

Présentation d'un programme Basic d'analyse log-probit pour micro-ordinateur

Michel RAYMOND ⁽¹⁾

Résumé

L'analyse statistique des tests insecticides fait appel à des calculs log-probit qu'il est difficile de traiter manuellement. Nous

Mots-clés : Probit — Informatique — Test biologique — Statistique.

Summary

LOG-PROBIT ANALYSIS BASIC PROGRAM FOR MICROCOMPUTER. *The statistical analysis of bioassays necessitates log-probit calculations which are too strenuous to compute by hand. We have developed a set of Basic microsoft IV programs running on a microcomputer Tandy (Radio shack) TRS-80, model III or IV, that can easily be adapted to other computers. The main steps of the principal program are described as well as some of the results it provides.*

Key words : Probit — Informatics — Bioassay — Statistics.

L'analyse statistique des tests insecticides, ou plus généralement de tous les tests suivant la loi du « tout ou rien » (réponse quantale), a fait l'objet d'une étude complète dans l'ouvrage de Finney (1971). Elle implique toutefois une masse de calculs fastidieux qu'il est possible en revanche de réaliser rapidement grâce à un traitement informatique. Nous présentons ici les principales étapes d'un programme qui effectue rapidement la plupart des calculs de l'analyse log-probit et nous donnons quelques exemples de ses capacités.

Après la mise en fichier des résultats bruts des

tests insecticides sous la forme d'un tableau (tabl. I) comprenant le nombre de morts observés à chaque concentration, le nombre total d'individus testés et la dose d'insecticide correspondante, le programme principal calcule :

— la mortalité relative à chaque concentration, corrigée en fonction de la mortalité observée dans les témoins, selon la formule d'Abbott (1925),

$$M = \frac{M^* - C}{1 - C}$$

(1) Laboratoire de génétique, Institut des Sciences de l'Évolution, U.S.T.L., Place E. Bataillon, 34060 Montpellier, France.

où M^* est la mortalité observée à la concentration considérée, C la mortalité observée dans le lot témoin et M la mortalité corrigée ;

— la meilleure droite de régression pondérée de la mortalité sur les doses. Ce calcul comprend dans l'ordre :

- la transformation en probit de chaque mortalité,
- la transformation éventuelle de la dose en unité logarithmique arbitraire,
- le calcul de la droite de régression non pondérée,
- le calcul de la meilleure droite de régression pondérée par itérations successives.

Ainsi, à partir des données présentées au tableau I, les caractéristiques des différentes droites obtenues lors de l'itération sont :

Droite de régression non pondérée :

$$Y = 4,89681 + 8,12961 * (X - 7,20167)$$

TABLEAU I

Résultat d'un test insecticide : une souche du moustique *Culex pipiens* L. a été testée pour sa sensibilité au chlorpyrifos (Dursban). Pour chaque dose d'insecticide on relève le nombre de morts et le total d'individus qui ont été soumis à cette dose.

MORTE	TOTAL	DOSE (ppm)
1	80	0,0010
4	80	0,0012
23	120	0,0014
49	140	0,0016
72	120	0,0018
109	160	0,0020
85	100	0,0025
120	120	0,0030
20	20	0,0100

Droite de régression pondérée :

$$Y = 4,88604 + 8,91393 * (X - 7,20258)$$

— une fois la droite calculée, les mortalités observées sont comparés aux mortalités attendues selon un test du χ^2 , suivant la formule (Finney, 1971, p. 66) :

$$\chi^2 = \frac{N}{\sum_{j=1}^N} \frac{(r_j - n_j * P_j)^2}{n_j * P_j * (1 - P_j)}$$

où N est le nombre de concentrations testées et où, pour le j^e point, r_j est le nombre de morts observés parmi les n_j individus testés, et P_j la mortalité attendue. Cette comparaison permet de vérifier que la droite obtenue représente correctement les données.

Si c'est le cas, comme dans l'exemple traité au tableau I ($\chi^2 = 9,9756$ pour 6 ddl, $P > 0,12$), la DL_{50} ainsi que ses limites de confiance, et, éventuellement, d'autres DL (tabl. II) sont calculées d'après la formule exacte (Finney, 1971, p. 79) :

$$DLx = m + \frac{g * (m - \bar{x})}{b * (1 - g)} \pm \frac{t}{b * (1 - g)} *$$

$$\sqrt{\left(\frac{1 - g}{S_{nw}} + \frac{(m - \bar{x})^2}{S_{xx}} \right) * H},$$

$$avcc \ g = \frac{t^2}{b^2 * S_{xx} * H},$$

où b est la pente de la droite, H le facteur d'hétérogénéité (voir ci-dessous), \bar{x} la moyenne pondérée des abscisses, S_{nw} la somme des poids, S_{xx} la variance pondérée des abscisses, t la valeur de la loi de Student suivant le niveau de confiance choisi et le nom-

TABLEAU II

Doses létales et limites de confiance : pour chaque niveau de mortalité choisie — ici 50, 80 et 90 — la dose létale (DL) correspondante est calculée ainsi que ses limites de confiance (le niveau de confiance désiré peut être choisi pour chaque DL). Les doses sont en ppm.

DL	NIVEAU DE CONFIANCE CHOISIE	LIMITES DE CONFIANCE DE LA DL.
DL 50 = 0,00177	0,95	0,00172-0,00181
DL 90 = 0,00245	0,95	0,00235-0,00258
DL 50 = 0,00177	0,98	0,00171-0,00182
DL 50 = 0,00177	0,999	0,00169-0,00185
DL 80 = 0,00219	0,95	0,00212-0,00228

bre de degrés de liberté ($ddl = \infty$ si $H = 1$) et m l'abscisse correspondante à x % de mortalité.

Le calcul de la variance des points de la droite et des limites de confiance des mortalités attendues (Finney, 1971, p. 76) permet de voir si les points expérimentaux sont bien compatibles avec la droite, avec un degré de confiance choisi (tabl. III).

TABLEAU III

Test de compatibilité des points vis à vis de la droite : pour chaque point de la droite, le calcul de la variance permet de déterminer une valeur inférieure (MINI) et supérieure (MAX), d'après le niveau de confiance accordé (ici 98 %). Si la mortalité observée (MORT.) n'est pas contenue dans ces limites, le test rejette le point. Les points numéro 5 et 7 peuvent être rejetés, avec au plus deux chances sur 100 de se tromper pour chacun d'eux. Les points extrêmes (0 et 100 % de mortalité) ne sont pas pris en compte pour ce test.

N	DOSE	MINI	MORT.	MAX	TEST
1	0,0010	0,48949	0,5	2,96754	OUI
2	0,0012	3,81396	4,3	10,175	OUI
3	0,0014	13,6753	18,6	23,0663	OUI
4	0,0016	30,0109	34,5	39,8518	OUI
5	0,0018	48,0594	59,7	57,6854	NON
6	0,0020	63,5911	67,9	73,4482	OUI
7	0,0025	87,1717	84,9	94,4627	NON

Si le test du χ^2 rejette l'hypothèse de la bonne représentation des données par la droite de régression, on peut, soit réaliser un regroupement des valeurs observées et attendues des mortalités des points extrêmes afin de diminuer leur contribution au χ^2 (Finney, 1971, p. 73), soit calculer l'indice d'hétérogénéité, $H = \chi^2/ddl$ (Finney, 1971, p. 70-72), qui interviendra alors dans les calculs des DL et de l'intervalle de confiance des mortalités. La valeur

du χ^2 correspondant à la droite de régression obtenue lors du test insecticide présenté au tableau IV est très élevée ($\chi^2 = 944$ pour 2 ddl, $P < 0,0001$). Le point numéro 1 contribue fortement au χ^2 du fait de la très faible valeur attendue du nombre de morts. Le regroupement de ce point avec le point suivant (signalé par la lettre R, tabl. V) rend le χ^2 égal à 2,278 pour 1 ddl ; on ne rejette plus l'hypothèse que la droite représente bien les données.

Un sous-programme testant le parallélisme de plusieurs droites de régression est associé au programme principal ; il permet de comparer la variance relative (la pente) d'un nombre quelconque de droites de régression (seule la capacité mémoire de l'ordinateur est un facteur limitant) selon les formules de Finney (1971, p. 100-124). Enfin, deux programmes de tracé de courbes et de droites de régression nécessitant une table traçante (TRS-80, CGP-115 ou Multi-Pen Plotter) sont aussi disponibles. La figure 1 montre un exemple de tracé sur la Multi-Pen Plotter.

Les programmes effectuent toutes les opérations de manière automatique et conversationnelle. Le regroupement des différentes options en menu (tabl. VI) permet de conduire l'analyse complète d'un ensemble de données suivant chaque cas particulier. Le menu principal (menu 1) permet d'appeler le sous-programme testant le parallélisme et les programmes annexes (programme de mise en fichier et programmes pour les tables traçantes), d'atteindre les fichiers de données déjà stockés et d'en initier l'analyse. Après le calcul de la meilleure droite de régression pondérée, le menu 2 permet de réaliser ou non tous les calculs décrits plus haut suivant les caractéristiques des données (homogènes ou non) et les desiderata de l'expérimentateur (par exemple calcul de la DL 99 au niveau de confiance 1 %).

Les programmes sont écrits en Basic microsoft niveau IV et sont conçus pour fonctionner sur un

TABLEAU IV

Contribution au χ^2 de chaque point. Résultat de l'analyse d'un test de sensibilité au chlorpyrifos sur la souche S5 de *Culex pipiens* L. Les données proviennent de Pasteur et Sinègre (1978). La faible valeur attendue du premier point explique sa forte contribution au χ^2 total. $P(\chi^2 < 943,776 ; ddl = 2) = 1$.

DOSE (ppm)	MORTALITE	PROBIT	TOTAL	MORTS		CONTRIBUTION AU χ^2
				OBSERVES	ATTENDUS	
0,003	1	2,67321	100	1	0,00105	940,552
0,006	6	3,4449	100	6	10,558	2,20997
0,012	98	7,05419	100	98	96,039	1,01469
0,024	100	7,57624	100	100	99,99	9,9474E-5

TABLEAU V

Contribution au χ^2 de chaque point après regroupement des deux premiers points du tableau IV. Deuxième analyse des résultats de Pasteur et Sinègre (1978) sur *Culex pipiens* L. Après le regroupement des deux premiers points le χ^2 n'est plus significatif : on ne rejette plus l'hypothèse que la droite représente bien les données. $P(\chi^2 < 2, 2777 ; \text{ddl} = 1) = 0,8687$.

DOSE ($\mu\text{g/l}$)	MORTALITE	PROBIT	TOTAL	MORTS		CONTRIBUTION AU χ^2
				OBSERVES	ATTENDUS	
0,003	1	2,67321	100	1	0,00105	0 (R)
0,006	6	3,4449	100	6	10,558	1,27421 (R)
0,012	98	7,05419	100	98	96,039	1,01469
0,024	100	7,57624	100	100	99,99	9,9474E-5

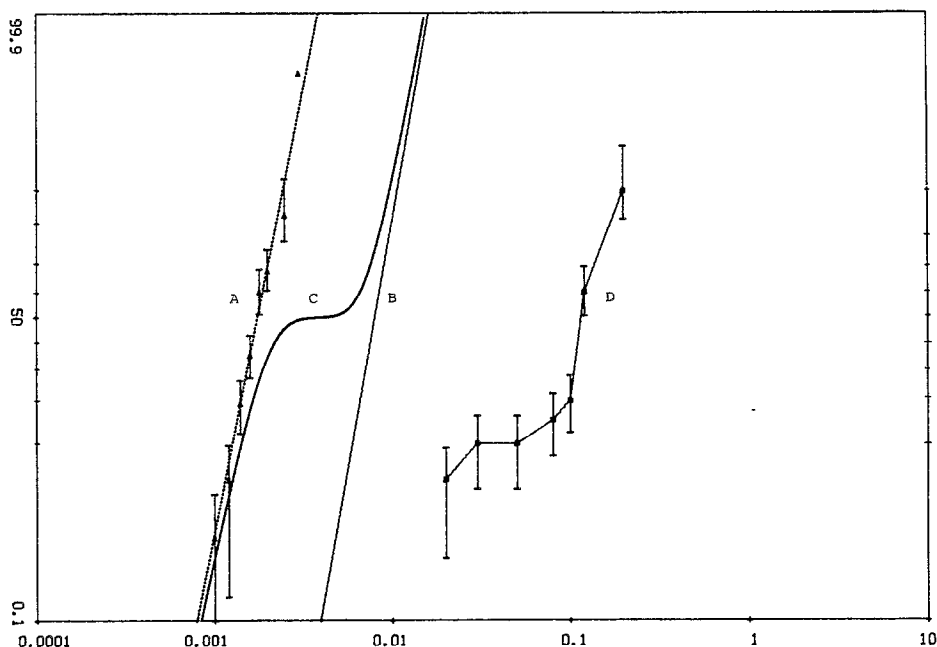


FIG. 1. — Exemple de représentation de divers tests obtenus grâce à la table traçante TRS-80 Multi-Pen Plotter. On peut représenter de façon indépendante les points correspondant aux résultats bruts, reliés ou non entre eux par un trait, la droite de régression pondérée, et l'intervalle de confiance associé à chaque point. Les ordonnées sont en pourcentage de mortalité et pour cet exemple, les abscisses sont en ppm de chlorpyrifos. A : Données correspondant au tableau I et droite de régression pondérée en pointillé. Les intervalles de confiance au niveau 95 % sont représentés pour chaque point. B : Droite de régression correspondant aux données du tableau IV (d'après Pasteur et Sinègre, 1978). C : Données d'un test au chlorpyrifos sur une souche hétérogène de *Culex pipiens*. L'option de la représentation des points reliés entre eux a été choisie. D : Courbe théorique de mortalité d'une population composée de x % de la souche représentée en A et de 100 - x % de la souche S5 représentée en B. Pour cet exemple, x = 50 %.

TRS-80, modèle III ou IV avec disquette, mais peuvent être adaptés aisément à n'importe quel micro-ordinateur moderne.

Le listing de ces programmes, une disquette contenant les programmes et d'éventuelles explications supplémentaires pourront être demandés à l'auteur.

REMERCIEMENTS :

Je tiens à remercier M^{me} Nicole Pasteur et M. Henri Croset pour leur aide et leurs précieux conseils.

Manuscrit accepté par le Comité de Rédaction le 12-6-1985.

TABLEAU VI

Détail des menus 1 et 2 du programme principal. L'option 1 initie le calcul de base de l'analyse (transformation logarithmique des doses, correction des mortalités, transformation probit et calcul de la meilleure droite de régression pondérée), puis renvoie au menu 2 pour une analyse complète « à la carte ». L'option 2 (menu 1) enregistre toutes les données analysées sous forme de fichier, rééditable à tout moment par l'option 3. Les options 5, 6 et 7 servent à appeler le sous-programme testant le parallélisme et les programmes annexes de création de fichier et de tracé de courbes.

MENU 1	MENU 2
ANALYSE COMPLETE D'UN FICHIER..... 1	IMPRESSION DU TABLEAU GENERAL
MEMORISATION DES ANALYSES..... 2	D'ANALYSE..... 1
LECTURE DES DONNEES ANALYSEES..... 3	REGROUPEMENT DE VALEURS POUR LE
MENU 2..... 4	CALCUL DU χ^2 2
TEST PARALLELISME..... 5	CALCUL DE L'HETEROGENEITE..... 3
CREATION, MODIFICATION DE FICHIER.... 6	VARIANCE DES PARAMETRES DE LA DROITE. 4
TRACE DE COURBES..... 7	SUPPRESSION DE L'HETEROGENEITE..... 5
	CALCUL DES DL..... 6
	TEST DES POINTS..... 7
	MENU 1..... 8

BIBLIOGRAPHIE

ABBOTT (W. S.), 1925. — A method of computing the effectiveness of an insecticide. *J. Econ. Ent.*, 18 : 265-267.
 FINNEY (D. J.), 1971. — Probit analysis. Third edition, Cambridge University Press, 333 p.

PASTEUR (N.) et SINÈGRE (G.), 1978. — Chlorpyrifos (Dursban) resistance in *Culex pipiens* L. from Southern France : inheritance and linkage. *Experientia*, 34 : 709-710.