CHAPITRE III

DECOUVERTE DES ESTIMATEURS LS ET LAD

III.1- HISTORIQUE DE L'ESTIMATION LAD (1757-1955) :

Parmi les estimateurs robustes, les estimateurs LAD ont probablement l'histoire la plus ancienne. En effet, Ronchetti (1987) mentionne qu'on en retrouve des traces dans l'oeuvre de Galilée (1632), intitulée *"Dialogo dei massimi sistemi"*, Le problème était alors de déterminer la distance de la terre à une étoile récemment découverte à cette époque. C'est cependant à Boscovich (1757) que l'on reconnaît généralement l'introduction du critère d'estimation LAD (Harter,1974 ; Ronchetti, 1987, Dielman,1992).

L'un des problèmes qui excita le plus, la curiosité des hommes de science du *XVIII*^{ème} siècle fut celui de la détermination de l'ellipticité de la terre. C'est dans ce contexte, près d'un demi-siècle avant l'annonce par (Legendre,1805) du principe des moindres carrés et vingt ans avant la naissance de Gauss en 1777, que Roger Joseph Boscovich (1757) proposa une procédure pour déterminer les paramètres du modèle de régression linéaire simple

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1, \ldots, n$$

Pour obtenir la droite $y = \hat{\beta}_0 + \hat{\beta}_1 x$ décrivant au mieux les observations, il proposa le critère de l'estimation LAD :

$$\text{Min} \sum_{i=1}^{n} \left| y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right|$$

En imposant que la droite estimée passe par le centroïde des données $\left(\overline{x}; \overline{y}\right)$, en ajoutant la condition :

$$\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) = 0$$

Boscovich justifia cette approche de la manière suivante. Le critère de l'estimation LAD comme étant nécessaire pour que la solution soit aussi proche que possible des observations, et la condition supplémentaire pour que les erreurs positives et négatives soient de probabilité égales.

En effet cette condition signifie que la somme des erreurs positives et négatives doit être la même. De plus, elle peut se mettre sous la forme :

$$\overline{y} - \hat{\beta}_0 - \hat{\beta}_1 \overline{x} = 0$$

(1)

d'où

$$\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}$$

(2)

Et le problème se réduit alors à minimiser :

$$\sum_{i=1}^{n} \left| \left(y_i - \overline{y} \right) - \widehat{\beta}_1 \left(x_i - \overline{x} \right) \right|$$

Par conséquent, la détermination de la "droite de Boscovich" satisfaisant les deux critères revient à déterminer la pente $\widehat{\beta}_1$ de l'équation (2), puis à évaluer l'ordonnée à l'origine $\widehat{\beta}_0$ par l'équation (1). Ce n'est cependant que trois ans plus tard, en 1760, que Boscovich donna une procédure géométrique permettant de résoudre l'équation (2). Cette procédure est décrite en détails dans un article d'Eisenhart (1961).

Sept ans avant de s'intéresser aux estimateurs LAD , Laplace (1786) proposa une procédure permettant d'estimer les paramètres d'un modèle de régression linéaire simple en se basant sur le critère L_∞ .En d'autres termes, il proposa une solution pour trouver $\left(\widehat{\beta}_0 , \widehat{\beta}_1 \right)$ qui minimise :

$$\max_{1 \le i \le n} \left| y_i - \widehat{\beta}_0 - \widehat{\beta_1} x_i \right| = \max_{1 \le i \le n} \left| \widehat{e}_i \right|$$

Dans une publication ultérieure, Laplace (1793) proposa une procédure qu'il qualifie lui-même de plus simple. Cette procédure, basée sur les deux critères qu'avait proposé Boscovich en 1757, a l'avantage d'être analytique alors que celle proposée par Boscovich était géométrique.

L'intérêt de cette procédure analytique réside dans la facilité à obtenir les paramètres $\widehat{\beta}_0 \; et \; \widehat{\beta}_1$ lorsque le nombre d'observations augmente. Cette solution analytique de Laplace est élégante et mérite d'être rappelée ici. En adoptant les notations suivantes

$$Y_i = y_i - \overline{y} \; et \; X_i = x_i - \overline{x}$$

Le problème revient à trouver la valeur de ß qui minimise la fonction

$$f\ (\beta) = \sum_{i=1}^{n} \left| y_i - \beta X_i \right|$$

(3)

Notons que les valeurs X_i peuvent être supposées non nulles $\left(X_I \neq 0 \right)$ puisque $f\ (\beta) = \sum \left| y_i \right| + \sum \left| y_i - \beta X_i \right|$, la première somme étant prise pour

les *Xi* nuls et la seconde somme pour les X_i non nuls. Le minimum de la fonction f étant atteint pour la même valeur de β que celle rendant la seconde somme minimale. La fonction (3) peut s'écrire :

$$f(\beta) = \sum_{i=1}^{n} f_i(\beta) \text{ avec } f_i(\beta) = \left| Y_i - \beta X_i \right|$$

Chaque fonction $f_i(\beta)$ est continue, linéaire par morceaux et convexe.

Elle est formée de deux droites avec un minimum en $\left(\frac{y_i}{x_i}; 0\right)$. Sa pente est donnée par

$$f(\beta) = \begin{cases} -\left|X_i\right| & \text{si } \beta < \frac{y_i}{x_i} \\ +\left|X_i\right| & \text{si } \beta > \frac{y_i}{x_i} \end{cases}$$

Pour étudier la pente de $f(\beta)$, il s'agit d'ordonner en ordre croissant les rapports $\frac{Y_i}{X_i}$ de manière à ce que :

$$\frac{Y_1}{X_1} \leq \frac{Y_2}{X_2} \leq \ldots\ldots\ldots \leq \frac{Y_n}{X_n}$$

Ceci peut toujours être fait en renumérotant les observations. Ces rapports $\frac{Y_k}{X_k}$ seront désignés par $\beta_{(k)}, k = 1, \ldots n$ Pour $\beta < \beta_{(1)}$, chacune des fonctions $f_i(\beta)$ a une pente de $-\left|X_i\right|$ et par conséquent la pente de la fonction (I.3) est donnée par :

$$f'(\beta) = -\sum_{i=1}^{n} \left|X_i\right|$$

En chaque point $\frac{Y_k}{X_k}$ la pente de *f (ß)* augmente de $2\left|X_k\right|, k = 1, \ldots\ldots, n$.

f (ß) étant continue, linéaire par morceaux et convexe, elle atteint son minimum lorsque sa pente change de signe, c'est-à-dire pour $\beta_{(r)}$ tel que :

$$-\sum_{i=1}^{n}\left|X_i\right| + 2\sum_{i=1}^{r-1}\left|X_K\right| < 0 \quad \text{et} \quad -\sum_{i=1}^{n}\left|X_i\right| + 2\sum_{k=1}^{r}\left|X_K\right| \geq 0$$

Dans le cas où : $-\sum_{i=1}^{n}\left|X_i\right| + 2\sum_{k=1}^{r}\left|X\right| = 0$

La solution n'est *pas unique;* dans ce cas, pour toute valeur $\widehat{\beta}$ telle que :

$$\beta_{(r)} \leq \widehat{\beta} \leq \left(\beta_{(r+1)}\right), f(\beta)x \qquad \text{soit minimale.}$$

Notons encore que $\widehat{\beta} = \frac{Y_r}{X_r}$ est appelée ***médiane pondérée*** des $\frac{Y_i}{X_i}$, avec poids $\left|X_i\right|$**.** Ainsi, dans le cas où la droite LAD doit satisfaire le second critère de Boscovich, elle passe par le centroïde des données et par l'une des observations au moins.

C'est à Gauss (1809) que l'on doit une étape importante de la caractérisation des estimateurs LAD. Contrairement à Boscovich, il étudia la méthode consistant à minimiser la somme des erreurs en valeur absolue sans la restriction que leur somme soit nulle (appliquant uniquement le critère 1). A cette époque, Gauss ne semblait d'ailleurs pas savoir que cette restriction avait été introduite par Boscovich, puisqu'il l'attribue à Laplace. D'autre part, Gauss s'intéressa à l'estimation LAD dans un modèle de régression linéaire multiple, en cherchant le vecteur de paramètres $\left(\widehat{\beta}_1,....,\widehat{\beta}_p\right)$ qui minimise :

$$\sum_{i=1}^{n}\left|y_i - x_{i1}\widehat{\beta}_1 - ....x_{ip}\widehat{\beta}_p\right| = \sum_{i=1}^{n}\left|\widehat{e}_i\right|$$

Il mentionna que cette méthode fournit nécessairement *p* résidus nuls et qu'elle n'utilise les autres $(n-p)$ résidus que dans la détermination du choix des ***p*** résidus nuls. De plus, il mentionne que la solution obtenue par cette méthode n'est pas modifiée si la valeur des y_i est augmentée ou diminuée sans que les résidus changent de signe. Gauss remarqua également que la méthode consistant à minimiser

$\sum_{i=1}^{n}\left|\widehat{e}_i\right|$ avec la restriction que $\sum_{i=1}^{n}\left|\widehat{e}_i\right| = 0$ fournit nécessairement *(p* - 1) résidus nuls. Dans le cas de la régression linéaire simple *(p* = 2) traitée par Laplace avec la restriction que la somme des résidus soit nulle, on obtient effectivement une droite passant par l'une des observations, c'est-à-dire qu'un des résidus est nul.

Bloomfield et Steiger (1983) prouvent ce résultat et indiquent qu'il pourrait bien être l'un des premiers en programmation linéaire, mais pas assez profond pour que Gauss le démontre.

Avec l'annonce par Legendre (1805) de la méthode des moindres carrés et son développement par Gauss (1809, 1823, 1828) et Laplace (1812) basé sur la théorie des probabilités, la méthode d'estimation LAD joua un rôle secondaire durant la plus grande partie du *XIX*^{ème} siècle. Ce n'est qu'en 1887 que cette méthode refait surface grâce au travail d'Edgeworth.

En effet, Edgeworth supprima la restriction faite par Boscovich que la somme des résidus soit nulle. Il présenta d'un point de vue géométrique une procédure générale décrite ici dans le cas de la régression linéaire simple *(p = 2).*

Les n observations sont notées $p_1\left(x_1; y_1\right), \ldots, P_n\left(x_n; y_n\right)$.En posant $m_0 = 1$ et en traitant Pm_0 comme l'origine (en soustrayant $\boldsymbol{P_i}$ des autres observations), la procédure de Laplace peut s'appliquer. Or la droite ainsi forcée de passer par Pm_0 contiendra l'une des autres observations, disons Pm_1. Traitant cette nouvelle observation comme l'origine, on trouve une droite passant par une autre observation Pm_2 et ainsi de suite. Cette procédure ne requiert pas plus de $\boldsymbol{r = n - 1}$ étapes, chacune représentant le calcul d'une médiane pondérée, puisqu'elle se termine lorsque $Pm_r = Pm_{r-2}$

Lorsque $p > 2$, l'algorithme, bien que plus compliqué, est analogue et revient à fixer $(p - 1)$ des paramètres à estimer puis à utiliser la procédure de Laplace pour déterminer la valeur optimale du paramètre restant.

Notons finalement que l'algorithme décrit ci-dessus dans le cas de la régression linéaire simple présente certains défauts. Par exemple, il se peut que sur l'une des droites obtenues, il y ait trois observations (d'indice $\boldsymbol{i_1, i_2}$ *et* $\boldsymbol{i_3}$) conduisant la procédure à faire un cycle de la façon suivante $i_1 \rightarrow i_2 \rightarrow i_3 \rightarrow i_1 \ldots\ldots$ sans conduire pour autant à diminuer la somme des résidus en valeur absolue comme l'indique Sposito (1976). Karst (1958) mentionne que cette procédure peut s'arrêter prématurément. C'est le cas lorsque par exemple la droite forcée à passer par p_{i1} contient l'observation p_{i2} et vice-versa, mais n'est pas optimale.

Une implantation en langage Fortran de cet algorithme a été faite par Sadovski (1974) permettant d'éviter les problèmes décrits ci-dessus.

Rhodes (1930) trouva la solution graphique d'Edgeworth difficile à appliquer en pratique. Il proposa alors une procédure itérative que l'on peut résumer ainsi :

(i) Choisir arbitrairement $p - 1$ équations.

(ii) Les utiliser pour éliminer les $p - 1$ premiers paramètres du problème.

(iii) Utiliser la procédure de Laplace pour estimer le paramètre restant.

(iv) Associer l'équation correspondant au point 3 à l'ensemble des $p - 1$ équations.

(v) Si l'ensemble des p équations se répète p fois, on s'arrête. Sinon, on élimine l'équation la plus ancienne et l'on retourne au point 2.

Cette procédure reste cependant difficile à utiliser dans des problèmes pratiques compte tenu des moyens de l'époque.

C'est grâce au travail de Charnes, Cooper et Fergusson (1955) que l'intérêt porté aux estimateurs LAD a été le plus stimulé. Comme alternative à la méthode des moindres carrés, ils proposèrent l'utilisation de la programmation linéaire pour calculer les estimateurs LAD.

Charnes, Cooper et Fergusson (1955) ont montré que le problème de régression linéaire multiple basé sur la norme LAD peut se mettre sous la forme d'un problème de programmation linéaire; pour cela, ils considèrent les résidus comme la différence de deux variables non négatives.

En posant $e_i = u_i - v_i$ où $u_i ; v_i \geq 0$ représentent les déviations positives et négatives respectivement, le problème devient :

$$\text{minimiser} \sum_{i=1}^{n} \left(u_i + v_i \right)$$

$$s.c \sum_{i=1}^{p} x_{ij} + u_i - v_i = y_i$$

$$et \; u_i , v_i \geq 0, \quad i = 1, \ldots\ldots, n$$

Où $\widehat{\beta}_1, \ldots\ldots, \widehat{\beta}_p$ sont sans restriction de signe. Ainsi, le problème de régression linéaire multiple basé sur la norme LAD peut être formulé comme un problème de programmation linéaire avec $2n + p$ variables et n contraintes.

Cette formulation du problème correspond à la forme primale et a pu être résolu en utilisant la méthode du simplexe. Cependant, il a rapidement été reconnu que la structure de ce type de problème pouvait être prise en compte pour améliorer la performance des algorithmes. En effet, Wagner (1959) proposa une formulation de ce problème basée sur le dual, ce qui permit à Barrodale et Young (1966) de mettre au point un algorithme relativement rapide. Par la suite, de nombreuses publications furent consacrées à l'élaboration d'algorithmes de plus en plus performants.

III.2- DECOUVERTE DE L'ESTIMATION LS *:*

La découverte de l'estimation LS *(*méthode des moindres carrés) mérite d'être rappelée ici puisqu'elle fut à l'origine de l'une des plus grandes disputes dans l'histoire de la statistique. Adrien Marie Legendre (1805) publia le premier la méthode des moindres carrés. Il donna une explication claire de la méthode en donnant les équations normales et en fournissant un exemple numérique.

Selon Stigler (1981), Robert Adrain, un américain, publia la méthode vers la fin de l'année 1808 ou au début de l'année 1809. Selon Stigler (1977, 1978), il se pourrait que Robert Adrain

ait "découvert" cette méthode dans l'ouvrage de Legendre (1805). Cependant, quatre ans après la publication de Legendre, Gauss (1809) a le courage de réclamer la paternité de la méthode des moindres carrés, en prétendant l'avoir utilisée depuis 1795. La revendication de Gauss déclencha l'une des plus grandes disputes scientifiques dont les détails sont présentés et résumés dans un article de Plackett (1972). Bien que le doute subsiste, plusieurs faits troublants semblent indiquer que Gauss a effectivement utilisé la méthode des moindres carrés avant 1805. En particulier, Gauss prétend qu'il a parlé de cette méthode à certains astronomes (Olbers, Lindenau et von Zach) avant 1805. De plus, dans une lettre de Gauss datant de 1799, il est fait mention de "ma méthode", sans qu'un nom y soit donné. Il semble difficile de ne pas le croire, vu l'extraordinaire compétence reconnue à Gauss comme mathématicien.

Il reste cependant une question très importante : quelle importance attachait Gauss à cette découverte ? La réponse pourrait être que Gauss, bien que jugeant cette méthode utile, n'a pas réussi à communiquer son importance à ses contemporains avant 1809. En effet, dans sa publication de 1809, Gauss est allé bien plus loin que Legendre dans ses développements autant conceptuels que techniques. C'est dans cet article qu'il lie la méthode des moindres carrés à la loi normale (Gaussienne) des erreurs. Il propose également un algorithme pour le calcul des estimateurs. Son travail a d'ailleurs été discuté par plusieurs auteurs comme Seal (1967), Eisenhart (1968), Goldstine (1977), Sprott (1978) et Sheynin (1979).

Gauss a certainement été le plus grand mathématicien de cette époque, mais c'est Legendre qui a cristallisé l'idée de la méthode des moindres carrés sous une forme compréhensible par ses contemporains.

IV.1- INTRODUCTION :

La méthode des moindres écarts en valeurs absolues a plusieurs notations, comme LAD (Least absolute deviations), MAD (Minimum absolute deviations), LAR (Least absolute residuals), la norme L_1 et LAV (Least absolute values), dans ce travail nous utiliserons la notation LAD.

IV.2- METHODES D'ESTIMATIONS :

Il existe plusieurs méthodes d'estimation des paramètres; les méthodes considérées par la suite, concernent principalement l'estimation LAD, et l'estimation LS.

IV.2.1- L'ESTIMATION LAD :

La méthode des moindres écarts en valeurs absolues, dite méthode LAD**.** (En anglais : least absolute déviations), est l'une des principales alternatives à la méthode des moindres carrés lorsqu'il s'agit d'estimer les paramètre d'un modèle de régression. Elle a été introduite presque cinquante ans avant la méthode des moindres carrés, en 1757 par Roger Joseph Boscovich. Il utilisa cette procédure dans le but de concilier des mesures incohérentes dans le cadre de l'estimation de la forme de la terre. Pierre Simon Laplace adopta cette méthode trente ans plus tard, mais elle fut ensuite éclipsée par la méthode des moindres carrés développées par Legendre et Gauss, la popularité de la méthode des moindres carrés repose principalement sur la simplicité des calculs Mais aujourd'hui, avec les progrès de l'informatique, la méthode LAD, peut être utilisée presque aussi simplement.

IV.2.1.1- LE MODELE DE REGRESSION LINEAIRE SIMPLE :

Le modèle de régression simple est donné par :

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (4)$$

A partir des n observations (x_i, y_i), il s'agit d'estimer les paramètres β_0 et β_1 du modèle par $\widehat{\beta_0}$ et $\widehat{\beta_1}$. On obtient ainsi des valeurs estimées :

$$\widehat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

qui ne doivent pas être trop éloignées de y_i, du moins si le modèle est correct. Cela signifie que les estimateurs $\widehat{\beta_0}$ et $\widehat{\beta_1}$ doivent être choisis de telle sorte que les résidus du modèle :

$$e_i = y_i - \widehat{y}_i$$

soient petits .Le critère utilisé par la méthode des moindres carrés est de minimiser la somme des carrés des résidus :

$$\Sigma e_i^2 = \Sigma \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right)^2$$

(5)

Le critère utilisé par la méthode LAD est de minimiser la somme des valeurs absolues des résidus :

$$\Sigma \left| e_i \right| = \Sigma \left| y_i - \widehat{\beta_0} - \widehat{\beta_1} x_i \right|$$

(6) Il s'agit donc ici de choisir $\hat{\beta}_0$ et $\hat{\beta}_1$ de telle sorte que (6) soit minimale .Dans un certain sens, il peut paraître plus naturel de vouloir minimiser (6) plutôt que (5). Pourtant, le calcul des estimateurs LAD, est plus complexe. En particulier, on ne dispose pas de formule explicite pour calculer $\hat{\beta}_0$ et $\hat{\beta}_1$ comme c'est le cas avec la méthode des moindres carrés, puisque la fonction valeur absolue n'est pas dérivable. Les estimateurs LAD, peuvent toutefois être calculés par des algorithmes itératifs.

IV.2.1.2- TEST D'HYPOTHESE SUR LA PENTE β_1:

On va voir dans cette section comment tester l'hypothèse nulle :

$$H_0 : \beta_1 = 0$$

contre l'hypothèse alternative :

$$H_1 : \beta_1 \neq 0$$

en utilisant la régression LAD.

Il s'agit tout d'abord d'estimer les paramètres du modèle β_0 et β_1 par les estimateurs LAD $\hat{\beta}_0$ et $\hat{\beta}_1$. On calcule alors les valeurs estimées LAD.

$$\widehat{y}_i = \beta_0 + \beta_1 x_i$$

Et les résidus LAD :

$$e_i = y_i - \widehat{y}_i$$

Comme la droite LAD passe par deux points, on a 2 résidus nuls (si l'on n'est pas dans un cas de dégénérescence). On classe les $m = n - 2$ résidus non nuls par ordre croissant de telle

sorte que e_1 soit le plus petit résidus non nul, e_2 le second plus petit résidu non nul, et ainsi de suite.

Soit k_1, l'entier le plus proche de $(m+1)/2 - \sqrt{m}$ et soit k_2, l'entier le plus proche de $(m+1)/2 + \sqrt{m}$, on définit :

$$\hat{\tau} = \frac{\sqrt{m}\left(e_{k_2} - e_{k_1}\right)}{4}$$

(7) L'écart type de l'estimateur LAD $\hat{\beta}_1$ est alors estimé par :

$$s(\hat{\beta}_1) = \frac{\hat{\tau}}{\sqrt{\sum (x_i - \overline{x})^2}}$$

(8)

Et la statistique pour tester H_0 est donnée par :

$$t_{LAD} = \frac{|\hat{\beta}_1|}{s(\hat{\beta}_1)}$$

(9)

On rejette H_0 à un seuil de signification α si cette statistique t_{LAD} est plus grande que la valeur critique $t_{\alpha/2,n-2}$ que l'on trouve dans une table de Student.

On remarque que cette procédure est très similaire à celle utilisée par la méthode des moindres carrés, où rappelons-le, l'écart type de l'estimateur de β_1 était estimé par :

$$s(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\sum (x_i - \overline{x})^2}}$$

(10) Il s'agit donc de la même formule que (8) sauf que le $\hat{\tau}$ défini par (7) est remplacé dans (10) par l'estimateur habituel $\hat{\sigma}$ de l'écart type σ des erreurs ε_i du modèle le rapport entre les écarts type de l'estimateur LAD et celui des moindres carrés est donc égal à τ/σ.

IV.2.2- L'ESTIMATION LS :

IV.2.2.1- LE MODELE DE REGRESSION LINEAIRE SIMPLE :

Considérons un échantillon de n observations paires $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$. Nous avons vu que le modèle de régression linéaire, appelé ici *modèle de régression simple*, suppose, pour tout $i = 1, ..., n$, la relation:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Où les ε_i sont des quantités aléatoires inobservables, que nous appellerons dorénavant les *erreurs.* Ainsi les y_i sont des variables aléatoires (car elles dépendent des ε_i), alors que les x_i sont considérés comme des nombres fixés. En supposant l'espérance des ε_i nulle, on a ainsi:

$$\begin{aligned} E(y_i) &= \beta_0 + \beta_1 x_i + E(\varepsilon_i) \\ &= \beta_0 + \beta_1 x_i \\ Var(y_i) &= Var(\beta_0 + \beta_1 x_i + \varepsilon_i) = Var(\varepsilon_i) \end{aligned}$$

Afin de pouvoir faire de l'inférence sur la droite de régression :

$$\mu_y(x) = \beta_0 + \beta_1 x$$

à partir de la droite des moindres carrés :

$$\widehat{y}(x) = \widehat{\beta}_0 + \widehat{\beta}_1 x$$

on fait généralement les hypothèses supplémentaires suivantes sur ces variables aléatoires ε_i .

La variance de ε_i est égale à une quantité σ^2 (inconnue) ne dépendant pas de x_i. On a donc pour tout $i = 1, ..., n$:

$$Var(\varepsilon_i) = Var(y_i) = \sigma^2$$

- Les ε_i sont indépendantes.
- Les ε_i sont normalement distribuées.

Ces trois conditions reviennent à dire que les variables aléatoires ε_i sont indépendantes et identiquement distribuées (i.i.d.) selon une loi normale d'espérance nulle et de variance σ^2.
On note parfois:

$$\varepsilon_i \; i.i.d \; .\mathrm{N}(0, \sigma^2)$$

(11)

Remarquons que : la normalité des ε_i implique la normalité des y_i (un y_i s'obtenant d'un ε_i par une simple addition de la constante $(\beta_0 + \beta_1 x_i)$ de même que l'indépendance des ε_i implique l'indépendance des y_i. On a en effet, si $i \neq j$:

$$Cov(y_i, y_j) = Cov(\beta_0 + \beta_1 x_i + \varepsilon_i, \beta_0 + \beta_1 x_j + \varepsilon_j)$$
$$= Cov(\varepsilon_i, \varepsilon_j) = 0$$

Rappelons à ce sujet que l'indépendance entre deux variables dont la distribution conjointe est normale bivariée est équivalente à la nullité de leur covariance.

Lorsque l'on utilise un modèle de régression simple, on suppose donc que l'on a tiré un échantillon de ε_i distribués selon (11). Cependant, on n'observe pas ces ε_i **,** on observe à la place les variables aléatoires y_i définies par (4) à partir de ces ε_i, pour des valeurs x_i fixées par avance et de paramètres β_0 et β_1, inconnus.

IV.2.2.2- ESTIMATION DE LA VARIANCE DES ERREURS :

Cette quantité est toutefois inconnue et doit être estimée.

Si les erreurs ε_i pouvaient être observées, un estimateur non biaisé de σ^2 serait donné par la formule habituelle :

$$\frac{\sum (\varepsilon_i - \bar{\varepsilon})^2}{n - 1}$$

où $\bar{\varepsilon}$ serait la moyenne des ε_i . Or ces quantités ne sont pas observables. Mais nous avons vu que les erreurs ε_i peuvent être estimées par les résidus du modèle :

$$e_i = y_i - \hat{y}_i$$

où les :

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$
$$= \bar{y} + \hat{\beta}_1 (x_i - \bar{x})$$

sont les valeurs estimées des y_i. Ainsi, nous allons estimer σ^2 en utilisant la somme des carrés des résidus comme estimateur de la somme des carrés des erreurs :

$$\sum (e_i - \bar{e})^2 = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2$$

où $\bar{e}$ désigne la moyenne des e_i (on a vu que la somme, donc la moyenne, des résidus est nulle).

or, comme on a vu que

$$\sum (y_i - \widehat{y}_i)^2 = \sum y_i^2 - \sum \widehat{y}_i^2$$

on a :

$$\begin{aligned} E\left(\sum (y_i - \widehat{y}_i)^2\right) &= \sum E(y_i^2) - \sum E(\widehat{y}_i^2) \\ &= \sum \left(Var(y_i) + E^2(y_i)\right) - \sum \left(Var(\widehat{y}_i) + E^2(\widehat{y}_i)\right) \end{aligned}$$

or, on a :

$$\begin{aligned} \mathrm{E}(\mathrm{y}) &= \mathrm{E}\left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i\right) \\ &= E\left(\widehat{\beta}_0\right) + x_i E\left(\widehat{\beta}_1\right) \\ &= \beta_0 + x_i \beta_1 \\ &= E\left(y_i\right) \end{aligned}$$

on obtient ainsi :

$$\begin{aligned} E\left(\sum (y_i - \widehat{y}_i)^2\right) &= \sum \left(Var(y_i) - \sum Var(\widehat{y}_i)\right) \\ &= n\sigma^2 - \sum Var(\widehat{y}_i) \end{aligned}$$

d'autre part, on a :

$$\begin{aligned} Var(\widehat{y}_i) &= Var\left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i\right) \\ &= Var\left(\widehat{\beta}_0\right) + x\, Var\left(\widehat{\beta}_1\right) + 2x_i Cov\left(\widehat{\beta}_0, \widehat{\beta}_1\right) \\ &= \frac{\sigma^2 \sum x_j^2}{n\sum\left(x_j - \bar{x}\right)^2} + \frac{x_i^2 \sigma^2}{\sum\left(x_j - \bar{x}\right)^2} - \frac{2x_i \bar{x}\sigma^2}{\sum\left(x_j - \bar{x}\right)^2} \\ &= \frac{\sigma^2}{\sum\left(x_j - \bar{x}\right)^2}\left(\frac{\sum x_j^2}{n} + x_i^2 - 2x_i\bar{x}\right) \\ &= \frac{\sigma^2}{\sum\left(x_j - \bar{x}\right)^2}\left(\frac{\sum x_j^2}{n} - \frac{n\bar{x}^2}{n} + x_i^2 - 2x_i\bar{x} + \bar{x}^2\right) \\ &= \frac{\sigma^2}{\sum\left(x_j - \bar{x}\right)^2}\left(\frac{\sum\left(x_j - \bar{x}\right)^2}{n} + \left(x_i - \bar{x}\right)^2\right) \\ &= \sigma^2\left(\frac{1}{n} + \frac{\left(x_i - \bar{x}\right)^2}{\sum\left(x_j - \bar{x}\right)^2}\right) \end{aligned}$$

on obtient ainsi :

$$E\left(\sum\left(x_j-\overline{x}\right)^2\right)=N\,\sigma^2-\sigma^2\left(\frac{n}{n}+\frac{\sum\left(x_i-\overline{x}\right)^2}{\sum\left(x_j-\overline{x}\right)^2}\right)$$

$$=\sigma^2(n-2)$$

on peut ainsi définir un estimateur sans biais de σ^2 en posant :

$$s^2=\frac{\sum(y_i-\widehat{y}_i)^2}{n-2}=\frac{SC_{res}}{n-2}=\frac{SC_{tot}-SC_{reg}}{n-2}=\frac{\sum(y_i-\widehat{y}_i)^2-\widehat{\beta}_1{}^2\sum\left(x_i-\overline{x}\right)^2}{n-2}$$

on estime par ailleurs l'écart type des erreurs par :

$$s=\sqrt{\frac{\sum(y_i-\widehat{y}_i)^2}{n-2}}\,.$$

IV.2.2.3- TEST SUR LA PENTE :

L'estimateur $\widehat{\beta}_1$ est normalement distribué, d'espérance β_1 et de variance que l'on note ici par :

$$\sigma^2\left(\widehat{\beta}_1\right)=\frac{\sigma}{\sum\left(x_i-\overline{x}\right)}$$

il s'ensuit que la quantité :

$$\frac{\widehat{\beta}_1-\beta_1}{\sigma\left(\widehat{\beta}_1\right)}$$

où $\sigma\left(\widehat{\beta}_1\right)$ désigne la racine carrée de $\sigma^2\left(\widehat{\beta}_1\right)$, suit une loi normale standardisée. Or cette quantité ne peut pas être utilisée pour un problème de test d'hypothèses puisque on ne connaît pas la valeur de σ. En pratique, on estime $\sigma^2\left(\widehat{\beta}_1\right)$ par :

$$s^2\left(\widehat{\beta}_1\right)=\frac{s^2}{\sum\left(x_i-\overline{x}\right)^2}$$

où s^2 est l'estimateur sans biais de σ^2 défini ci-dessus. On peut alors montrer que la quantité :

$$\frac{\widehat{\beta}_1-\beta_1}{s\left(\widehat{\beta}_1\right)}$$

où $s\left(\widehat{\beta}_1\right)$ désigne la racine carrée de $s^2\left(\widehat{\beta}_1\right)$, suit une loi de Student avec $(n-2)$ degrés de liberté.

Il s'ensuit que dans un problème de test d'hypothèses bilatéral où l'on désire tester l'hypothèses nulle :

$$H_0 : \beta_1 = b_1$$

contre l'hypothèse alternative :

$$H_t : \beta_1 \neq b_1$$

on peut utiliser la statistique :

$$t_c = \frac{\hat{\beta}_1 - b_1}{s\left(\hat{\beta}_1\right)}$$

on rejette au seuil de signification α si :

$$| t_c | \rangle t_{(\alpha/2\,n-2)}$$

où la valeur critique $t_{(\alpha/2n-2)}$ est le $(1-\alpha/2)$ quantile d'une loi de Student avec $(n-2)$ degrés de liberté que l'on trouve dans une table de Student.

Un test d'hypothèse particulièrement intéressant est le test de l'hypothèse nulle :

$$H_0 : \beta_1 = 0$$

contre l'hypothèse alternative :

$$H_1 : \beta_1 \neq 0$$

En effet, le non-rejet de l'hypothèse nulle implique un modèle avec un seul paramètre :

$$y_i = \beta_0 + \varepsilon_i$$

par contre si cette hypothèse H_0 est rejetée, c'est-à-dire si :

$$| t_c | = \frac{\hat{\beta}_1 - b_1}{s\left(\hat{\beta}_1\right)} \rangle t_{(\alpha/2n-2)}$$

on dit que la relation entre les x_i et les y_i est significative au seuil de signification α.

IV.2.2.4- INTERVALLE DE CONFIANCE :

Un intervalle de confiance au niveau $(\alpha-1)$ pour un paramètre β_j est défini par :

$$\left[\hat{\beta}_j - t_{(\alpha/2,n-p)}.s\left(\hat{\beta}_j\right); \hat{\beta}_j + t_{(\alpha/2,n-p)}.s\left(\hat{\beta}_j\right)\right]$$

c'est -à- dire par :

$$\hat{\beta}_j \pm t_{(\alpha/2,n-p)}.s\left(\hat{\beta}_j\right)$$

cet intervalle est construit de telle sorte qu'il contienne le paramètre inconnu β_j avec une probabilité de $(\alpha-1)$.

IV.2.2.5- COEFFICIENT DE CORRELATION :

Le coefficient de corrélation est défini comme suite :

$$r_{xy} = \frac{\sum\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sqrt{\left(x_i - \overline{x}\right)^2}\sqrt{\left(y_i - \overline{y}\right)^2}} = \frac{\sum\left(\widehat{y}_i - \overline{y}\right)\left(y_i - \overline{y}\right)}{\sqrt{\left(\widehat{y}_i - \overline{y}\right)^2}\sqrt{\left(y_i - \overline{y}\right)^2}} = r_{\widehat{y}y}$$

où :

$$\overline{x} = \sum x_i \Big/ n$$

$$\overline{y} = \sum y_i \Big/ n$$

IV.2.2.6- LIEN ENTRE LE COEFFICIENT DE CORRELATION ET LE COEFFICIENT DE DETERMINATION :

Il faut mettre en évidence la relation fondamentale qui existe entre le coefficient de corrélation et le coefficient de détermination R^2, qui donne le pourcentage de la variance totale des Y_i expliquée par le modèle de régression simple et le coefficient de corrélation r_{xy}. Rappelons à ce propos que l'on avait :

$$R^2 = \frac{\left(\widehat{y}_i - \overline{y}\right)^2}{\left(y_i - \overline{y}\right)^2}$$

avec :

$$\begin{aligned}\widehat{y}i &= \widehat{\beta}_0 + \widehat{\beta}_1 x_i \\ &= \overline{y} + \widehat{\beta}_1\left(x_i - \overline{x}\right)\end{aligned}$$

ce qui donne :

$$R^2 = \frac{\left(\widehat{y}_i + \widehat{\beta}_1\left(x_i - \overline{x}\right) - \overline{y}\right)^2}{\left(y_i - \overline{y}\right)^2}$$

$$= \widehat{\beta}_1^2 \frac{\sum\left(x_i - \overline{x}\right)^2}{\sum\left(y_i - \overline{y}\right)^2}$$

En reprenant les notions introduites ci-dessus, on a ainsi :

$$R^2 = \widehat{\beta}_1^2 \frac{s_x^2}{s_y^2}$$

et donc :

$$R^2 = r_{xy}^2$$

IV.3- COMPARAISON DE DEUX DROITES DE REGRESSION :

Soit $Y = a + bx = \overline{y} + b(x - \overline{x})$ l'équation d'une droite de régression, les deux droites comparées seront distinguées par les indices Y_{lad} et Y_{LS}, il faut calculer les coefficients $\overline{y}$ et b, ainsi que les variances résiduelles S. Nous nous plaçons dans le cas où l'un des nombres de couples de résultats N_I ou N_{II} est inférieur à 30.

IV.4- COMPARAISON DES ORDONNEES DE DEUX DROITES AU POINT MOYEN :

Choisir une valeur de x_0 appartenant au domaine de toutes les droites et aussi prés que possible du point moyen. Calculer la valeur de Y correspondant sur chaque droite à l'abscisse x_0 et comparer les valeurs de Y de la façon indiquée ci-*après* ; on calcule une estimation S de la variance résiduelle commune aux deux droites en faisant une moyenne pondérée de $S^2_{2(I)}$ et $S^2_{2(II)}$ suivant le nombre de degrés de liberté :

$$S_2^2 = \frac{(N_I - 2)S^2_{2(I)} + (N_{II} - 2)S^2_{2(II)}}{N_I + N_{II} + 4}$$

(12)

Cette estimation est utilisée pour calculer S^2_{YI} et S^2_{YII} par la formule :

$$S^2_{YI} = S_2^2 \left[\frac{1}{N_I} + \frac{(x_0 - \overline{x}_I)}{\sum (x_{iI} - \overline{x}_I)^2} \right]$$

(13)

La formule ci-dessous permet d'en déduire une valeur t**.**

$$t = \frac{|Y_I - Y_{II}|}{\sqrt{S^2_{Y_I} - S^2_{Y_{II}}}}$$

(14)

Cette variable suit la loi de Student à (NI +NII - 4) degrés de liberté dans le cas où les deux droites de régression vraies sont confondues.

La valeur expérimentale t est donc comparée à la limite $t_{1-\alpha/2}$ donnée par la table de Student.

Si la valeur de t est supérieure à la limite donnée par la table, on peut admettre, au niveau de confiance choisi, que les deux droites se déplacent parallèlement l'une par rapport à l'autre.

IV.5- COMPARAISON DES PENTES DES DEUX DROITES :

Utiliser l'estimation commune (12) de la variance résiduelle pour calculer

$$S^2_{b(I)} = \frac{S^2_{2(I)}}{\sum (x_{iI} - \overline{x}_I)^2}$$

(15)

En déduire t par la formule :

$$t = \frac{|b_I - b_{II}|}{\sqrt{S^2_{b_I} - S^2_{b_{II}}}}$$

(16)

Cette variable suit la loi de Student à (NI+NII - 4) degrés de liberté dans le cas où les deux droites de régression vraies sont confondues. Comparer la valeur de **t** à la limite $t_{1-\alpha/2}$ donnée par la table de Student. Si le test est significatif, on peut conclure qu'il y a rotation d'une droite par rapport à l'autre.

IV.6- COMPARAISON DES VARIANCES RESIDUELLES :

Comparer les valeurs S^2_2 par le test de Snedecor en formant le rapport :

$$F = \frac{S^2_{2I}}{S^2_{2II}}$$

Comparer ce rapport expérimental aux limites données par la table de Snedecor pour $Y_I = (N_I - 2)$ et $Y_{II} = (N_{II} - 2)$ soit, au niveau de confiance $(1-\alpha)$

$$F_{1-\alpha/2}(N_I, N_{II}) \quad \text{et} \quad F_{\alpha/2} = \frac{1}{F_{1-\alpha/2}(N_I, N_{II})}$$

L'hypothèse contrôlée : $\sigma^2_I = \sigma^2_{II}$

sera refusée si : $F > F_{1-\alpha/2}$

ou si : $F < F_{1-\alpha}$

ne sera pas refusée si : $F_\alpha < F < F_{1-\alpha/2}$.

V.1.1- Droites de régression LS et LAD_1 :

Après l'entrée des données dans l'application programmée en langage Pascal et exécutable sous Windows, nous avons calculé les paramètres du modèle LAD_1 donnés par **l'algorithme des moindres carrés re-pondérés itérativement**, le R^2 a été calculé via Excel selon la méthode décrite au chapitreIV, et les paramètres du modèle LS ont été calculés via Minitab, le tableau 2 reproduit les paramètres des modèles LS et LAD_1 :

Tableau 2 : Paramètres des modèles LS et LAD_1.

	m	**b**	**R^2**	**Variance**	$\sum\|e_i\|$
LS	0.834	- 2.070	95.20%	σ^2_{LS} =0.105	6.98682
LAD_1	0.841	- 2.149	97.54%	τ^2_{LAD} =0.089	6.55319

La figure 4 reproduit les représentations graphiques des deux droites de régression LS et LAD_1 réalisées sous Excel :

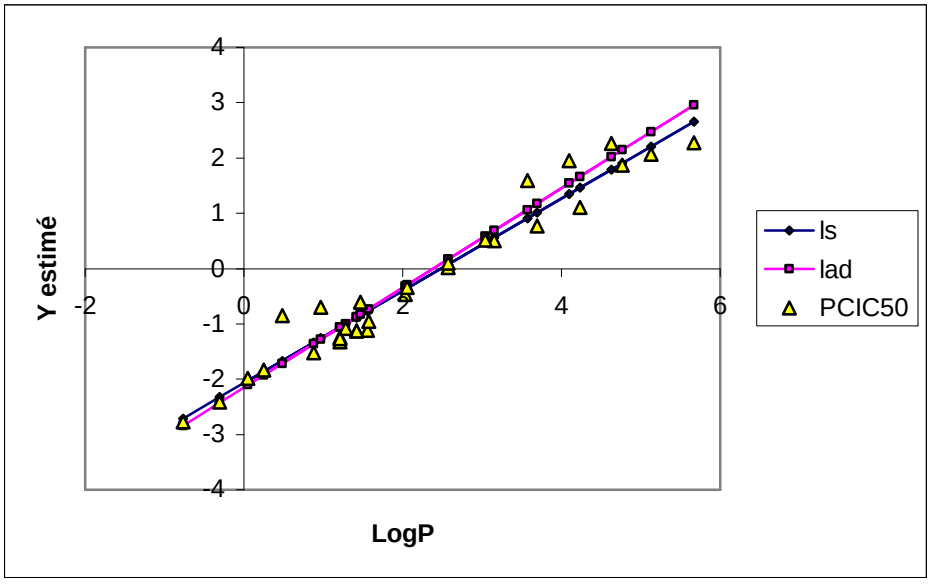

Figure 4 : Droites de régression LS, et LAD_1.

Puis nous avons calculé par les deux méthodes LS et LAD_1, les valeurs estimées de la toxicité pCIC50 et les erreurs résiduelles correspondant à chaque coefficient de partage logP des composés. Le tableau 3 reproduit la toxicité estimée et l'erreur résiduelle obtenues par régressions LS et LAD_1 pour chaque composé, la figure 5 reproduit l'histogramme des erreurs résiduelles.

Tableau 3 : Erreurs résiduelles pour la LS et la LAD_1.

N°	Composé	pCIC50	logP	$\hat{Y}_{LS}$	$\hat{Y}_{LAD1}$	$e_{i(LS)}$	$e_{i(LAD1)}$
1	Méthanol	-2.77	-0.77	-2.71018	-2.79657	-0.05982	0.02657
2	Ethanol	-2.41	-0.31	-2.32654	-2.40971	-0.08346	-0.00029
3	Propan-1-ol	-1.84	0.25	-1.8595	-1.93875	0.0195	0.09875
4	Butan-1-ol	-1.52	0.88	-1.33408	-1.40892	-0.18592	-0.11108
5	Pentan-1-ol	-1.12	1.56	-0.76696	-0.83704	-0.35304	-0.28296
6	Hexan-1-ol	-0.47	2.03	-0.37498	-0.44177	-0.09502	-0.02823
7	Heptan-1-ol	0.02	2.57	0.07538	0.01237	-0.05538	0.00763
8	Octan-1-ol	0.5	3.15	0.5591	0.50015	-0.0591	-0.00015
9	Nonan-1-ol	0.77	3.69	1.00946	0.95429	-0.23946	-0.18429
10	Decan-1-ol	1.1	4.23	1.45982	1.40843	-0.35982	-0.30843
11	Undecan-1-ol	1.87	4.77	1.91018	1.86257	-0.04018	0.00743
12	Dodecan-1-ol	2.07	5.13	2.21042	2.16533	-0.14042	-0.09533
13	Tridecan-1-ol	2.28	5.67	2.66078	2.61947	-0.38078	-0.33947
14	Propan-2-ol	-1.99	0.05	-2.0263	-2.10695	0.0363	0.11695
15	Pentan-2-ol	-1.25	1.21	-1.05886	-1.13139	-0.19114	-0.11861
16	Pentan-3-ol	-1.33	1.21	-1.05886	-1.13139	-0.27114	-0.19861
17	2-methylbutan-1-ol	-1.13	1.42	-0.88372	-0.95478	-0.24628	-0.17522
18	3-methylbutan-1-ol	-1.13	1.42	-0.88372	-0.95478	-0.24628	-0.17522
19	3-methylbutan-2-ol	-1.08	1.28	-1.00048	-1.07252	-0.07952	-0.00748
20	(ter) pentanol	-1.27	1.21	-1.05886	-1.13139	-0.21114	-0.13861
21	(neo) pentanol	-0.96	1.57	-0.75862	-0.82863	-0.20138	-0.13137
22	1-propylamine	-0.85	0.48	-1.66768	-1.74532	0.81768	0.89532
23	1-butylamine	-0.7	0.97	-1.25902	-1.33323	0.55902	0.63323
24	1-amylamine	-0.61	1.47	-0.84202	-0.91273	0.23202	0.30273
25	1-hexylamine	-0.34	2.06	-0.34996	-0.41654	0.00996	0.07654
26	1-heptylamine	0.1	2.57	0.07538	0.01237	0.02462	0.08763
27	1-octylamine	0.51	3.04	0.46736	0.40764	0.04264	0.10236
28	1-nonylamine	1.59	3.57	0.90938	0.85337	0.68062	0.73663
29	1-decylamine	1.95	4.1	1.3514	1.2991	0.5986	0.6509
30	1-undecylamine	2.26	4.63	1.79342	1.74483	0.46658	0.51517
					$\sum\|e_i\|$	**6.98682**	**6.55319**

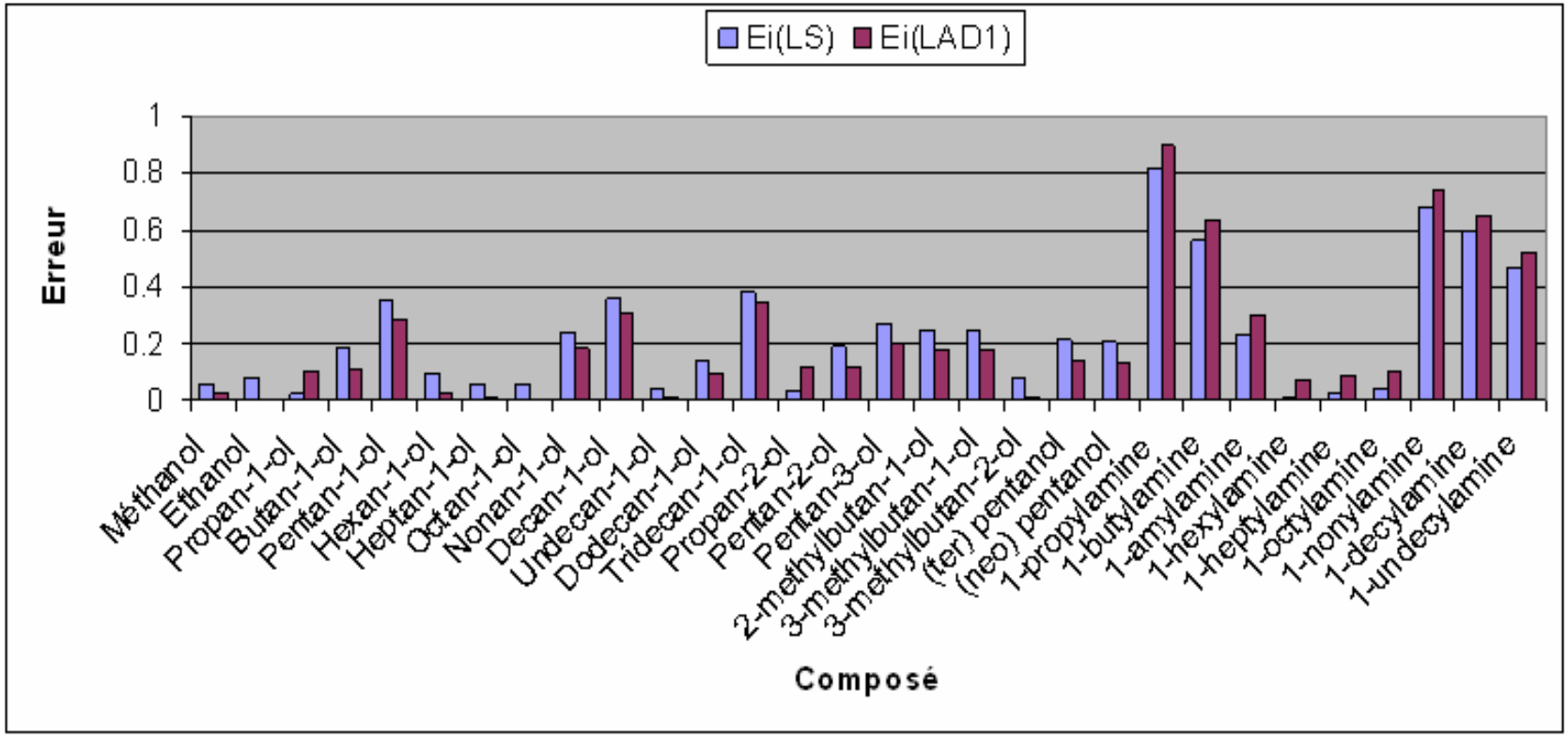

Figure 5 : Histogramme des erreurs résiduelles pour la LS et la LAD_1.

LS: $\hat{Y}_{LS}$= 0.834 logP-2.070 R^2=95.20% σ^2_{LS} =0.105

LAD_1: $\hat{Y}_{LAD1}$=0.841 logP-2.149 R^2=97.54% σ^2_{LAD} =0.089