2 \text{var } \delta + \text{var } \varepsilon}{a^2 + 1} \quad (\text{minimale pour } \lambda^2 = 1)$$

Sur la figure 2 nous avons représenté cette variance résiduelle apparente pour des valeurs particulières de $\text{var } x$, $\text{var } y$ et $\text{covar}(x, y)$ en fonction de λ^2 ou de $1/\lambda^2$. L'allure générale de la courbe est indépendante des valeurs particulières choisies.

En supposant les erreurs δ et ε réparties à peu près normalement, la moitié des points $M(x_i, y_i)$ doit se trouver entre les deux droites parallèles à la droite déterminée à une distance orthogonale de

$$\frac{2}{3} \sqrt{\text{var}(\text{res})}$$

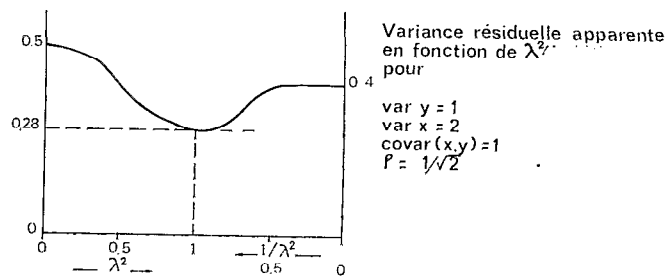


Fig. 2

Les distances de ces droites à la droite déterminée sont parallèlement à l'axe des Y :

$$\pm \frac{2}{3} \sqrt{a^2 \text{ var } \delta + \text{ var } \varepsilon}$$

parallèlement à l'axe des X :

$$\pm \frac{2}{3} \sqrt{\text{ var } \delta + \text{ var } \varepsilon / a^2}$$

On peut ainsi vérifier que les conditions d'homoscedasticité sont respectées : vérification grossière, mais suffisante dans la plupart des cas.

2.2.8. Intervalle de confiance (X, Y, δ et ε répartis normalement).

2.2.8.1. En posant

$$r = \text{covar}(x, y) / \sqrt{\text{var } x \text{ var } y}$$

la statistique

$$t = [(n-2)r^2 / (1-r^2)]^{1/2}$$

est distribuée suivant une loi de Student à $(n-2)$ degrés de liberté, et peut être utilisée pour tester l'hypothèse $a = 0$ (qui est l'hypothèse d'absence de liaison linéaire entre X et Y ou entre ζ et η).

2.2.8.2. Après avoir fait le changement de variable

$$y' = y / \sqrt{\text{var } \varepsilon / \text{var } \delta}$$

d'où

$$\text{var } y' = \text{var } y / (\text{var } \varepsilon / \text{var } \delta)$$

$$\text{covar}(x, y') = \text{covar}(x, y) / \sqrt{\text{var } \varepsilon / \text{var } \delta}$$

$$a' = (\text{var } y' - \text{var } x + \sqrt{(\text{var } y' - \text{var } x)^2 + 4 \text{covar}(x, y')}) / 2 \text{covar}(x, y')$$

les intervalles de confiance de $\theta' = \text{arc } t_g a'$ sont donnés par

$$\theta_c = \theta' \pm \frac{1}{2} \text{arc } \sin \left[2 t_c \sqrt{\frac{\text{var } x \text{ var } y' - \text{covar}^2(x, y')}{(n-2) [(\text{var } y' - \text{var } x)^2 + 4 \text{covar}(x, y')]} \right]$$

t_c étant la valeur de la variable de Student à $(n-2)$ degrés de liberté pour l'intervalle de confiance choisi, et à condition que la taille de l'échantillon soit assez grande pour que les deux angles θ' et θ_c soient, en valeurs absolues, inférieures à $\pi/4$.

2.2.9. Variance des variables « reconstituées »

Dans l'échantillon, nous pouvons écrire

(2)

$$\text{var } y = a^2 (\text{var } x - \text{var } \delta) + \text{var } \varepsilon$$

Après avoir choisi la valeur de $\lambda^2 = \text{var } \varepsilon / \text{var } \delta$ et calculé la pente a de la liaison linéaire (2.2.2.), nous pouvons calculer des valeurs y d'après des valeurs x soit observées, soit tirées au hasard dans la loi de distribution de X . Si la quantité $\text{var } x$ est conservée, de même que la valeur moyenne de x , la variance des valeurs y calculées est

$$\text{var}_c y = a^2 \text{var } x$$

La différence entre cette valeur et la variance de l'échantillon est

$$\text{var}_c y - \text{var } y = \text{var } \varepsilon \left(\frac{a^2}{\lambda^2} - 1 \right)$$

qui est nulle lorsque

$$\lambda^2 = a^2 = \text{var } y / \text{var } x$$

de valeurs absolues maximales pour

$$1/\lambda^2 = 0 \text{ régression de } Y \text{ en } X$$

$$\text{var}_c y - \text{var } y = -(1 - \rho^2) \text{var } y$$

$$\lambda^2 = 0 \text{ régression de } X \text{ en } Y$$

$$\text{var}_c y - \text{var } y = \frac{1 - \rho^2}{\rho^2} \text{var } y$$

ρ coefficient de corrélation linéaire entre x et y

Sur la figure 3, nous montrons la différence $\text{var}_c y - \text{var } y$ (pour les valeurs particulières de $\text{var } x$, $\text{var } y$ et $\text{covar}(x, y)$ de la figure 2) en fonction de λ^2 ou de $1/\lambda^2$.

L'allure générale de la courbe est indépendante des valeurs particulières choisies.

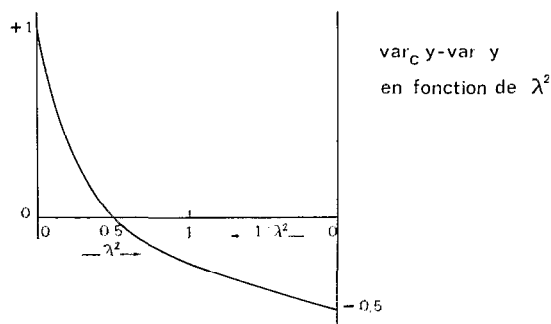


Fig. 3

3. ESTIMATEUR DE WALD

Il est souvent difficile de faire des hypothèses sur la valeur de λ . La méthode de Wald permet d'estimer λ sans préjuger de la valeur de λ qui peut ensuite se calculer.

3.1. Cet estimateur a été étudié dans le cas où le nombre d'observations n est pair et où on peut les partager en deux groupes de taille $n/2$.

groupe inférieur toutes les valeurs des x_i^- inférieures à la valeur médiane des x ,
toutes les valeurs correspondantes des y_i^- inférieures à la valeur médiane des y .
groupe supérieur toutes les valeurs des x_i^+ supérieures à la valeur médiane des x ,
toutes les valeurs correspondantes des y_i^+ supérieures à la valeur médiane des y .
(figures 4).

3.1.1. a se calcule aisément :

$$a = \frac{\text{valeur moyenne des } y_i^+ - \text{valeur moyenne des } y_i^-}{\text{valeur moyenne des } x_i^+ - \text{valeur moyenne des } x_i^-}$$

et la droite représentant la relation fonctionnelle passe par les

$$\text{points} \left\{ \begin{array}{l} \bar{x}^-, \bar{y}^- \\ \text{valeur moyenne des } x_i^-, \text{ valeur moyenne des } y_i^- \\ \bar{x}^+, \bar{y}^+ \\ \text{valeur moyenne des } x_i^+, \text{ valeur moyenne des } y_i^+ \end{array} \right.$$

avec $b = \bar{y} - a \bar{x}$

3.1.2. Il faut, pour pouvoir effectuer les calculs des 2.2.6, 2.2.7 et 2.2.9, que a satisfasse les inégalités

$$\frac{\text{covar } x, y}{\text{var } x} \leq a \leq \frac{\text{var } y}{\text{covar } x, y}$$

3.2. On ne peut pas toujours remplir les conditions de 3.1 pour le partage en deux groupes égaux. Si on peut faire deux groupes : (figures 4).

groupe inférieur m valeurs x_i^- inférieures à une valeur x_M ,
toutes les valeurs correspondantes y_i^- inférieures à une valeur y_M ,

groupe supérieur $n - m$ valeurs x_i^+ supérieures à la valeur x_M ,
toutes les valeurs correspondantes y_i^+ supérieures à la valeur de y_M .

il nous semble que si la valeur absolue de la différence entre m et $n - m$ soit $|2m - n|$ est inférieure ou égale à $1 \sqrt{n/5}$ on puisse calculer a comme en 3.1.1.

La droite représentant la relation fonctionnelle passe par le point \bar{x}, \bar{y} car

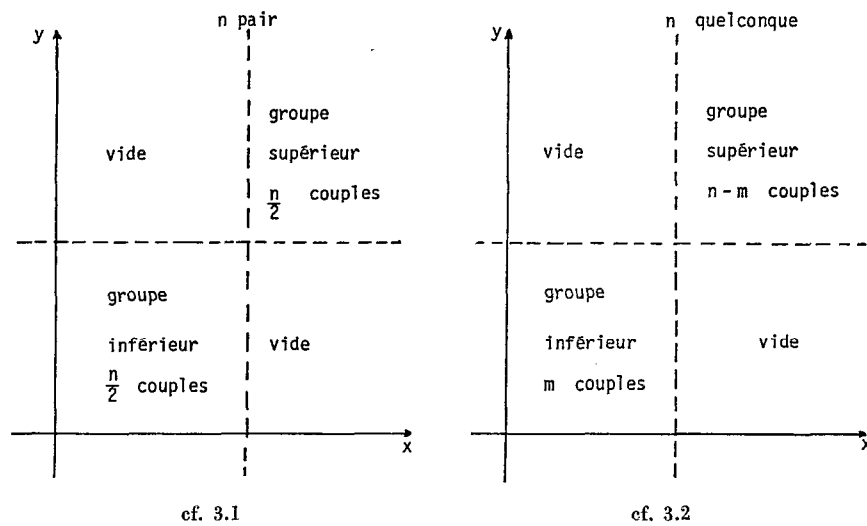
$$\frac{\text{somme des } y^+ - (n - m) \bar{y}}{\text{somme des } x^+ - (n - m) \bar{x}} = \frac{m \bar{y} - \text{somme des } y^-}{m \bar{x} - \text{somme des } x^-}$$

et les remarques de 3.1.2. sont valables.

3.3. Lorsque les conditions de 3.2. ne peuvent pas être remplies, il nous semble possible d'éliminer quelques-uns des couples médians pour obtenir deux groupes : (figures 4)

groupe inférieur m valeurs x_i^- inférieures à une valeur x_M ,
toutes les valeurs correspondantes y_i^- inférieures à une valeur y_M ,

groupe supérieur m' valeurs x_i^+ supérieures à la valeur x_M ,
toutes les valeurs correspondantes y_i^+ supérieures à la valeur y_M .



Nota : Le principe de l'élimination de quelques couples est, compte tenu des variances des erreurs δ et ϵ , de risquer le moins possible de mettre dans le groupe supérieur des couples pouvant appartenir au groupe inférieur et vice versa.

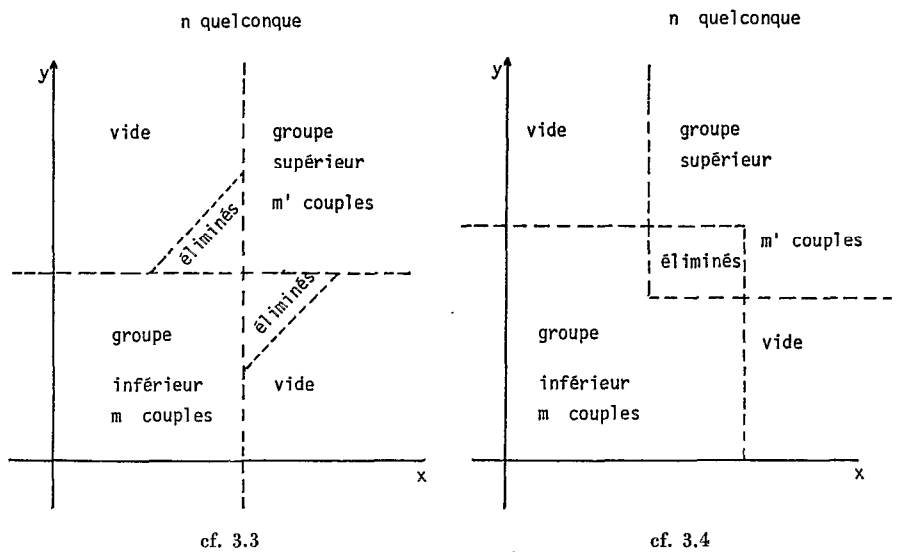


Fig. 4

3.3.1. Nous pensons qu'il faut éliminer le moins de points possible

$$n - (m + m') \leq \sqrt{n/5}$$

et que la valeur absolue de $(m - m')$ doit être inférieure ou égale à $\sqrt{n/5}$
 a peut alors être calculé comme en 3.1.1.

3.3.2. La droite représentant la relation fonctionnelle ne passera par le point \bar{x}, \bar{y} que si

$$\frac{\text{somme des } y^+ - m' \bar{y}}{\text{somme des } x^+ - m' \bar{x}} = \frac{m \bar{y} - \text{somme des } y^-}{m \bar{x} - \text{somme des } x^-}$$

Il suffit pratiquement de choisir les valeurs éliminées et les tailles m et m' des groupes pour vérifier graphiquement que la droite passe, à la précision du dessin, par le point \bar{x}, \bar{y} ; les remarques de 3.1.2 sont valables.

3.4. Lorsque les conditions de 3.3. ne peuvent pas être remplies, il nous semble possible d'éliminer quelques uns des couples médians pour obtenir deux groupes (figures 4) :

groupe inférieur m valeurs x_i^- inférieures à la plus forte valeur des couples éliminés, toutes les valeurs y_i^- correspondantes inférieures à la plus forte valeur des couples éliminés.

groupe supérieur m' valeurs x_i^+ supérieures à la plus faible valeur des couples éliminés, toutes les valeurs y_i^+ correspondantes supérieures à la plus faible valeur des couples éliminés.

Les paragraphes 3.3.1 et 3.3.2 restent alors valables.

3.5. Remarques très importantes sur l'estimation de Wald. Les conditions

$$\frac{\text{covar } x, y}{\text{var } x} \leq a \leq \frac{\text{var } y}{\text{covar } x, y}$$

sont facilement remplies lorsque la valeur de $|\rho|$ est voisine de 0,9. Si les conditions ne sont pas remplies, lorsque la valeur de $|\rho|$ est voisine de 1, on a avantage à faire les hypothèses 2.3 ou 2.4.

Les conditions 3.1. ou 3.2. sont facilement remplies lorsque $|\rho|$ est compris entre 0,9 et 0,95.

Si $|\rho|$ est inférieur à 0,85 il est possible qu'on ne pourra pas remplir même les conditions 3.4. car les variances des erreurs δ et ε sont trop importantes en valeurs relatives.

Si $|\rho|$ est supérieur à 0,95, on pourra probablement remplir les conditions 3.3. sauf si $|\rho|$ est supérieur à 0,98 il faudra alors probablement utiliser la méthode 3.4.

3.6. Intervalles de confiance (X, Y, δ et ε répartis normalement).

3.6.1. Intervalle de confiance pour la pente de la droite représentative. Dans le cas où les n (pair) observations sont partagées en deux groupes égaux (3.1.), en posant :

- x'_i une valeur du groupe inférieur
- \bar{x}' moyenne des $n/2$ valeurs de ce groupe
- y'_i la valeur correspondant à x'_i
- \bar{y}' moyenne des $n/2$ valeurs y'_i
- x''_i une valeur du groupe supérieur
- \bar{x}'' moyenne des $n/2$ valeurs de ce groupe
- y''_i la valeur correspondant à x''_i
- \bar{y}'' moyenne des $n/2$ valeurs y''_i

et en calculant :

$$S_x^2 = \frac{1}{n-2} [\Sigma (x'_i - \bar{x}')^2 + \Sigma (x''_i - \bar{x}'')^2]$$

$$S_y^2 = \frac{1}{n-2} [\Sigma (y'_i - \bar{y}')^2 + \Sigma (y''_i - \bar{y}'')^2]$$

$$S_{xy}^2 = \frac{1}{n-2} [\Sigma (x_i' - \bar{x}') (y_i' - \bar{y}') + \Sigma (x_i'' - \bar{x}'') (y_i'' - \bar{y}'')] \\ a = \frac{\bar{y}' - \bar{y}''}{\bar{x}' - \bar{x}''}$$

la quantité

$$t = \frac{(\bar{x}' - \bar{x}'') (\alpha - a)}{2} \sqrt{\frac{n}{S_y^2 + \alpha^2 S_x^2 - 2 \alpha S_{xy}^2}}$$

suivrait une distribution de Student à $n - 2$ degrés de liberté. Pour un intervalle de confiance choisi t est connu et les bornes (en valeur de la pente) sont les racines de l'équation en α

$$4 t^2 (S_y^2 + \alpha^2 S_x^2 - 2 \alpha S_{xy}^2) - n (\bar{x}' - \bar{x}'') (\alpha - a) = 0$$

3.6.2. Région de confiance pour les droites représentatives.

En calculant

$$b = \bar{y} - a \bar{x} = (\bar{y}' \bar{x}' - \bar{y}'' \bar{x}'') / (\bar{x}' - \bar{x}'')$$

la quantité

$$F = \frac{n}{2} \left[\frac{(\bar{x}' - \bar{x}'')^2}{4} (a - \alpha)^2 + (b - \beta)^2 \right] / (S_y^2 + \alpha^2 S_x^2 - 2 \alpha S_{xy}^2)$$

serait distribuée suivant une loi de Fischer-Snedecor avec 2 et $(n - 2)$ degrés de liberté, ce qui permet de calculer β et donc les intervalles de confiance du terme constant de la droite représentative.

Pour un intervalle de confiance choisi, F est connu, et en partant de l'équation d'une droite :

$$Y = \alpha X + \beta$$

$$F (S_y^2 + \alpha^2 S_x^2 - 2 \alpha S_{xy}^2) = \frac{n}{2} \left[\frac{(\bar{x}' - \bar{x}'')^2}{4} (a - \alpha)^2 + (b + \alpha X - Y)^2 \right]$$

en éliminant α entre l'équation ci-dessus et sa dérivée par rapport à α

$$2 F (\alpha S_x^2 - S_{xy}^2) = n \left[- \frac{(\bar{x}' - \bar{x}'')^2}{4} (a - \alpha) + X (b + \alpha X - Y) \right]$$

on obtient l'enveloppe contenant, pour l'intervalle de confiance choisi, toutes les droites représentatives de la liaison structurelle linéaire entre les couples $x_i y_i$.

Les intersections de cette enveloppe avec une droite $X = x_u$ (ou une droite $Y = y_u$) donnent l'intervalle de confiance (celui choisi pour la valeur de F) en y lié à la valeur x_u (ou en x lié à la valeur y_u).

Il serait possible d'utiliser ce procédé pour calculer le gain de précision sur la valeur moyenne de Y lorsqu'en plus des n observations des couples X, Y on dispose de k observations supplémentaires des X .

3.7. Il semble possible d'obtenir une limite inférieure de ce gain en utilisant son estimation obtenue par M. Veron dans le cas de la régression linéaire de Y en X (variables réparties normalement). Dans ce cas, toutes les erreurs d'observations sont affectées à Y alors que l'estimateur de Wald affecte une partie de ces erreurs à X .

3.7.1. L'estimation de M. Veron est la suivante : échantillon de n couples X, Y d'après lequel on a calculé

$$\bar{x}, \bar{y}, \text{var } y, \text{var } x, \text{covar } (x, y)$$

et le coefficient de corrélation

$$\rho = \text{covar } (x, y) / \sqrt{\text{var } x \text{ var } y}$$

la variance de \bar{y} est $\text{var } \bar{y} = \frac{\text{var } y}{n}$

La prise en compte de k valeurs supplémentaires de x permet de calculer la valeur moyenne \bar{x}_t sur l'échantillon de $(n + k)$ valeurs, et de calculer une valeur moyenne, sans distorsion

$$\bar{y}_t = a' \bar{x}_t + b \quad a' = \rho \sqrt{\frac{\text{var } y}{\text{var } x}} \quad b = \bar{y} - a' \bar{x}$$

dont la variance est

$$\text{var } \bar{y}_t = \frac{\text{var } y}{n} \left[1 - \rho^2 \frac{k}{n + k} + \frac{k}{(n + k)(k - 3)} (1 - \rho^2) \right]$$

L'efficacité relative de \bar{y}_t par rapport à \bar{y} est

$$E = \frac{\text{var } \bar{y}_t}{\text{var } \bar{y}}$$

et l'estimateur \bar{y}_t a la même efficacité que si cette moyenne avait été calculée sur un échantillon observé de $\frac{n}{E}$ valeurs.

3.7.2. On peut aussi obtenir une limite supérieure du gain : en partant des définitions du paragraphe 2.2.

$$\text{var}' \bar{y}_t = \frac{1}{n + k} [a^2 (\text{var } x - \text{var } \delta) + \text{var } \varepsilon]$$

expression qui ne tient pas compte des variances d'estimation des paramètres a et b de la droite.

3.7.3. Il semble qu'on puisse se contenter, lorsque λ^2 est voisin de 1, c'est-à-dire lorsque $\frac{\text{var } \delta}{\text{var } \varepsilon}$ est peu différent de $\frac{\text{var } x}{\text{var } y}$ dans l'échantillon de taille n , de calculer une variance approchée de la moyenne \bar{y}_t calculée avec la prise en compte de K valeurs supplémentaires de x par

$$\text{var}'' \bar{y}_t = \frac{1}{n + k} (a^2 \text{var } x + \text{var } \varepsilon)$$

4. RELATION LINÉAIRE STRUCTURELLE ENTRE TROIS VARIABLES OU PLUS

4.1. AVEC DEUX VARIABLES EXPLICATIVES

4.1.1. La relation linéaire entre une variable et deux autres est fonctionnelle lorsque pour tous les couples $\zeta_{1i} \zeta_{2i}$ η_i de ces variables, on peut écrire :

$$\eta_i = a_1 \zeta_{1i} + a_2 \zeta_{2i} + b = a (\zeta_{1i} + \beta \zeta_{2i}) + b = a \zeta_i + b$$

a_1 a_2 et b étant des paramètres.

La relation devient structurelle lorsque la variable observée x_{1i} diffère de la valeur exacte ζ_{1i} qu'on aurait dû observer d'une erreur δ_{1i}

$$x_{1i} = \zeta_{1i} + \delta_{1i}$$

de même

$$x_{2i} = \zeta_{2i} + \delta_{2i}$$

et

$$y_i = \eta_i + \varepsilon_i$$

d'où

$$y_i - \varepsilon_i = a_1 (x_{1i} - \delta_{1i}) + a_2 (x_{2i} - \delta_{2i}) + b$$

que l'on peut écrire plus simplement

$$y_i - \varepsilon_i = a_1 x_{1i} + a_2 x_{2i} - \delta_i + b$$

La régression linéaire multiple de Y en X_1 et X_2 est un cas très particulier dans lequel on suppose δ_{1i} et δ_{2i} identiquement nuls quel que soit i.

4.1.2. Hypothèse de base

Les erreurs ε_i et δ_i sont des variables aléatoires indépendantes entre elles, indépendantes des valeurs observées x_{1i} , x_{2i} et y_i , de moyennes nulles, de variances indépendantes des valeurs δ_i , ε_i , x_{1i} , x_{2i} et y_i (homoscedasticité).

En posant :

$$a_1 x_{1i} + a_2 x_{2i} = a (x_{1i} + \beta x_{2i}) = a x_i$$

β et a étant des paramètres ; il s'ensuit que, dans une représentation graphique, la distance du point $x_i y_i$ à la droite représentant la relation structurelle

$$Y = a X + b = a (X_1 + \beta X_2) + b$$

sera doublée de la qualité d'homoscedasticité.

Pour l'utilisation pratique de l'estimateur de Wald, il faut ajouter les hypothèses, assez peu restrictives en fait, que les distributions des erreurs δ_i et ε_i puissent être considérées comme à peu près symétriques, et de variances nettement inférieures respectivement à celle de $(X_1 + \beta X_2)$ et à celle de Y, que les variables X_1 puissent être considérées comme indépendantes entre elles, de même que les variables X_2 et que les variables Y. S'il y a une liaison entre les variables X_1 et X_2 , il faut que cette liaison puisse être considérée comme linéaire.

4.2. PRINCIPE DU CALCUL DES PARAMÈTRES

4.2.1. On voit que, lorsque β a été déterminé, on est ramené à la relation linéaire structurelle entre deux variables Y et X et la détermination de a peut être faite par l'estimateur de Wald : la valeur de b est

$$b = \bar{y} - a_1 \bar{x}_1 - a_2 \bar{x}_2 = \bar{y} - a \bar{x}$$

Supposons β et a déterminés : nous pouvons écrire (2.2.6)

$$\text{var } \delta = \text{var } x - \text{covar } (x, y)/a$$

$$\text{var } \varepsilon = \text{var } y - a \text{ covar } (x, y)$$

avec

$$\text{var } x = \text{var } x_1 + \beta^2 \text{var } x_2 + 2 \beta \text{covar } (x_1, x_2)$$

$$\text{covar } (x, y) = \text{covar } (x_1, y) + \beta \text{covar } (x_2, y)$$

Les variances expliquées par la relation structurelle sont :

$$\text{pour } Y : \text{var } y - \text{var } \varepsilon$$

$$\text{pour } X : \text{var } x - \text{var } \delta$$

le rapport de la variance généralisée expliquée à la variance généralisée globale est :

$$\frac{(\text{var } y - \text{var } \varepsilon) (\text{var } x - \text{var } \delta)}{\text{var } y \text{ var } x} = \frac{[(\text{covar } (x, y))]^2}{\text{var } x \text{ var } y}$$

4.2.2. La valeur maximale de ce rapport est celle du carré du coefficient de corrélation multiple entre Y et X_1, X_2 . Nous proposons donc de calculer β comme dans les régressions multiples :

$$\beta = \frac{\text{var } x_1 \text{ covar } (x_2, y) - \text{covar } (x_1, y) \text{ covar } (x_1, x_2)}{\text{var } x_2 \text{ covar } (x_1, y) - \text{covar } (x_2, y) \text{ covar } (x_1, x_2)}$$

La détermination de a se fera ensuite par l'estimateur de Wald (3) en utilisant les couples $x_i y_i$ (où $x_i = x_{1i} + \beta x_{2i}$).

4.3. AUTRE DÉTERMINATION DE a

Si $\lambda^2 = \frac{\text{var } \varepsilon}{\text{var } \delta}$ est connu,

avec $\text{var } \delta = \text{var } \delta_1 + \beta^2 \text{var } \delta_2 + 2 \beta \text{covar } (\delta_1, \delta_2)$

ce qui suppose β connu a priori, on peut utiliser l'équation du paragraphe 2.2.2. pour calculer a .

Dans les cas qui nous intéressent, cette démarche ne semble pas réaliste... et on peut faire bien d'autres hypothèses sans intérêt.

4.4. RELATION LINÉAIRE STRUCTURELLE, ESTIMATEUR DE WALD AVEC PLUS DE DEUX VARIABLES « EXPLICATIVES »

La généralisation à plus de deux variables explicatives du cas de deux variables explicatives est immédiate : en écrivant

$$\begin{aligned} y_i - \varepsilon_i &= a (x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}) - \delta_i + b \\ &= a x_i - \delta_i + b \end{aligned}$$

On en déduit, lorsque a et $\beta_2 \dots \beta_k$ ont été déterminés

$$b = \bar{y} - a (\bar{x}_1 + \beta_2 \bar{x}_2 \dots + \beta_k \bar{x}_k) = \bar{y} - a \bar{x}$$

Les β sont calculés comme dans les régressions multiples en cherchant le minimum de

$$\sum [y_i - a' (x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}) - b]^2$$

a' étant ici un paramètre auxiliaire différent de a

a est déterminé par l'estimateur de Wald en utilisant les couples $x_i y_i$ avec l'indice

$$x_i = x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}$$

5. EXEMPLE D'APPLICATION

5.1. PREMIER EXEMPLE

Un très bon exemple d'application, avec justification de la méthode décrite en 2.2.4. « Régression orthogonale », est donné dans l'article Nouvelot. *Cahier Hydrologie*, vol. IX, n° 1, 1972. Nous n'avons malheureusement pas pu retrouver les valeurs originales pour reprendre et donner ici l'ensemble des calculs.

5.2. DEUXIÈME EXEMPLE D'APPLICATION

Etude des lames écoulées du bassin intermédiaire entre les stations de Bab Ouender et d'Ourtzagh (Ouergha) en fonction des précipitations à Jbel Ourka en années hydrologiques. Le bassin intermédiaire est de 2 646 km², les lames écoulées sont calculées d'après la différence des modules aux deux stations qu'il est inutile de corriger des variations des réserves naturelles entre le début et la fin de l'année hydrologique. Le poste pluviométrique de Jbel Ourka est situé au milieu de la frontière ouest du bassin

n = 19 années communes

	<i>Groupe inférieur</i>			<i>Groupe supérieur</i>	
	<i>Jbel Ourka</i>	<i>lame</i>		<i>Jbel Ourka</i>	<i>lame</i>
	mm	mm		mm	mm
1956-57	1 056	99	1955-56	2 433	968
57-58	1 626	379	58-59	1 899	638
60-61	1 452	270	59-60	2 400	1 067
64-65	1 645	501	61-62	2 003	663
65-66	1 572	463	62-63	3 171	1 429
66-67	1 209	155	63-64	2 365	879
67-68	1 480	389	68-69	2 508	1 008
71-72	1 625	431	69-70	2 489	963
72-73	1 062	157	70-71	2 005	687
73-74	1 679	486			

5.2.1. On se trouve dans les conditions de (3.2.) et il n'y a aucun point à éliminer (figure 5)

<i>Sur 19 ans</i>		<i>Jbel Ourka X</i>		<i>Bassin Y</i>
Valeurs	} somme } moyenne	35 679	1 877,8421	11 632 612,2105
Somme des carrés	variance	72 735 491	$(564,5038)^2 = 318 664,5850$	9 465 874 $(360,9125)^2 = 130 257,8421$
Somme des produits	covariance		25 479 132 202 004,0350	
Coefficient de corrélation		0,9915		
10 valeurs inférieures	} somme } moyenne	14 406	1 440,6000	3 330 333,0000
9 valeurs supérieures	} somme } moyenne	21 273	2 363,6667	8 302 922,4444

5.2.2. En choisissant la représentation

$$\text{Lame bassin} = a (\text{précipitation Jbel Ourka} - P_0)$$

$$\text{on doit avoir (2.2.6, 3.1.2)} \frac{\text{covar}(x, y)}{\text{var } x} = 0,6339 \leq a \leq 0,6448 = \frac{\text{var } y}{\text{covar}(x, y)}$$

5.2.3. L'estimateur de Wald conduit à

$$\text{Lame} = 0,6385 (\text{Jbel Ourka} - 919,0)$$

La valeur de « a » vérifiant les inégalités ci-dessus.

Les variances des erreurs (2.2.6.) sont

$$\text{sur précipitation Jbel Ourka } \text{var } x - \text{covar } (x, y)/a = 2\,291,78 = \text{var } \delta = (47,87)^2$$

$$\text{sur lame } \text{var } y - a \text{ covar } (x, y) = 1\,278,26 = \text{var } \varepsilon = (35,75)^2$$

5.2.4. Sur les 19 années observées, la moyenne de la lame est de 612,2 mm, avec un écart-type de $\frac{360,91}{\sqrt{19}} = 82,8$ mm

En tenant compte de toutes les années observées à Jbel Ourka (41 ans = 19 + 22), la moyenne des précipitations est de 1 869,5 mm et la variance des précipitations est de $(574,58)^2$: on peut admettre que la lame moyenne sur 41 ans est de

$$0,6385 (1\,869,5 - 919,0) = 606,9 \text{ mm}$$

avec un écart-type (3.7.3) de $\sqrt{\frac{1}{41} (0,6385^2 \text{ var } x + \text{var } \varepsilon)} = 56,6$ mm.

5.2.5. Remarque

La précipitation à Jbel Ourka doit être considérée comme un « indice pluviométrique » : ce n'est pas la précipitation moyenne sur le bassin. Cet indice est en liaison qui semble linéaire et assez étroite avec la lame écoulée, et probablement, pour les valeurs autour de sa moyenne, en liaison linéaire avec la pluviométrie moyenne sur le bassin, ce qui justifie son emploi pour améliorer la connaissance de la valeur moyenne de la lame écoulée.

Pour tenir compte des effets d'« abattement » et pour éviter de calculer des lames écoulées trop fortes ou négatives d'après des indices pluviométriques de fréquences rares (fortes ou faibles) je propose, pour des cas semblables à celui de cet exemple, une approximation de la liaison lame écoulée Q et indice pluviométrique P par une cubique de calculs simples

$$a^3 P^3 = 2 Q^3 + 6 a^2 P_0^2 Q$$

a et P_0 ayant les valeurs calculées d'après l'estimateur de Wald (5.2.2., 5.2.3).

Cette cubique, bi-univoque, est symétrique par rapport à $P = 0$, $Q = 0$ (point d'inflexion à tangente nulle). Elle admet un point d'inflexion en $P = 2 P_0$, $Q = a P_0$ dont la tangente est la droite d'estimation de Wald $Q = a (P - P_0)$ et la cubique est confondue avec cette droite sur une bonne distance (figure 5). L'asymptote de la cubique a pour équation : $Q = \frac{a P}{2^{1/3}}$.

Les solutions de cette cubique sont :

pour Q connu

$$P = \frac{1}{a} (2 Q^3 + 6 a^2 P_0^2 Q)^{1/3}$$

pour P connu

$$Q = a P_0 \left([\sqrt{1+z^2} + z]^{1/3} - [\sqrt{1+z^2} - z]^{1/3} \right), \text{ avec } z = \frac{P^3}{4 P_0^3}$$

5.3. TROISIÈME EXEMPLE D'APPLICATION

Etude des modules à Pont du M'dez (Sebou) en fonction des précipitations à Fez Ain Kaddous, en années hydrologiques. Les modules au Pont du M'dez sont corrigés des variations des réserves naturelles entre le début

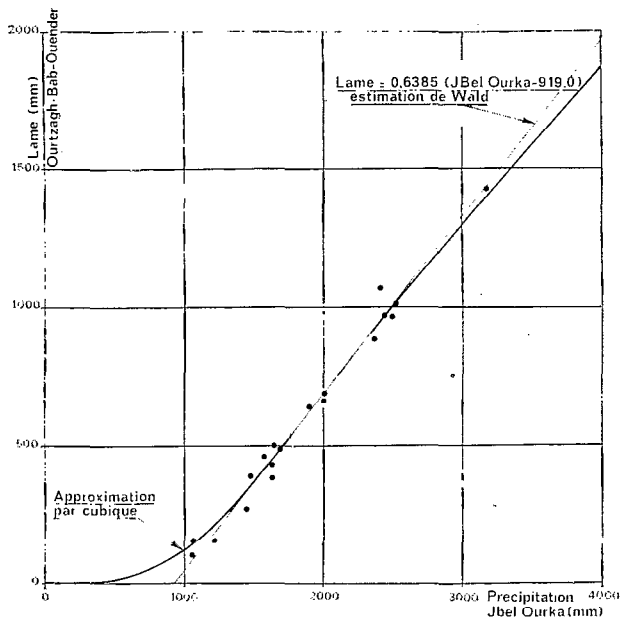


Fig. 5. — Lames écoulées du bassin intermédiaire Ourtzagh-Bab Ouender en fonction des précipitations à Jbel-Ourka

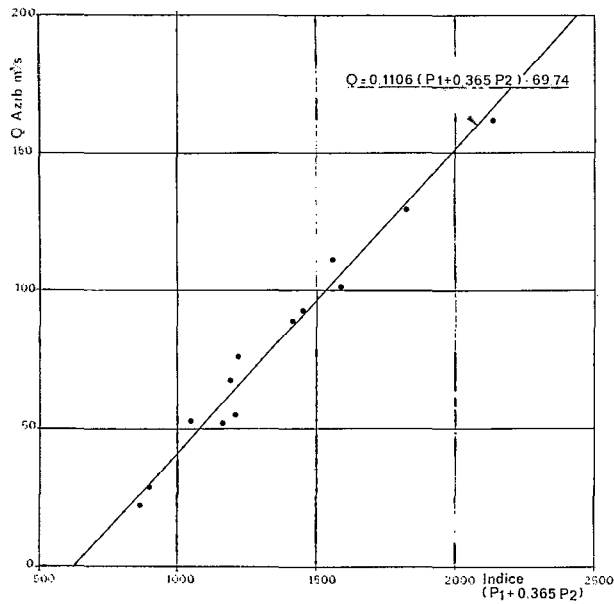


Fig. 7. — Module du Sebou à Azib es Soltame en fonction d'un indice pluviométrique (P_1 = Fez Ain Kaddous, P_2 = Jbel Ourka)

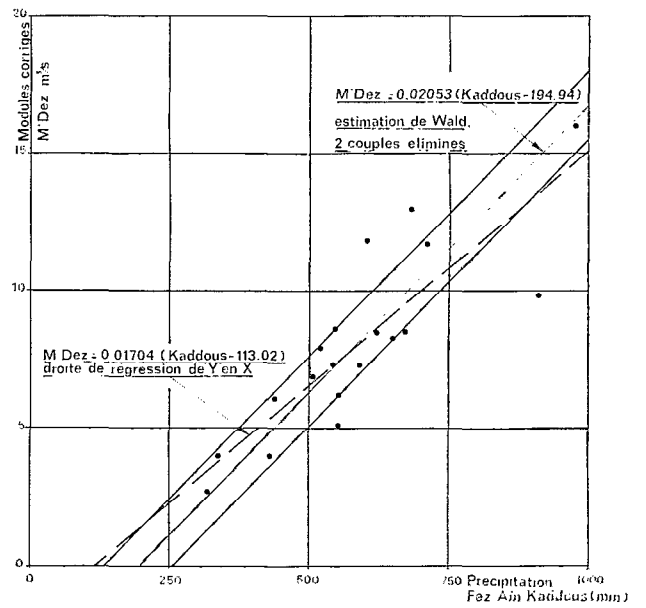


Fig. 6. — Modules à pont du M'Dez en fonction des précipitations à Fez Ain Kaddous

et la fin de l'année hydrologique. Le poste pluviométrique de Fez Aïn Kaddous se trouve à 60 km au nord nord-ouest du point le plus proche du bassin, à 80 km de son centre (bassin de 3 109 km²).

n = 19 années communes.

	Groupe inférieur		Éliminés		Groupe supérieur			
	Kaddous	M'dez	Kaddous	M'dez	Kaddous	M'dez		
	mm	m ³ /s	mm	m ³ /s	mm	m ³ /s		
1956-57	337	4,0	1963-64	591	7,3	1959-60	711	11,7
57-58	553	6,2	72-73	547	8,6	62-63	976	16,0
58-59	552	5,1				64-65	622	8,5
60-61	522	7,9				67-68	672	8,5
61-62	437	6,1				68-69	910	9,8
63-64	591	7,3				69-70	650	8,3
65-66	321	2,7				70-71	683	13,0
66-67	430	4,0				71-72	603	11,8
73-74	544	7,3				72-73	547	8,6
74-75	506	6,9						

5.3.1. Il semble préférable d'éliminer deux couples pour se placer dans les conditions de (3.3.). On pourrait ne rien éliminer et se placer dans les conditions de (3.4.) car, étant donné la forte variance d'erreur δ sur la précipitation, il est probable que le point 1963-64 appartient au groupe inférieur et que le point 1972-73 appartient au groupe supérieur.

Kaddous « X » sur 63 ans

moyenne	576,4 mm
variance	19 126,89
écart type	138,3 mm

Sur 19 ans	Kaddous X		M'dez Y	
Valeurs } somme	11 165		153,7	
} moyenne		587,6316		8,0895
Somme des carrés	7 051 119		1 442,07	
variance		27 234,0103		11,0397
Somme des produits			98 674,4	
covariance			464,1902	
Coefficient de corrélation	0,8466			
9 valeurs } somme	4 200		50,2	
inférieures } moyenne		466,6667		5,5778
2 valeurs } somme	1 138		15,9	
éliminées } moyenne		569,0000		7,9500
8 valeurs } somme	5 827		87,6	
supérieures } moyenne		728,3750		10,9500
10 valeurs } somme	4 791		57,5	
inférieures } moyenne		479,1000		5,7500
9 valeurs } somme	6 374		96,2	
supérieures } moyenne		708,2222		10,6889

5.3.2. En choisissant la représentation

Module (corrigé) M'dez = a (précipitation Kaddous — P₀) on doit avoir

$$\frac{\text{covar}(x, y)}{\text{var } x} = 0,0170 \leq a \leq 0,0238 = \frac{\text{var } y}{\text{covar}(x, y)} \tag{2.2.6}$$

5.3.3. (Deux couples éliminés), l'estimation de Wald conduit à

$$M'dez = 0,02053 \text{ (Kaddous — 194,94)} \tag{figure 6}$$

cette droite passe près du point $\bar{x} \bar{y}$ des 19 valeurs

pour $\bar{x} = 587,6316$		M'dez = 8,06
	elle donne	
pour $\bar{y} = 8,0895$		Kaddous = 588,97

Les variances des erreurs sont :

sur Kaddous	var x — covar(x, y)/a = 4 623,67 = var δ = (68,00) ²	
sur M'dez	var y — a covar(x, y) = 1,51 = var ε = (1,23) ²	

la variance résiduelle apparente est, aux échelles de la figure 6 :

$$\frac{a \text{ var}' \delta + \text{var}' \varepsilon}{a^2 + 1} = 1,684 \text{ cm}^2$$

On peut tracer deux droites parallèles à la droite représentative à des distances, en Y, de

$$\pm \frac{2}{3} \sqrt{a^2 \text{ var } \delta + \text{var } \varepsilon} = \pm 1,24 \text{ m}^3/\text{s} \tag{2.2.7.}$$

La figure 6 montre que les conditions d'homoscedasticité ne sont peut-être pas très bien remplies.

Sur 19 ans, l'écart-type de la moyenne observée à M'dez (8,09 m³/s) est $\sqrt{\text{var } y/19} = 0,76 \text{ m}^3/\text{s}$.

La moyenne à M'dez, d'après la relation et la moyenne de 63 ans à Kaddous, est de 7,83 m³/s à laquelle on peut affecter comme valeur supérieure d'écart-type (3.7.1. avec m = 44, n = 19) 0,54 m³/s, et comme valeur inférieure d'écart-type 0,42 m³/s.

5.3.4. Pas de couples éliminés : l'estimation de Wald conduit à :

$$M'dez = 0,02156 \text{ (Kaddous — 212,35)}$$

qui passe par $\bar{x} \bar{y}$, les variances des erreurs sont :

sur Kaddous	5 703,85 = var δ = (75,52) ²	
sur M'dez	1,03 = var ε = (1,016) ²	

la variance résiduelle apparente est de 1,703 cm².

5.3.5. La droite de régression de Y en X s'écrit (figure 6)

$$M'dez = 0,01704 \text{ (Kaddous — 113,02)}$$

passant par $\bar{x} \bar{y}$, et les variances des erreurs sont :

sur Kaddous	identiquement nulle	
sur M'dez	(1 — ρ ²) var y = 3,127 = (1,77) ² = var ε	

la variance résiduelle apparente est de 1,812 cm².

5.3.6. Nous proposerions de prendre la représentation 5.3.3. ou la représentation 5.3.2. (ou une moyenne entre les deux :

$$M^{\text{dez}} = 0,021 \text{ (Kaddous — 204,4)}$$

avec l'interprétation, presque exacte, suivante :

L'indice pluviométrique « précipitation à Fez Aïn Kaddous » représente (à un facteur c près) la précipitation annuelle moyenne (spatiale) sur le bassin du Sebou à Pont du M^{dez}, avec un écart-type $\sqrt{\text{var } \delta}$ (au même facteur c près). Le module annuel est affecté d'un écart-type d'estimation $\sqrt{\text{var } \varepsilon}$ intégrant, d'une part, les erreurs (notamment de tarages), d'autre part, les variations des conditions d'écoulement (intensités, fréquences, localisations des précipitations).

En reconstituant des modules (corrigés) à M^{dez} par l'utilisation de la longue série de précipitations à Fez Aïn Kaddous, la variance de ces modules serait $\left(\frac{0,01704}{0,02053}\right)^2 = 0,69$ fois plus faible en reconstituant par la régression linéaire de 5.3.5. qu'en reconstituant par la relation 5.3.3.

5.4. QUATRIÈME EXEMPLE

Modules du Sebou à Azib es Soltame (corrigés des variations de réserves d'une année sur l'autre) en fonction d'un indice pluviométrique calculé d'après les précipitations annuelles à Fez Aïn Kaddous et à Jbel Ourka.

	Modules Azib $Q \text{ m}^3/\text{s}$	Précipitations Fez $P_1 \text{ mm}$	Jbel Ourka $P_2 \text{ mm}$	Indice $P_1 + 0,365 P_2$
1959-60	101	711	2 400	1 587,00
60-61	52,4	522	1 452	1 051,98
61-62	51,8	437	2 003	1 168,10
62-63	162	976	3 171	2 133,42
63-64	92,4	591	2 365	1 454,23
64-65	76,0	622	1 645	1 222,43
65-66	29,2	321	1 572	894,78
66-67	22,8	430	1 209	871,29
67-68	54,4	672	1 480	1 212,20
68-69	130	910	2 508	1 825,42
69-70	111	650	2 489	1 558,49
70-71	88,4	684	2 005	1 415,83
71-72	67,6	603	1 625	1 196,13

Moyennés sur 13 ans	$Q = 79,9231$	$P_1 = 625,3077$	$P_2 = 1 994,1538$
Variances	$Q = 1 590,4036$	$P_1 = 33 014,8923$	$P_2 = 317 288,4075$
Racine carrée variance	$Q = 39,8799$	181,7000	563,2836
Covariances	$QP_1 = 6 625,8756$	$QP_2 = 20 912,7462$	$P_1P_2 = 76 287,2300$
Coefficient de corrélation	$\rho_{QP_1} = 0,9144$	$\rho_{QP_2} = 0,9310$	$\rho_{P_1P_2} = 0,7454$

5.4.1. $\beta(4.2.2.) = 0,365$

$$\text{Coefficient de corrélation multiple : } \rho^2 = \frac{\rho^2_{QP_1} + \rho^2_{QP_2} - 2 \rho_{QP_1} \rho_{QP_2} \rho_{P_1P_2}}{1 - \rho^2_{P_1P_2}}$$

d'où $\rho = 0,9880$

var (indice) 130 975,3183 = (361,9051)², moyenne indice 1 353,17

covar (Q, indice) 14 259,0280

moyennes des six valeurs les plus fortes Q 114,1333 indice 1 662,40

moyennes des sept valeurs les plus faibles Q 50,6000 indice 1 088,13

d'où a = 0,1106 et Q = 0,1106 indice — 69,74 (figure 7)

$$\text{var } \delta = 2 051,0145 = (45,2881)^2$$

$$\text{var } \varepsilon = 13,3551 = (3,6545)^2$$

5.4.2. En n'utilisant que les précipitations à Jbel Ourka comme indice pluviométrique :

moyennes des six valeurs les plus fortes Q 114,1333 P₂ 2 489,667

moyennes des sept valeurs les plus faibles Q 50,6000 P₂ 1 569,4286

d'où a = 0,06904 et Q = 0,06904 P₂ — 57,75

$$\text{var } \delta' = 14 380,7279 = (119,9197)^2$$

$$\text{var } \varepsilon' = 146,5876 = (12,1073)^2$$

l'écart-type de ε' est 3,3 fois plus fort que celui de ε : on voit l'avantage obtenu en utilisant deux stations pluviométriques au lieu d'une.