# Régression de modèle II – Guide

Pierre Legendre

Département de sciences biologiques, Université de Montréal,
C.P. 6128, succursale Centre-ville, Montréal, Québec H3C 3J7, Canada
*E-mail:* Pierre.Legendre@umontreal.ca

*Septembre 2001*

Ce programme calcule des régressions de modèle II selon différentes méthodes: l'axe majeur (AM), l'axe majeur réduit (AMR), les moindres carrés ordinaires (MCO) et l'axe majeur des données cadrées (AMDC). Ces méthodes sont décrites à la section 10.3.2 de Legendre & Legendre (1998) ainsi qu'aux sections 14.13 et 15.7 de Sokal & Rohlf (1995)[*]. Le programme calcule l'intervalle de confiance paramétrique à 95% pour les estimations de la pente et de l'ordonnée à l'origine. Un test par permutation peut être réalisé pour établir la signification de la pente de AM, de OLS et de AMDC, ainsi que pour le coefficient de corrélation.

La méthode de régression de modèle II par les trois groupes de Bartlett, qui est décrite par les auteurs mentionnés ci-dessus, n'est pas calculée par le programme parce qu'elle présente plusieurs problèmes. Le principal est que la droite de régression n'est pas la même selon que le groupement (en trois groupes) est réalisé à partir de $x$ ou de $y$. La droite de régression ne passe pas nécessairement par le centroïde du nuage de points et l'estimateur de la pente n'est pas symétrique, i.e. la pente de la régression $y = f(x)$ n'est pas la réciproque de la pente de la régression $x = f(y)$.

Il faut utiliser la régression de modèle II lorsque les deux variables sont aléatoires, c'est-à-dire non contrôlées par le chercheur. La régression de modèle I par moindres carrés ordinaires sous-estime la pente de la relation linéaire lorsque les deux variables contiennent de l'erreur; voir l'exemple 4 ci-dessous. Des recommandations détaillées suivent.

### *Recommandations quant au choix d'une méthode de régression de modèle II*

Tenant compte de plusieurs études par simulation, Legendre & Legendre (1998) recommandent la démarche suivante pour choisir une méthode permettant d'estimer les paramètres de la relation fonctionnelle linéaire reliant deux variables aléatoires mesurées avec erreur (tableau 1).

---

[*] Dans Sokal & Rohlf (*Biometry, 2nd edition*, 1981: 551), le résultat fourni pour la régression AM pour les données de leur exemple est erroné. L'erreur a été corrigée dans l'édition de 1995.

**Tableau 1** Application des méthodes de régression de modèle II. Les numéros dans la colonne de gauche se réfèrent aux paragraphes correspondants dans le texte.

| Par. | Méthode | Conditions d'application | Test possible |
|---|---|---|---|
| 1 | MCO | L'erreur de $y$ >> l'erreur de $x$ | Oui |
| 3 | AM | La distribution est binormale<br>Les variables sont dans les mêmes unités physiques ou sans dimension<br>La variance de l'erreur est à peu près la même pour $x$ et $y$ | Oui |
| 4 | | La distribution est binormale<br>La variance de l'erreur de chaque axe est proportionnelle à la variance de la variable correspondante | |
| 4.1 | AMDC | Vérifier le diagramme de dispersion: pas de valeurs aberrantes ou extrêmes | Oui |
| 4.2 | AMR | La corrélation $r$ est significativement différente de zéro | Non |
| 5 | MCO | La distribution n'est pas binormale<br>La relation entre $x$ et $y$ est linéaire | Oui |
| 6 | MCO | On désire calculer des valeurs $\hat{y}$ prévues ou prédites par l'équation<br>(L'équation de régression et les intervalles de confiance ne peuvent être utilisés) | Oui |
| 7 | AM | Pour comparer des observations aux prédictions d'un modèle | Oui |

1. Si la variation aléatoire (i.e. la variance de l'erreur[*]) de la variable-réponse $y$ est beaucoup plus forte (plus de trois fois) que celle de la variable explicative $x$, utiliser la méthode des moindres carrés ordinaires (MCO). Sinon:

2. Vérifier si les données sont approximativement binormales par l'examen d'un diagramme de dispersion ou par un test de signification. Sinon, tenter une transformation pour rendre la distribution binormale. Si les données sont ou peuvent être rendues binormales, voir les recommandations 3 et 4. Sinon passer à 5.

3. Lorsque les deux variables sont exprimées dans les mêmes unités physiques ou sont sans dimension (p. ex. des variables ayant subi une transformation log) et que la variance de l'erreur est à peu près la même pour les deux variables, utiliser l'axe majeur (AM).

Lorsqu'on ne possède aucune information sur le rapport des variances des erreurs alors qu'il n'y a pas de raison de croire que ce rapport puisse différer de 1, on peut employer AM si les résultats sont interprétés avec prudence. La méthode AM produit des estimations non biaisées de la pente et des intervalles de confiance aux bornes exactes (Jolicoeur, 1990).

[*] Contrairement à la variance de l'échantillon, la variance de l'erreur ne peut être estimée à partir des données. On ne peut l'estimer qu'en considérant la méthode qui a été employée pour mesurer les variables $x$ et $y$.

La méthode AM peut être employée avec des variables qui sont dimensionnellement hétérogènes si l'objectif de l'analyse est (1) de comparer la pente de la relation fonctionnelle entre deux variables (les mêmes) mesurées dans des conditions différentes (p. ex. à deux ou plusieurs sites d'échantillonnage), ou encore (2) de tester l'hypothèse que l'axe majeur ne diffère pas significativement d'une valeur fournie par hypothèse (p. ex. la relation $E = b_1 m$ où, selon la célèbre équation d'Einstein, $b_1 = c^2$, c étant la vitesse de la lumière dans le vide).

4. Pour des données binormales, si la régression AM ne peut pas être employée parce que les variables ne sont pas exprimées dans les mêmes unités physique ou parce que la variance de l'erreur des deux variables diffère, il reste deux méthodes pour estimer les paramètres de l'équation fonctionnelle linéaire si on peut raisonnablement supposer que la variance de l'erreur de chaque axe est proportionnelle à la variance de la variable correspondante, i.e. si (la variance de l'erreur de *y* / la variance de *y*) ≈ (la variance de l'erreur de *x* / la variance de *x*). Cette condition est souvent satisfaite par des dénombrements d'organismes (p. ex. le nombre de plantes ou d'animaux) ou par des variables ayant subi une transformation log (McArdle, 1988).

4.1. On peut employer l'axe majeur des données cadrées (AMDC). La méthode est décrite ci-dessous. Attention, cependant, aux valeurs aberrantes ou extrêmes; on peut identifier celles-ci grâce à un diagramme de dispersion des objets.

4.2. On peut employer l'axe majeur réduit (AMR). Il faut d'abord tester la signification de la corrélation (*r*) entre les deux variables afin de déterminer si l'hypothèse de l'existence d'une liaison entre les deux variables est supportée par les données. Il n'y a pas lieu de calculer une équation AMR si la corrélation n'est pas significative.

L'AMR demeure une solution boiteuse puisqu'on ne peut pas tester la signification de sa pente. L'intervalle de confiance doit aussi être employé avec circonspection car, comme on l'a montré à l'aide de simulations, à mesure que la pente réelle de la distribution s'éloigne de ±1, l'estimation de la pente par l'AMR devient de plus en plus biaisée et l'intervalle de confiance inclut la valeur réelle de la pente de moins en moins souvent. Même lorsque la pente de la distribution est proche de ±1 (p. ex. dans l'exemple 5), l'intervalle de confiance AMR est trop étroit si l'effectif (*n*) est très petit ou si la corrélation est faible.

5. Si la distribution des données n'est pas et ne peut être rendue binormale (p. ex. si la distribution possède deux ou plusieurs modes), il faut se demander si la pente d'une droite de régression est un modèle adéquat pour décrire la relation fonctionnelle entre les deux variables. Puisque la distribution n'est pas binormale, il est incorrect d'employer AM, AMR ou AMDC puisque ces modèles décrivent la première composante principale d'une distribution binormale. (a) Si la relation est linéaire, il vaut mieux utiliser les moindres carrés ordinaires (MCO); il faut alors tester la pente par permutation puisque la condition de binormalité n'est pas remplie. (2) Si une ligne droite ne semble pas être un modèle adéquat pour décrire la relation entre les variables, il vaut mieux employer la régression polynomiale ou la régression non-linéaire.

6. Si le but de l'étude n'est pas d'estimer les paramètres d'une relation fonctionnelle, mais plutôt de prévoir ou prédire les valeurs de *y* à partir de *x*, il faut employer la méthode MCO.

C'est la seule méthode qui minimise la somme des carrés des résidus en *y*. La droite de régression MCO elle-même n'a aucune signification, de même que l'erreur type du coefficient de régression et son intervalle de confiance, à moins que *x* n'ait été mesuré sans erreur (Sokal & Rohlf, 1995: 545, tableau 14.3); cet avertissement s'applique en particulier à l'intervalle de confiance de la pente MCO calculé par ce programme.

7. Dans certaines recherches on désire comparer des observations aux prédictions d'un modèle. Si le modèle contient des variables sujettes à fluctuation aléatoire, il faut utiliser l'axe majeur (MA) pour cette comparaison puisque les valeurs observées de même que les prédictions du modèle devraient être dans les mêmes unités physiques.

Si le modèle colle bien aux données, on s'attend à trouver une pente de 1 et une ordonnée à l'origine de 0. Si la pente diffère significativement de 1, cela indique que la *différence* entre les valeurs observées et simulées est proportionnelle aux valeurs observées. Pour des variables à échelle de variation relative à un vrai zéro, si l'ordonnée à l'origine diffère significativement de 0, cela suggère l'existence d'une différence systématique entre les valeurs observées et simulées (Mesplé *et al.*, 1996).

8. Attention: dans toutes ces méthodes, l'intervalle de confiance est grand lorsque le nombre d'observations est faible. Il diminue à mesure que *n* augmente jusqu'à 60 environ; après cela, il change beaucoup plus lentement. La régression de modèle II ne devrait idéalement être employée qu'avec des jeux de données de 60 observations ou plus. Les exemples ci-dessous contiennent parfois moins d'observations; ils ne sont présentés qu'à titre d'illustration.

### *Axe majeur des données cadrées*

La régression selon l'axe majeur des données cadrées (AMDC) n'est décrit que dans Legendre & Legendre (1998: 511-512). Cette méthode se calcule comme suit:

1. Cadrer les variables *y* et *x*, ce qui consiste à les transformer en de nouvelles variables *y*' et *x*' dont l'étendue de variation est 1. Il y a deux formules possibles pour le cadrage, selon la nature des variables:

- Pour les variables notées par rapport à un zéro arbitraire (variables à échelle de variation par intervalle, p. ex. la température en °C), la formule pour le cadrage est:

$$y'_i = \frac{(y_i - y_{min})}{(y_{max} - y_{min})} \quad \text{ou} \quad x'_i = \frac{(x_i - x_{min})}{(x_{max} - x_{min})} \qquad \textbf{(1)}$$

- Pour les variables notées par rapport à un vrai zéro (variables à échelle de variation relative, p. ex. les abondances d'espèces ou encore la température en °K), la formule de cadrage suppose que la valeur minimum de la variable est 0. L'éq. 1 se réduit donc à:

$$y'_i = \frac{y_i}{y_{max}} \quad \text{ou} \quad x'_i = \frac{x_i}{x_{max}} \qquad \textbf{(2)}$$

2. Calculer la régression AM entre les variables cadrées *y*' et *x*'. Tester la signification de la pente par permutation si cela est nécessaire.

3. Rétro-transformer la pente estimée, ainsi que son intervalle de confiance, dans les unités physiques originelles des variables en les multipliant par le rapport des plages de variation:

$$b_1 = b'_1 \frac{y_{max} - y_{min}}{x_{max} - x_{min}} \quad \textbf{(3)}$$

4. Recalculer l'ordonnée à l'origine $b_0$ ainsi que son intervalle de confiance, à l'aide du centroïde originel $(\bar{x}, \bar{y})$ du nuage de points ainsi que l'estimation de la pente $b_1$ et des bornes de son intervalle de confiance:

$$b_0 = \bar{y} - b_1 \bar{x} \quad \textbf{(4)}$$

L'estimateur de la pente AMDC possède plusieurs propriétés intéressantes lorsque les variables *x* et *y* ne sont pas dans les mêmes unités physiques ou lorsque la variance de l'erreur sur les deux axes diffère. (1) L'estimateur de la pente suit les changements des unités physiques des deux variables: la position de la droite de régression dans le nuage de points demeure la même quelle que soit le changement linéaire d'échelle que subissent les variables. AMR possède également cette propriété, mais pas AM. (2) L'estimateur de la pente répond à la covariance des variables; tel n'est pas le cas de AMR. (3) Enfin, et contrairement à AMR, on peut tester l'hypothèse que la pente de AMDC est égale à une valeur posée par hypothèse, en particulier 0 ou 1. Comme pour AM, ce test peut se faire par permutation, ou alors en comparant l'intervalle de confiance de la pente à la valeur qui nous intéresse. Donc, lorsqu'on ne peut pas employer AM parce que les deux variables ne sont pas dans les mêmes unités physiques ou encore parce que la variance de l'erreur sur les deux axes diffère, on peut alors employer la régression AMDC. Il n'y a cependant pas de raison d'employer AMDC lorsqu'on peut utiliser AM.

Avant de calculer une régression AMDC, il importe de tracer un diagramme de dispersion pour chercher si le jeu données comporte des valeurs aberrantes ou extrêmes. Il ne faut pas employer AMDC avec de telles valeurs car elles changent l'estimation de l'étendue de variation des variables. Des données aberrantes qui ne sont pas alignées dans l'axe de l'ellipse de dispersion des points peuvent avoir une influence indésirable sur l'estimation de la pente. L'identification et le traitement des données aberrantes est discuté par Sokal & Rohlf (1995, section 13.4). Celles-ci peuvent, dans certains cas, être éliminées du jeu de données ou encore subir la procédure de "winsorization" décrite par ces auteurs.

### *Fichier de données*

Le fichier de données contient un tableau rectangulaire dont les lignes correspondent aux objets et les colonnes aux variables. Il ne doit y avoir d'identificateurs ni pour les lignes,

ni pour les colonnes. On peut placer la variable explicative *x* ou encore la variable réponse *y* en premier; le programme demande à l'utilisateur quelle est la position des variables dans le tableau (*x* d'abord ou *y* d'abord). Les colonnes sont séparées par un tabulateur ou encore par une ou des espaces; les espaces en début de ligne sont ignorées. Ce fichier doit être sauvegardé en format ASCII (texte).

***Fichier de résultats***

Ce fichier contient les résultats suivants:

1. Équation de régression linéaire simple ($\hat{y} = b_0 + b_1 x$) selon l'axe majeur (AM). Comme son nom l'indique, cette droite est l'axe majeur de l'ellipse de dispersion des points; on l'appelle aussi la première composante principale. Les détails du calcul se trouvent dans Sokal & Rohlf (1995, Box 15.6). Le programme calcule l'équation de régression de même que l'intervalle de confiance paramétrique à 95% de la pente et de l'ordonnée à l'origine.

2. Même résultat pour l'axe majeur réduit (AMR). L'axe majeur réduit est l'axe majeur calculé à partir des données centrées réduites; la pente est rétro-transformée en fonction des unités physiques originelles des variables. La pente est estimée par le rapport $b_{\mathrm{AMR}} = \pm s_y/s_x$. Si les deux variables ont à peu près la même variance, $b_{\mathrm{AMR}} \approx \pm 1$.

3. Même résultat pour la régression par moindres carrés ordinaires (MCO). Le programme calcule également les coefficients de corrélation ($r$) et de détermination ($r^2$). Dans le cas de données aléatoires, $b_{\mathrm{MCO}} \approx 0$.

4. Même résultat pour l'axe majeur des données cadrées (AMDC). L'axe majeur des données cadrées est l'axe majeur calculé à partir des données cadrées. Les détails du calcul sont présentés à la section précédente. La méthode AMDC est proposée comme option du programme parce que sa mise en oeuvre demande à l'usager de répondre à des questions supplémentaires à propos du cadrage des variables; il peut ne pas vouloir prendre ces décisions à moins d'être intéressé à obtenir le résultat de la régression AMDC.

5. Des tests par permutation sont proposés pour la pente de AM, MCO et AMDC ainsi que pour le coefficient de corrélation *r*. Les résultats des tests de $b_{\mathrm{MCO}}$ et de la corrélation *r* sont toujours identiques puisque ces deux tests sont équivalents. La pente de AMR ne peut pas être testée par permutation (Legendre & Legendre, 1998: 511). Lorsque la pente de AM et AMDC est plus grande que 1 (ou < –1), le test par permutation est réalisé sur la pente $b' = 1/b$ de la régression de *x* sur *y* qui est, elle, < 1 (ou > –1) (Legendre & Legendre, 1998: 508). La même chose se passe pour AMDC; dans ce cas, le test concerne la pente des données cadrées et non la pente rétro-transformée dans les unités physiques originelles des variables.

Pour la pente de AM, MCO et AMDC, le test par permutation utilise comme statistique l'estimation *b* de la pente elle-même. En régression linéaire simple par MCO, un

test par permutation basé sur $r$ est équivalent à un test par permutation basé sur la statistique pivotale $t$ associée à $b_{MCO}$ (Legendre & Legendre, 1998: 21). Par ailleurs, au cours des permutations, l'estimation de la pente ($b_{MCO}$) ne diffère de $r$ que par une constante ($s_y/s_x$) puisque $b_{MCO} = r_{xy}\, s_y/s_x$, si bien que $b_{MCO}$ et $r$ sont des statistiques équivalentes pour un test par permutation. Par conséquent, un test par permutation basé sur $b_{MCO}$ est équivalent à un test par permutation basé sur la statistique $t$ associée à $b_{MCO}$. Cela n'est cependant pas le cas en régression linéaire multiple, comme l'ont montré Anderson & Legendre (1999).

Si l'étude a simplement comme objectif d'établir s'il existe une relation significative entre deux variables, il suffit de calculer la corrélation $r$ et de tester sa signification. On peut utiliser un test paramétrique lorsque les données sont approximativement binormales ou un test par permutation lorsque ce n'est pas le cas.

L'intervalle de confiance de l'ordonnée à l'origine des MCO est calculé à l'aide des formules habituelles qui se trouvent dans les manuels de statistique; les résultats sont identiques à ceux des programmes commerciaux. Il n'existe cependant pas de telle formule, garantissant une erreur de type I correcte, pour les trois autres méthodes. Dans le programme, l'intervalle de confiance de l'ordonnée à l'origine de AM, AMR et AMDC est calculé en projetant les bornes de l'intervalle de confiance de la pente sur l'ordonnée; cela produit une sous-estimation de ces intervalles de confiance.

En régression AM et AMDC, les bornes de l'intervalle de confiance de la pente peuvent parfois se trouver en dehors des quadrants I et IV du plan centré sur le centroïde de la distribution bivariable. Lorsque la *borne inférieure* de l'intervalle de confiance correspond à une ligne dans le quadrant III (figure 1a), elle a une pente positive; la droite de régression AMDC de l'exemple 5 fournit un exemple de ce phénomène. De même, lorsque la *borne supérieure* de l'intervalle de confiance correspond à une ligne dans le quadrant II (figure 1b), elle a une pente négative. Dans d'autres cas, l'intervalle de confiance de la pente peut occuper les 360° du plan. Elle n'a alors pas de bornes et celles-ci sont notées 0.00000; voir l'exemple 5.

En régression AMR et MCO, les bornes de l'intervalle de confiance ne peuvent se trouver à l'extérieur des quadrants I et IV. En AMR, la droite de régression a toujours un angle de +45° ou –45° dans l'espace des données centrées réduites; la pente AMR est une rétro-transformation de ±45° dans les unités physiques originelles des variables. En MCO, la pente se trouve toujours à un angle plus proche de zéro que l'axe majeur; en d'autres termes, cette droite sous-estime toujours, en valeur absolue, la pente de l'axe majeur.

Une "commande cachée" permet d'obtenir la liste des valeurs de référence et des valeurs permutées obtenues au cours des tests de la pente de AM et AMDC. En réponse à la question:

```
Calculer l'axe majeur des données cadrées (AMDC)?
(0) non,  (1) oui
```

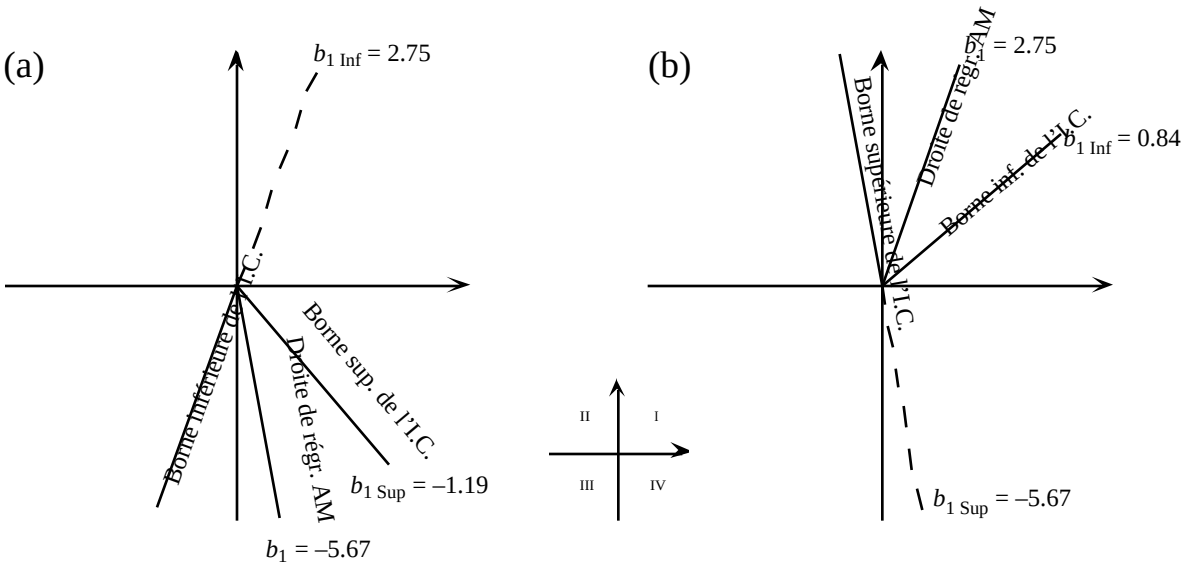


**Figure 1** (a) Si la borne inférieure de l'intervalle de confiance (I.C.) de la pente de MA se trouve dans le quadrant III, cette borne a une pente positive (+2.75 dans l'exemple). (b) De la même façon, si la borne supérieure de l'intervalle de confiance de la pente de MA se trouve dans le quadrant II, cette borne a une pente négative (–5.67 dans l'exemple).

si on répond –1 plutôt que 1, le programme inscrit dans le fichier "Permutations" les valeurs de *bma* (pente de AM, qui est convertie en $1/b_{\mathrm{AM}}$ par le programme si $b_{\mathrm{AM}} > 1$), *brma* (pente de AMDC dans l'espace des variables cadrées, qui est onvertie en $1/b_{\mathrm{AMDC}}$ par le programme si $b_{\mathrm{AMDC}} > 1$) et *brmaC* (pente de AMDC rétro-transformée dans les unités physiques originelles des variables *x* et *y*). Ces valeurs sont calculées par le sous-programme *Regression.* Les valeurs de référence (non permutées) se trouvent sur la première ligne de ce fichier.

### *Exemple 1*

Fichier de données

Dans cet exemple, on compare des observations aux prévisions d'un modèle. Le département de chirurgie d'un hôpital désirait prévoir la survie de patients ayant à subir une certaine intervention au foie. Quatre variables explicatives furent mesurées sur les patients. La variable-réponse *Y* était la durée de survie; celle-ci subit une transformation $\log_{10}$. Les données sont décrites en détail à la section 8.2 de Neter *et al.* (1996); on trouvera d'ailleurs les données elles-mêmes dans ce manuel. Les données furent divisées en deux groupes de 54 patients. Le premier groupe de données a été utilisé pour construire des modèles prévisionnels alors que le second groupe fut mis de côté pour permettre la validation de ces

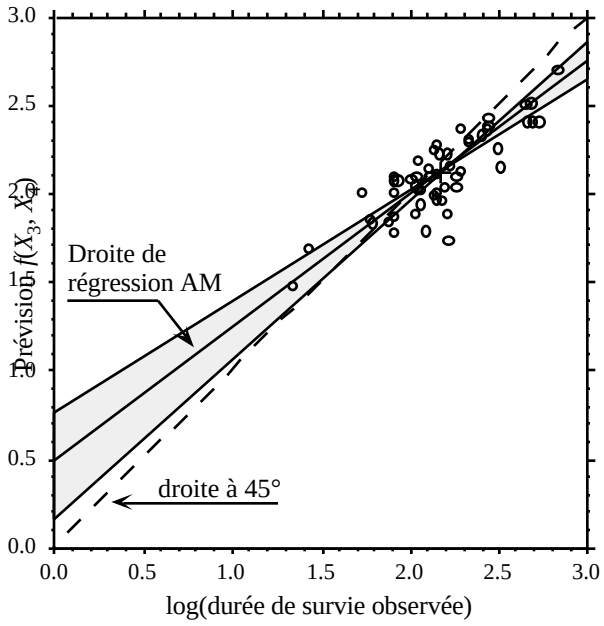


**Figure 2** Diagramme de dispersion des données de l'exemple 1 montrant la droite de régression de l'axe majeur (AM) ainsi que son intervalle de confiance à 95% (en gris). La droite à 45° (tirets) est dessinée à titre de référence. La croix indique le centroïde du nuage de points. La droite de régression AM passe par ce centroïde.

modèles. Plusieurs modèles de régression furent examinés. L'un d'eux, qui est construit à partir des variables $X_3$ = résultat d'un test des fonctions enzymatiques et $X_4$ = résultat d'un test de fonctionnement du foie, est utilisé dans le présent exemple. L'équation de régression multiple est la suivante:

$$\hat{Y} = 1.388778 + 0.005653\, X_3 + 0.139015\, X_4$$

Cette équation fut appliquée au second jeu de données, constitué également de 54 patients, pour produire des durées de survie prévues par le modèle. Dans notre exemple, ces valeurs sont comparées aux survies observés. La figure 2 présente le diagramme de dispersion. Les $\log_{10}$(durée de survie observée) sont en abscisse alors que les valeurs prévues par le modèle sont en ordonnée. Pour fins de comparaison, la figure montre la droite à 45° qui correspondrait à des prévisions parfaites de la survie des patients.

Fichier de résultats: Équations de régression AM, AMR et MCO, intervalles de confiance, tests de signification. La méthode AMDC, qui est optionnelle, n'a pas été calculée puisque seule la méthode AM est appropriée pour cet exemple.

```
Régression de modèle II
pour les cas où y et x sont deux variables aléatoires.

Estimation des paramètres  de l'équation fonctionnelle
y = b0 + b1*x
                    © Pierre Legendre, 1994, 1999, 2000
Département de sciences biologiques, Univ. de Montréal

                ------------------------
```

```
Méthodes:   axe majeur (AM),  axe majeur réduit (AMR),
moindres carrés ordinaires (MCO),
axe majeur des données cadrées (AMDC).

Tests par permutation: r et pentes de AM, MCO et AMDC.
On ne peut pas tester la pente de AMR par permutation.
-----------------------------------------------------

Fichier de données: EX1_54x2.txt

---------------------------------------------------------------------

Axe majeur (AM,"MA"):
Valeurs propres: lambda 1 =       0.13324  lambda 2 =       0.01090

b0 =   0.48720    b1 =   0.74921    angle (°) =  36.84093

I.C. 95% de la pente = [   0.62166,    0.89456]
I.C. 95% de l'ordonnée à l'origine
                     = [   0.17258,    0.76331]   (sous-estimation)

---------------------------------------------------------------------

Axe majeur réduit (AMR,"SMA"):

b0 =   0.41155    b1 =   0.78416    angle (°) =  38.10197

I.C. 95% de la pente = [   0.67428,    0.91193]
I.C. 95% de l'ordonnée à l'origine
                     = [   0.13496,    0.64939]   (sous-estimation)

I.C. de la pente selon Jolicoeur & Mosimann (1968),  McArdle (1988)
---------------------------------------------------------------------

Moindres Carrés Ordinaires (MCO,"OLS"):
          r =   0.83873   coeff. de détermination (r^2) =   0.70347

b0 =   0.68530    b1 =   0.65770    angle (°) =  33.33276

I.C. 95% de la pente = [   0.53887,    0.77652]
I.C. 95% de l'ordonnée à l'origine
                     = [   0.42569,    0.94490]

---------------------------------------------------------------------

b0=ord. à l'origine, b1=pente, xbar=     2.16466, ybar=     2.10898

Code 999.99999 si la pente est infinie (angle = 90°)

Code   0.00000 si les bornes de l'I.C. ne peuvent être
calculées; quand les deux valeurs propres sont trop
semblables, l'I.C. incorpore les 360° du plan.
```

```
----------------------------------------
Test par permutation des pentes et de r
----------------------------------------

Nombre de permutations aléatoires =   999

-----------------------------------------------
Méthode   Stat.    PP    EG    PG  p unilatérale
-----------------------------------------------

AM      0.74921   999     1     0     0.00100

MCO     0.65770   999     1     0     0.00100

Corr    0.83873   999     1     0     0.00100
-----------------------------------------------
```

L'aspect le plus intéressant de l'équation de régression AM est que la droite de régression n'est pas parallèle à la droite à 45° dessinée dans la figure 2. La droite à 45° n'est pas incluse dans l'intervalle de confiance à 95% de la pente AM, qui s'étend depuis $\tan^{-1}(0.62166) = 31.87°$ jusqu'à $\tan^{-1}(0.89456) = 41.81°$. La figure montre que l'équation prévisionnelle sur-estime les valeurs de survie en dessous de la moyenne et les sous-estime au dessus de la moyenne. Notez que la droite de régression MCO que des chercheurs utilisent souvent (à tort) pour des comparaisons de ce type, aurait montré une différence encore plus grande (angle de 33.3°), par rapport à la droite à 45°, que la droite de régression AM (angle de 36.8°).

***Exemple 2***

Fichier de données

Le tableau ci-dessous contient des observations réalisées à 20 sites lors d'une étude de relations prédateur-proie (Hines *et al.* 1998). *y* est le nombre de bivalves (*Macomona liliana*) de taille supérieure à 15 mm trouvés dans le sédiment dans des quadrats de 0.25 $m^2$; *x* est le nombre de traces de prédateurs, des raies pastenagues (*Myliobatis tenuicaudatus*), observées sur le sédiment dans un rayon de 15 m du quadrat des bivalves.

| N. prédateurs *x* | No. proies *y* | N. prédateurs *x* | No. proies *y* |
|---|---|---|---|
| 2 | 29 | 11 | 53 |
| 3 | 18 | 5 | 27 |
| 8 | 29 | 8 | 41 |
| 4 | 29 | 15 | 58 |
| 8 | 49 | 7 | 31 |
| 3 | 33 | 13 | 71 |
| 8 | 53 | 18 | 59 |
| 9 | 33 | 6 | 42 |
| 2 | 20 | 19 | 71 |
| 4 | 40 | 14 | 54 |

Les valeurs de *x* et *y* sont exprimées dans les mêmes unités, les deux variables contiennent de l'erreur et leur distribution est approximativement binormale. La variance de l'erreur n'est pas la même pour *x* et *y* mais, puisqu'il s'agit de dénombrements d'animaux, la variance de l'erreur de chaque axe est vraisemblablement proportionnelle à la variance de la variable correspondante. Le coefficient de corrélation linéaire est significatif: $r = 0.86$, $p < 0.001$. Les méthodes AMDC et AMR sont donc appropriées pour cet exemple; AM et MCO ne le sont pas. La figure 3 présente le diagramme de dispersion. Les différentes droites de régression sont présentées pour permettre leur comparaison.

Fichier de résultats: Équations de régression AM, AMR, MCO et AMDC, intervalles de confiance, tests de signification (en-tête éliminé). On remarque que l'intervalle de confiance de l'ordonnée à l'origine de AMR et AMDC n'inclut pas la valeur 0. Il peut y avoir plusieurs raisons à cela: (1) la relation entre les variables peut ne pas être linéaire; (2) l'étendue de l'intervalle de confiance de l'ordonnée à l'origine est sous-estimée; (3) les prédateurs (raies pastenagues) peuvent ne pas être attirés par les sites où se trouvent peu de proies (bivalves).

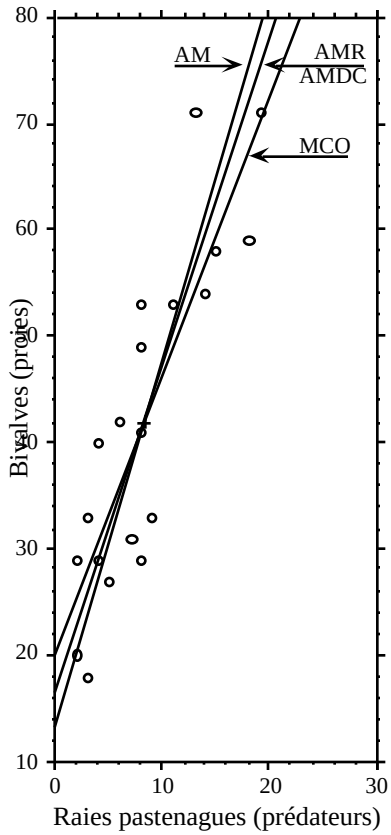


**Figure 3** Diagramme de dispersion des données de l'exemple 2 (nombre de bivalves en fonction du nombre de raies pastenagues) montrant les droites de régression de l'axe majeur (AM), l'axe majeur réduit (AMR), les moindres carrés ordinaires (MCO) et l'axe majeur des données cadrées (AMDC). AMR et AMDC sont les méthodes de régression appropriées pour cet exemple. La croix indique le centroïde du nuage de points. Les quatre droites de régression passent par ce centroïde.

```
-----------------------------------------------------

Fichier de données: EX2_20x2.txt

-----------------------------------------------------------------

Axe majeur (AM,"MA"):
Valeurs propres: lambda 1 =     269.82124  lambda 2 =       6.41823

b0 =  13.05968    b1 =   3.46591    angle (°) =  73.90584

I.C. 95% de la pente = [   2.66310,    4.86857]
I.C. 95% de l'ordonnée à l'origine
                     = [   1.34742,   19.76310]   (sous-estimation)

-----------------------------------------------------------------

Axe majeur réduit (AMR,"SMA"):

b0 =  16.45205    b1 =   3.05963    angle (°) =  71.90073

I.C. 95% de la pente = [   2.38281,    3.92871]
I.C. 95% de l'ordonnée à l'origine
                     = [   9.19529,   22.10353]   (sous-estimation)

I.C. de la pente selon Jolicoeur & Mosimann (1968),  McArdle (1988)
```

```
-----------------------------------------------------------------

Moindres Carrés Ordinaires (MCO,"OLS"):
          r =   0.86008   coeff. de détermination (r^2) =   0.73974

b0 =  20.02675    b1 =   2.63153    angle (°) =  69.19283

I.C. 95% de la pente = [   1.85858,    3.40448]
I.C. 95% de l'ordonnée à l'origine
                     = [  12.49099,   27.56251]

-----------------------------------------------------------------

Axe majeur des données cadrées (AMDC,"RMA"):
ymin =   0.00000 ymax =  71.00000 xmin =   0.00000 xmax =  19.00000
Valeurs propres: lambda 1 =       0.11509  lambda 2 =       0.00827

b0 =  17.25651    b1 =   2.96329    angle (°) =  71.35239

I.C. 95% de la pente = [   2.17426,    3.95653]
I.C. 95% de l'ordonnée à l'origine
                     = [   8.96300,   23.84493]   (sous-estimation)

-----------------------------------------------------------------

b0=ord. à l'origine, b1=pente, xbar=     8.35000, ybar=    42.00000

Code 999.99999 si la pente est infinie (angle = 90°)

Code   0.00000 si les bornes de l'I.C. ne peuvent être
calculées; quand les deux valeurs propres sont trop
semblables, l'I.C. incorpore les 360° du plan.


----------------------------------------
Test par permutation des pentes et de r
----------------------------------------

Nombre de permutations aléatoires =   999

-------------------------------------------------
Méthode   Stat.    PP    EG    PG  p unilatérale                [1]
-------------------------------------------------

AM      3.46591   999     1     0     0.00100

MCO     2.63153   999     1     0     0.00100

Corr    0.86008   999     1     0     0.00100

AMDC    2.96329   999     1     0     0.00100
-------------------------------------------------
```

[1] Dans ce tableau du fichier de résultats, les lignes correspondent respectivement aux pentes AM, MCO et AMDC ainsi qu'au coefficient de corrélation *r* ('Corr'). 'Stat.' est la

valeur de la statistique soumise au test. Tel qu'expliqué dans la section "Fichier de sortie", la statistique utilisée par le programme pour le test de la pente AM, dans cet exemple, est l'inverse de l'estimation de la pente $b_{MA}$ (1/3.46591 = 0.28852) puisque la valeur de référence de la statistique dans ce test par permutation ne doit pas excéder 1.

Les probabilités unilatérales ('p. unilatérale') sont calculées dans la direction du signe du coefficient. Pour un test unilatéral dans la queue supérieure de la distribution (c'est-à-dire pour un coefficient ayant un signe positif), p = (EG + PG)/(Nombre de permutations + 1). Pour un test dans la queue inférieure de la distribution (c'est-à-dire pour un coefficient ayant un signe négatif), p = (PP + EG)/(Nombre de permutations + 1), où

- PP est le nombre de valeurs obtenues au cours des permutations qui sont inférieures à la valeur de référence de la statistique;
- EG est le nombre de valeurs obtenues au cours des permutations qui sont égales à la valeur de référence de la statistique;
- PG est le nombre de valeurs obtenues au cours des permutations qui sont supérieures à la valeur de référence de la statistique.

***Exemple 3***

Fichier de données

Le tableau qui suit présente les données utilisées par Sokal & Rohlf (1995, Box 14.12) pour illustrer les méthodes de régression de modèle II. Elles concernent la masse ($x$) de femelles n'ayant pas encore frayé d'un poisson de la Californie, le cabezon (*Scorpaenichthys marmoratus*), ainsi que le nombre d'oeufs produits par chacune ($y$). On peut être intéressé à estimer l'équation fonctionnelle reliant le nombre d'oeufs à la masse des femelles matures. Les unités physiques des variables sont celles utilisées par Sokal & Rohlf (1995: 546) dans leur tableau.

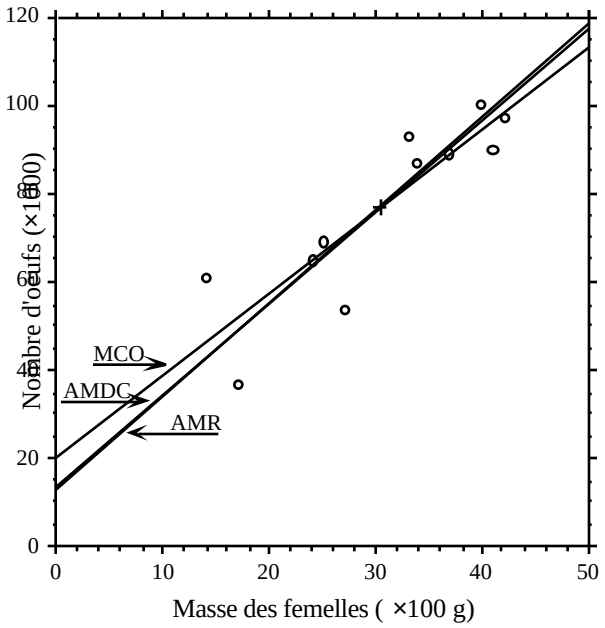


**Figure 4** Diagramme de dispersion des données de l'exemple 3 (nombre d'oeufs produits en fonction de la masse des femelles) montrant les droites de régression de l'axe majeur des données cadrées (AMDC), l'axe majeur réduit (AMR) et les moindres carrés ordinaires (MCO). AMDC et AMR sont les méthodes de régression appropriées pour cet exemple. La croix indique le centroïde du nuage de points. Les trois droites de régression passent par ce centroïde.

| Masse (×100 g) *x* | N. oeufs (×1000) *y* |
|---|---|
| 14 | 61 |
| 17 | 37 |
| 24 | 65 |
| 25 | 69 |
| 27 | 54 |
| 33 | 93 |
| 34 | 87 |
| 37 | 89 |
| 40 | 100 |
| 41 | 90 |
| 42 | 97 |

Puisque les variables sont exprimées dans des unités physiques différentes et sont estimées avec erreur et que leur distribution est approximativement binormale, AMDC et AMR sont appropriés pour cet exemple; AM est incorrect. La droite de régression MCO n'a pas de sens; en régression de modèle II, la méthode MCO ne doit être employée que pour calculer des valeurs prévues ou prédites par l'équation. Cette droite est tracée dans la figure 4 simplement pour fins de comparaison.

Les droites de régression AMDC et AMR sont pratiquement indistinguables l'une de l'autre dans cet exemple. On peut tester la signification de la pente AMDC ($H_0$: $b_{RMA} = 0$), mais pas celle de la pente AMR. Même s'ils sont sous-estimés, les intervalles de confiance à 95% de l'ordonnée à l'origine de AMDC et AMR incluent la valeur 0, comme on s'y attend si un modèle linéaire s'applique aux données: une femelle ayant une masse nulle ne devrait produire aucun oeuf.

Une autre propriété intéressante des méthodes AMDC et AMR est que l'estimation de leur pente et de leur ordonnée à l'origine change proportionnellement aux unités de mesure. On peut facilement vérifier cela en changeant la position du point décimal dans les données de l'exemple 2 et en recalculant les équations de régression. AMDC et AMR, de même que MCO, possèdent cette propriété. La régression AM ne la possède pas; c'est la raison pour laquelle cette méthode ne doit être employée qu'avec des variables exprimées dans les mêmes unités physiques, comme celles de l'exemple 1.

Fichier de résultats: Équations de régression AM, AMR, MCO et AMDC, intervalles de confiance à 95% et tests de signification (en-tête éliminé).

```
-----------------------------------------------------

Fichier de données: EX3_11x2.txt

-------------------------------------------------------------------

Axe majeur (AM,"MA"):
Valeurs propres: lambda 1 =     494.63400  lambda 2 =      17.49327

b0 =   6.65663    b1 =   2.30173    angle (°) =  66.51716

I.C. 95% de la pente = [   1.60430,    3.72440]
I.C. 95% de l'ordonnée à l'origine
                     = [ -36.54082,   27.83295]   (sous-estimation)

-------------------------------------------------------------------

Axe majeur réduit (AMR,"SMA"):

b0 =  12.19378    b1 =   2.11937    angle (°) =  64.74023

I.C. 95% de la pente = [   1.49672,    3.00104]
I.C. 95% de l'ordonnée à l'origine
                     = [ -14.57693,   31.09957]   (sous-estimation)

I.C. de la pente selon Jolicoeur & Mosimann (1968),  McArdle (1988)
```

```
-----------------------------------------------------------------

Moindres Carrés Ordinaires (MCO,"OLS"):
          r =   0.88232   coeff. de détermination (r^2) =   0.77849

b0 =  19.76682    b1 =   1.86996    angle (°) =  61.86337

I.C. 95% de la pente = [   1.11780,    2.62211]
I.C. 95% de l'ordonnée à l'origine
                     = [  -4.09838,   43.63201]

-----------------------------------------------------------------

Axe majeur des données cadrées (AMDC,"RMA"):
ymin =   0.00000 ymax = 100.00000 xmin =   0.00000 xmax =  42.00000
Valeurs propres: lambda 1 =       0.08926  lambda 2 =       0.00550

b0 =  13.17967    b1 =   2.08690    angle (°) =  64.39718

I.C. 95% de la pente = [   1.35993,    3.12944]
I.C. 95% de l'ordonnée à l'origine
                     = [ -18.47574,   35.25317]   (sous-estimation)

-----------------------------------------------------------------

b0=ord. à l'origine, b1=pente, xbar=    30.36364, ybar=    76.54545

Code 999.99999 si la pente est infinie (angle = 90°)

Code   0.00000 si les bornes de l'I.C. ne peuvent être
calculées; quand les deux valeurs propres sont trop
semblables, l'I.C. incorpore les 360° du plan.


----------------------------------------
Test par permutation des pentes et de r
----------------------------------------

Nombre de permutations aléatoires =   999

-------------------------------------------------
Méthode   Stat.    PP    EG    PG  p unilatérale
-------------------------------------------------

AM      2.30173   998     1     1     0.00200

MCO     1.86996   998     1     1     0.00200

Corr    0.88232   998     1     1     0.00200

AMDC    2.08690   998     1     1     0.00200
-------------------------------------------------
```

***Exemple 4***

Fichier de données

Mesplé *et al.* (1996) ont généré une variable $X$ contenant 100 valeurs tirées au hasard d'une distribution uniforme dans l'intervalle [0, 10]. Ils ont ensuite généré deux autres variables, $N_1$ et $N_2$, contenant des valeurs tirées au hasard d'une distribution normale $N(0, 1)$. Ces variables furent combinées pour créer deux nouvelles variables $x = (X + N_1)$ et $y = (X + N_2)$. La relation ainsi fabriquée entre $x$ et $y$ est parfaite et devrait avoir une pente de 1, malgré le fait qu'il y a de l'erreur normale ajoutée indépendamment à $x$ et à $y$.

Fichier de résultats

Les différentes méthodes de régression de modèle II ont donné les résultats qui suivent (en-tête éliminé):

```
-----------------------------------------------------

Fichier de données: E4_100x2.txt

-------------------------------------------------------------------

Axe majeur (AM,"MA"):
Valeurs propres: lambda 1 =      17.56697  lambda 2 =       0.95341

b0 =   0.29059    b1 =   0.96029    angle (°) =  43.83962

I.C. 95% de la pente = [   0.86957,    1.06006]
I.C. 95% de l'ordonnée à l'origine
                     = [  -0.21218,    0.74777]   (sous-estimation)

-------------------------------------------------------------------

Axe majeur réduit (AMR,"SMA"):

b0 =   0.27034    b1 =   0.96431    angle (°) =  43.95915

I.C. 95% de la pente = [   0.88261,    1.05358]
I.C. 95% de l'ordonnée à l'origine
                     = [  -0.17951,    0.68207]   (sous-estimation)

I.C. de la pente selon Jolicoeur & Mosimann (1968),  McArdle (1988)
```

```
-----------------------------------------------------------------

Moindres Carrés Ordinaires (MCO,"OLS"):
          r =   0.89690   coeff. de détermination (r^2) =   0.80443

b0 =   0.77135    b1 =   0.86489    angle (°) =  40.85618

I.C. 95% de la pente = [   0.77940,    0.95038]
I.C. 95% de l'ordonnée à l'origine
                     = [   0.26636,    1.27633]

-----------------------------------------------------------------

Axe majeur des données cadrées (AMDC,"RMA"):
ymin =  -1.51400 ymax =  10.58600 xmin =  -0.81100 xmax =  10.78200
Valeurs propres: lambda 1 =       0.12558  lambda 2 =       0.00678

b0 =  -0.46075    b1 =   0.95560    angle (°) =  43.69937

I.C. 95% de la pente = [   0.86511,    1.05464]
I.C. 95% de l'ordonnée à l'origine
                     = [  -0.18483,    0.77021]   (sous-estimation)

-----------------------------------------------------------------

b0=ord. à l'origine, b1=pente, xbar=     5.03918, ybar=     5.12968

Code 999.99999 si la pente est infinie (angle = 90°)

Code   0.00000 si les bornes de l'I.C. ne peuvent être
calculées; quand les deux valeurs propres sont trop
semblables, l'I.C. incorpore les 360° du plan.


----------------------------------------
Test par permutation des pentes et de r
----------------------------------------

Nombre de permutations aléatoires =   999

-------------------------------------------------
Méthode   Stat.    PP    EG    PG  p unilatérale
-------------------------------------------------

AM      0.96029   999     1     0     0.00100

MCO     0.86489   999     1     0     0.00100

Corr    0.89690   999     1     0     0.00100

AMDC    0.95560   999     1     0     0.00100
-------------------------------------------------
```

L'aspect le plus intéressant de ces résultats est qu'en régression MCO, l'intervalle de confiance de la pente ne contient pas la valeur 1 et l'intervalle de confiance de l'ordonnée à l'origine ne contient pas la valeur 0. La pente obtenue par MCO sous-estime la pente réelle de la relation qui est 1 par construction dans cet exemple. Cela illustre le fait que MCO, considérée comme méthode de régression de modèle I, est une méthode inadéquate pour estimer la pente de la relation fonctionnelle liant ces deux variables. En tant que méthode de régression de modèle II, MCO ne serait applicable que si l'on désire prédire les valeurs $\hat{y}$ à partir de $x$ (point 6 du tableau 1).

Avec les autres méthodes de régression de modèle II, au contraire, l'intervalle de confiance de la pente contient la valeur 1 et l'intervalle de confiance de l'ordonnée à l'origine contient la valeur 0, comme on peut s'y attendre pour ce jeu de données lorsque l'on considère leur mode de génération.

***Exemple 5***

Fichier de données

Deux vecteurs de 100 données aléatoires, tirées d'une normale (0, 1), ont été générés. On s'attend à trouver une corrélation nulle. Ces données ont été soumises au programme de régression de modèle II.

Fichier de résultats: Équations de régression AM, AMR, MCO et AMDC, intervalles de confiance, tests de signification (en-tête éliminé). La figure 5 présente le diagramme de dispersion. Les différentes droites de régression sont présentées pour permettre leur comparaison.

```
-----------------------------------------------------

Fichier de données: E5_100x2.txt

-------------------------------------------------------------------

Axe majeur (AM,"MA"):
Valeurs propres: lambda 1 =       1.07132  lambda 2 =       0.88571

b0 =   0.11293    b1 =  -0.60005    angle (°) = -30.96569

I.C. 95% de la pente = [   0.00000,    0.00000]
I.C. 95% de l'ordonnée à l'origine
                     = [   0.00000,    0.00000]   (sous-estimation)
```

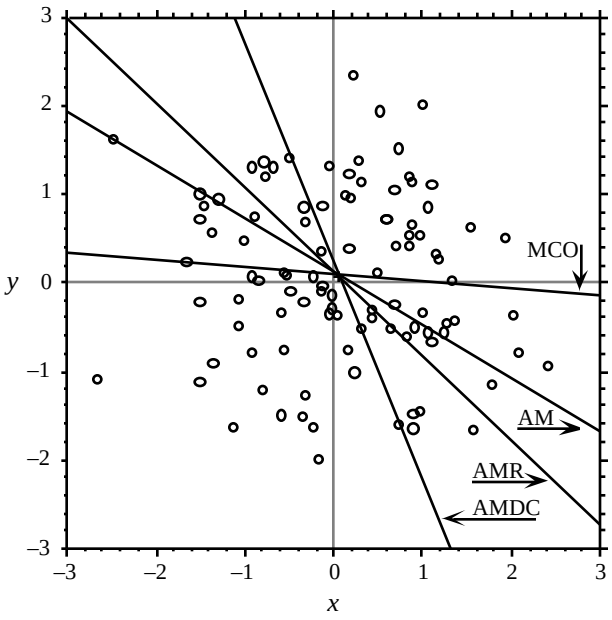


**Figure 5** Diagramme de dispersion des données de l'exemple 5 (nombres aléatoires) montrant les droites de régression de l'axe majeur (AM), l'axe majeur réduit (AMR), les moindres carrés ordinaires (MCO) et l'axe majeur des données cadrées (AMDC). Le coefficient de corrélation n'est pas significativement différent de zéro. La croix indique le centroïde du nuage de points. Les quatre droites de régression passent par ce centroïde.

```
------------------------------------------------------------------------

Axe majeur réduit (AMR,"SMA"):

b0 =   0.14184    b1 =  -0.95633    angle (°) = -43.72118

I.C. 95% de la pente = [  -1.16625,   -0.78419]
I.C. 95% de l'ordonnée à l'origine
                     = [   0.12788,    0.15888]   (sous-estimation)

I.C. de la pente selon Jolicoeur & Mosimann (1968),  McArdle (1988)
------------------------------------------------------------------------

Moindres Carrés Ordinaires (MCO,"OLS"):
          r =  -0.08377   coeff. de détermination (r^2) =   0.00702

b0 =   0.07074    b1 =  -0.08011    angle (°) =  -4.58017

I.C. 95% de la pente = [  -0.27114,    0.11092]
I.C. 95% de l'ordonnée à l'origine
                     = [  -0.12205,    0.26354]

------------------------------------------------------------------------

Axe majeur des données cadrées (AMDC,"RMA"):
ymin =  -1.98676 ymax =   2.35266 xmin =  -2.66496 xmax =   2.39702
Valeurs propres: lambda 1 =       0.05091  lambda 2 =       0.03863

b0 =   0.26978    b1 =  -2.53297    angle (°) = -68.45619

I.C. 95% de la pente = [   1.63300,   -0.39805]
I.C. 95% de l'ordonnée à l'origine
                     = [   0.09654,   -0.06827]   (sous-estimation)
```

```
-----------------------------------------------------------------

b0=ord. à l'origine, b1=pente, xbar=     0.08114, ybar=     0.06424

Code 999.99999 si la pente est infinie (angle = 90°)

Code   0.00000 si les bornes de l'I.C. ne peuvent être
calculées; quand les deux valeurs propres sont trop
semblables, l'I.C. incorpore les 360° du plan.


---------------------------------------
Test par permutation des pentes et de r
---------------------------------------

Nombre de permutations aléatoires =   999

-------------------------------------------------
Méthode   Stat.    PP    EG    PG  p unilatérale
-------------------------------------------------

AM     -0.60005   215     1   784     0.21600

MCO    -0.08011   215     1   784     0.21600

Corr   -0.08377   215     1   784     0.21600

AMDC   -2.53297   284     1   715     0.28500
-------------------------------------------------
```

Ni la corrélation ni aucune des pentes n'est significative, comme on pouvait s'y attendre à cause du mode de construction des données. On remarque que l'estimation de la pente diffère énormément d'une méthode à l'autre. L'axe majeur a une pente $b_{\mathrm{AM}} = -0.60005$ mais son intervalle de confiance, noté [0.00000, 0.00000], couvre les 360° du plan, comme l'indique la note sous le tableau des régressions. L'AMDC a une pente estimée $b_{\mathrm{AMDC}} = -2.53297$. La méthode MCO, qui ne devrait être employée que pour prédire les valeurs $\hat{y}$ à partir de $x$ (point 6 au tableau 1), a tendance à produire des pentes près de zéro pour des données aléatoires: $b_{\mathrm{MCO}} = -0.08011$.

Puisque la corrélation n'est pas significative, l'AMR n'aurait pas dû être calculé. Cette méthode a tendance à produire des pentes près de ±1; pour le présent exemple, la pente est effectivement près de ±1 ($b_{\mathrm{AMR}} = -0.95633$) puisque les écarts types des deux variables sont à peu près égaux ($s_x = 1.01103$, $s_y = 0.96688$). Cet exemple montre que AMDC ne produit pas nécessairement une pente voisine de l'estimation AMR.

Les bornes de l'intervalle de confiance de la pente et de l'ordonnée à l'origine de AMDC fournissent un exemple du phénomène d'inversion des bornes d'un intervalle de confiance, phénomène illustré à la figure 1.

### *Distribution du programme*

Un programme écrit par P. Legendre est distribué à partir de notre site WWWeb. On y trouve le fichier-source FORTRAN, la documentation en français et en anglais, des fichiers de données pour des essais, de même que les programmes exécutables pour MacOS (68k ou PowerPC) et DOS 32-bits (approprié pour des sessions DOS sous Windows 95/98/NT). L'adresse WWWeb est la suivante: http://www.fas.umontreal.ca/biol/legendre/.

### *Références*

Anderson, M. J. & P. Legendre. 1999. An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *Journal of Statistical Computation and Simulation* **62**: 271-303.

Hines, A. H., R. B. Whitlatch, S. F. Thrush, J. E. Hewitt, V. J. Cummings, P. K. Dayton & P. Legendre. 1997. Nonlinear foraging response of a large marine predator to benthic prey: eagle ray pits and bivalves in a New Zealand sandflat. *Journal of Experimental Marine Biology and Ecology* **216**: 191-210.

Jolicoeur, P. 1990. Bivariate allometry: interval estimation of the slopes of the ordinary and standardized normal major axes and structural relationship. *Journal of Theoretical Biology* **144**: 275-285.

Jolicoeur, P. & J. E. Mosimann. 1968. Intervalles de confiance pour la pente de l'axe majeur d'une distribution normale bidimensionnelle. *Biométrie-Praximétrie* **9**: 121-140.

Legendre, P. & L. Legendre. 1998. *Numerical ecology. 2nd English edition.* Elsevier Science BV, Amsterdam.

McArdle, B. 1988. The structural relationship: regression in biology. *Canadian Journal of Zoology* **66**: 2329-2339.

Mesplé, F., M. Troussellier, C. Casellas & P. Legendre. 1996. Evaluation of simple statistical criteria to qualify a simulation. *Ecological Modelling* **88**: 9-18.

Neter, J., M. H. Kutner, C. J. Nachtsheim & W. Wasserman. 1996. *Applied liear statistical models. 4th Edition.* Richard D. Irwin Inc., Chicago.

Sokal, R. R. & F. J. Rohlf. 1995. *Biometry – The principles and practice of statistics in biological research. 3rd edition.* W. H. Freeman, New York.

### *Citation du programme*

Les utilisateurs de ce programme peuvent y référer en citant le présent manuel:

Legendre, P. 2001. Régression de modèle II – Guide. Département de sciences biologiques, Université de Montréal. 24 p.
Distribué à partir du site WWWeb <http://www.fas.umontreal.ca/biol/legendre/>.