Journal de la société statistique de Paris

E. Morice

Méthodes d'analyse des observations par « tout ou rien »

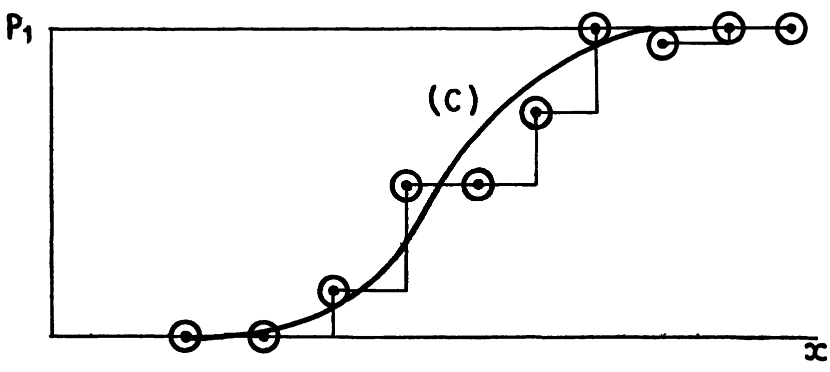
Journal de la société statistique de Paris, tome 100 (1959), p. 121-133
<http://www.numdam.org/item?id=JSFS_1959__100__121_0>

Méthodes d'analyse des observations par « tout ou rien »

I — *Method des* « *Probit* » (Finney et Bliss)

La méthode d'analyse des « probit », étudiée a peu près simultanément par Bliss et Finney (1) est appliquée dans des problèmes du type général suivant (problèmes de « tout ou rien »).

Les individus d'une population (d'un échantillon) sont soumis à un certain stimulus qui chez chacun d'eux, suivant la dose, provoque ou ne provoque pas de réaction à laquelle on s'intéresse : l'individu meurt ou ne meurt pas, un détonateur explose ou n'explose pas.



Graphique 1.

Les observations faites dans l'expérimentation donnent la fréquence des individus de l'échantillon chez lequels cette réaction a été observée.

---

(1) FINNEY. *Probit Analysis : a statistical treatment of the Sigmoid Response Curve.* Cambridge University Press 1947.

f est une estimation de la fréquence p des individus chez lesquels on noterait cette réaction dans la population.

Ceci revient à admettre que si x_p est la dose qui provoque la réaction chez une fraction p de la population, cette fréquence p représente la fraction de la population chez laquelle cette réaction peut être provoquée par une dose égale ou inférieure à x_p.

L'observation de p en fonction de x_p donne donc une estimation par points de la courbe de répartition de p en fonctiou de x_p, image plus ou moins précise de la vraie courbe de répartition dans la population, que l'on admet être une fonction monotone, continue $p\ (x_p)$, en général croissante dans la plupart des applications.

La connaissance de l'effet du stimulus est donc caractérisée par celle de cette courbe de répartition $p(x_p)$ qu'il s'agit d'ajuster au mieux aux observations $(x,\ f_x)$.

Théoriquement, le problème est donc un problème classique d'ajustement qui ne présenterait pas de difficultés particulières si l'on disposait d'un nombre suffisant d'observations.

Dans la réalité il s'agit souvent d'expériences coûteuses, destructrices (mort par effet d'un poison, explosion d'un détonateur en fonction d'un élément d'action : hauteur de chute, intensité d'un courant, ...).

L'allure du diagramme en escalier fourni par les observations, fort imprécise en raison du petit nombre des observations, de variations aléatoires fort importantes suivant les individus observés, rend cet ajustement difficile.

On peut légitimement admettre à priori que la courbe de répartition cherchée est une courbe sigmoïde (courbe en S) admettant deux asymptotes horizontales : effet nul (ou effet certain) pour des valeurs petites (ou grandes) de la variable x, avec, pour des raisons de continuité : un seul point d'inflexion, c'est-à-dire une courbe de répartition du type de celle qui correspond à une distribution continue unimodale, ou de la forme d'une courbe logistique.

L'assimilation, à priori, à la courbe de répartition d'une loi normale peut être assez arbitraire, mais un changement de variable convenable peut permettre d'essayer de se ramener à ce type (par exemple, $u = \log x$).

La méthode des « probits » est une solution de ce problème, basée sur ce point de vue qui tient compte des larges possibilités d'emploi des tables de la loi normale.

Notons en passant qu'une autre méthode basée sur l'emploi de la courbe logistique a été étudiée par Fisher et Yates et par Berkson.

Si l'on envisage une relation logistique :

$$p = [1 + e^{-\alpha-\beta x}]^{-1}$$

comme représentant approximativement la dépendance de p en fonction de x, soit encore :

$$\frac{p}{1-p} = e^{\alpha+\beta x},$$

Berkson (1944) a définit le « logit » de P comme étant :

$$Y = \log_e \frac{p}{1-p} = \alpha + \beta x$$

ce qui conduit à l'étude d'une régression linéaire de Y en fonction de x.

Le but de l'étude est d'estimer les paramètres caractéristiques de la répartition des effets du stimulus selon la dose employée, par exemple de déterminer la « dose effective médiane », c'est-à-dire celle qui provoque la réaction en moyenne dans 50 % des cas.

L'emploi de la méthode des probits suppose que la mesure de l'intensité x du stimulus est faite sur une échelle telle que les tolérances individuelles puissent être considérées comme normalement distribuées, x est donc une variable normale (m, σ) à laquelle correspond une variable réduite.

$$\text{(1)} \qquad X = \frac{x-m}{\sigma}$$

et afin d'éviter pratiquement l'emploi de nombres négatifs, un « probit » :

$$\text{(2)} \qquad Y = X + 5$$

La probabilité P d'observer la réaction considérée sur un individu choisi au hasard sera :

$$\text{avec} \begin{cases} \text{(3)} \quad P = \int_{-\infty}^{X} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \, du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Y-5} e^{-\frac{u^2}{2}} \, du \\ \\ \text{(4)} \quad Q = 1 - P = \frac{1}{\sqrt{2\pi}} \int_{Y-5}^{\infty} e^{-\frac{u^2}{2}} \, du \end{cases}$$

Considérons une valeur x_i du stimulus ($i = 1, 2 \ldots k$) à laquelle sont soumis n_i individus dont r_i donnent lieu à la réaction envisagée et soient :

$$\text{(5)} \qquad p_i = 1 - q_i = \frac{r_i}{n_i}$$

Les valeurs correspondantes y_i des probits seront les valeurs de Y données par l'équation (3) pour $P = p_i$.

Les points de coordonnées $(x_i\, y_1)$, si l'hypothèse de normalité est acceptable, devraient, aux fluctuations aléatoires près, être situés sur la droite :

$$\text{(6)} \qquad Y = 5 + \frac{x-m}{\sigma}$$

Le problème fondamental est d'estimer m et σ, c'est donc un simple problème de régression linéaire appliqué à une droite de Henri, mais dans lequel il faut tenir compte du fait que les y_i sont d'inégale précision.

Si donc on pose, de manière générale :

$$y = a + bx$$

(1) Cf. MORICE et CHARTIER. *Analyse statistique*, p. 199.

pour représenter la droite cherchée et si on désigne par $n_i\ \omega_i$ le poids attribué à une valeur y_i, la méthode classique des moindres carrés conduit à

$$(8)\begin{cases} \mathrm{b} = \left(\frac{1}{\sigma}\right) = \frac{\Sigma\, n\, \omega\, (x - \bar{x})\, (y - \bar{y})}{\Sigma\, n\, \omega\, (x - \bar{x})^2} = \frac{\Sigma\, n\, \omega\, \Sigma\, n\, \omega\, xy - \Sigma\, n\, \omega\, x\, \Sigma\, n\, \omega\, y}{\Sigma\, n\, \omega\, \Sigma\, n\, \omega\, x^2 - (\Sigma\, n\, \omega\, x)^2} \\ \mathrm{a} = \left(5 - \frac{m}{\sigma}\right) = \bar{y} - b\, \bar{x} \end{cases}$$

$$\text{avec}\begin{cases} \bar{x} = \frac{\Sigma\, n\, \omega\, x}{\Sigma\, n\, \omega} \\ \bar{y} = \frac{\Sigma\, n\, \omega\, y}{\Sigma\, n\, \omega} \end{cases}$$

La variance d'échantillonnage de y_i étant approximativement $\frac{\mathrm{P\ Q}}{n\ \mathrm{Z}^2}$ (1), avec

$$\mathrm{Z} = \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{x^2}{2}}$$

les poids utilisés seront :

$$(9) \qquad n_i\ \omega_i = \frac{n_i\ \mathrm{Z}_i^2}{\mathrm{P}_i\ \mathrm{Q}_i}$$

Les probits et les poids dépendant des observations, il est nécessaire d'employer une méthode d'itération qui pourra être la suivante :

1°) Des valeurs observées n_i et r_i on déduit p_i et y_i, à l'aide de la table de la loi normale.

2°) On peut tracer le graphique des points $(x_i\, y_i)$ et tracer approximativement une droite d'ajustement fournissant par simple lecture un premier ensemble de probits estimés Y_i' correspondant aux x_i.

3°) La différence entre deux probits y et Y et celle qui existe entre les valeurs correspondantes p et P (ou q et Q) des probabilités totales, étant liées approximativement par :

$$\mathrm{Z}\ (y - \mathrm{Y}) = \mathrm{Q} - q = p - \mathrm{P},$$

On peut à partir des probits estimés Y'_i (« expected probits ») calculer des valeurs corrigées des probits à utiliser dans le calcul (« working probits ») y_i' par l'une ou l'autre des formules :

$$y' = \mathrm{Y}' + \frac{\mathrm{Q}'}{\mathrm{Z}'} - \frac{q}{\mathrm{Z}'}$$

ou

$$y' = \mathrm{Y}' = \frac{\mathrm{P}'}{\mathrm{Z}'} + \frac{p}{\mathrm{Z}'},$$

$$p = 1 - q = \frac{r}{n}$$

$\mathrm{Z}', \frac{\mathrm{Q}'}{\mathrm{Z}'}, \left(\text{ou } \frac{\mathrm{P}'}{\mathrm{Z}'}\right)$ étant des valeurs données par la table de la loi normale pour le probit $y = \mathrm{Y}'$, p (ou q) les valeurs expérimentales données par (5).

Remarquer que pour $p = 0$ (ou $q = 0$) les probits de calcul seront simplement :

$$Y' = \frac{P'}{Z'} \left(\text{ou } Y' + \frac{Q'}{Z'}\right)$$

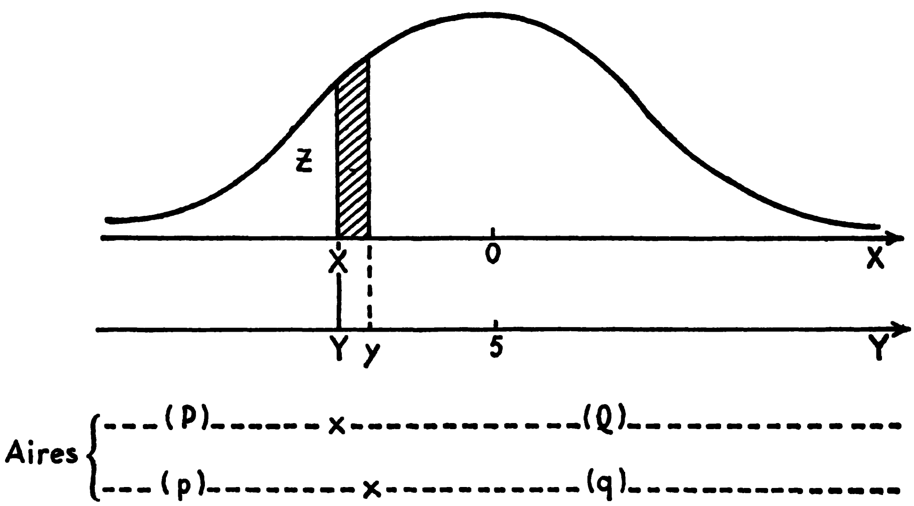
4°) A partir de l'équation (9) ;

$$\omega' = \frac{Z'^2}{P'\,Q'},$$

on pourra calculer les coefficients de pondération $n\ \omega'$

5°) En utilisant les poids $n\ \omega'$ et les probits de calcul y' les équations (8) fourniront une première approximation de a et $\bar{b}$.

6°) A partir de cette fonction linéaire $y = a + bx$, on pourra obtenir une



Graphique 2.

nouvelle estimation y' des probits à partir desquels on pourra recommencer le cycle d'itération, jusqu'à une convergence suffisante, qui ne sera évidemment obtenue que si l'hypotèse préalable de normalité est valable.

Si aucune des fréquences théoriques $n_2\ P_i$ (ou $n_i\ Q_i$) obtenues à partir des probits Y finalement estimés n'est trop petite, on pourra calculer :

$$\chi^2 = \frac{\Sigma^2\,(r_i - n_i\ P_i)^2}{n_i\ P_i\ Q_i}$$

et avoir un test de signification des accords entre observations et théorie, avec un nombre de degrés de liberté

$$\upsilon = k - 2$$

k = nombre de valeurs essayées de x.

Finney (1) a donné une autre méthode de calcul de χ^2, basée sur la somme pondérée des carrés des écarts entre les probits de calcul et la droite de régression.

Noter enfin que la dose médiane effective (« median effective dose ») c'est-à-dire la valeur de x telle qu'elle a 1 chance sur 2 de provoquer la réaction, est estimée par :

$$m = \frac{5 - a}{b} \left(= \text{E. D. } 50 = x_{50}\right) (1)$$

Finney (1) a donné pour l'écart type de x_{50} la formule :

$$\sigma\,(x_{50}) = \frac{1}{b} \left\{ \frac{1}{\Sigma\, n\, \omega} + \frac{(x_{50} - \bar{x})^2}{\Sigma\, n\, \omega\, (x - \bar{x})^2} \right\}^{1/2}$$

Afin de faciliter les calculs de l'analyse ci-dessus, Finney et Stevens, ont calculé des tables qui donnent directement, en fonction de P les valeurs de Y, Y + Q Z, Y — P Z, $\frac{1}{Z}$, et $\frac{Z^2}{P\,Q}$. Ces tables sont reproduites dans Pearson et Hartley « Tables for statisticians (Vol. I. 1956). On trouvera des tables analogues dans Fisher et Yates « Statistical tables for biological, agricultural and medical research ».

Exemple (2) — Une charge d'explosif a été placée à une distance x_i d'une cible de carton ($i = 1 \ldots 5$). L'expérience a été répétée 16 fois ($n_i = \ldots = n_5 = 16$) à chaque distance, et on a noté à chaque expérience si la cible était ou non perforée (r_i nombre de fois où la cible était perforée dans chaque groupe de $n_i = 16$ expériences).

On trouvera ci-après le tableau résumant partiellement les calculs, pour la première estimation Y′ ... et la cinquième itération Y′ ...

Le graphique de 5 points correspondant aux cinq séries, permet, à partir des trois points à distance finie, de tracer une première estimation de la droite de régression à partir de laquelle on lit approximativement les cinq valeurs de Y'_i soit :

i	1	2	3	4	5
Y'_i	4,75	5,02	5,30	5,57	5,85
x_i	0	1	2	3	4

x_i étant lié à la distance réelle par la formule :

$$x_i = \frac{1}{4}\,[53 - \text{distance (en pides)}].$$

Les calculs sont ensuite conduits comme il est indiqué ci-dessus, compte tenu de la simplification résultant de $n_i = \ldots = n_5 = 16$.

Ainsi pour :

1°) $x = 0$ (qui correspond à $y = -\infty$, point qui ne peut évidemment être utilisé dans le tracé) la droite tracée à l'œil donne :

$$Y' = 4{,}75$$

(1) E.D. *Observation anglaise de « effective dose ».*

(2) Extrait de Pearson et Hartley. « *Tables for statisticians* » — vol 1 — Cambridge University Press 1956.

d'où on déduit le probit correspondant :

$$y' = Y' - \frac{P'}{Z'} = 3{,}712$$

donné directement par la table qui donne en même temps la valeur de :

$$\omega' = 0{,}6223 \text{ pour } Y' = 4{,}75$$

2°) $x = 1$: on lit sur la droite Δ :

$$Y' = 5{,}02$$

Pour cette valeur du probit estimé Y', la table donne directement :

$$Y' - \frac{P'}{Z'} \text{ et } \frac{1}{Z'}$$

d'où on déduit, pour $p = 0{,}5625$

$$y' = Y' - \frac{P'}{Z'} + \frac{p}{Z'} = 5{,}157$$

et la valeur de

$$\omega' = 0{,}6365$$

Ayant fait ces calculs pour $x = 0, 1, 2, 3, 4$ on peut alors calculer (8)

$$a = 4{,}017 \qquad b = 0{,}619$$

L'équation :

$$y = 4{,}017 + 0{,}619\, x$$

donnera une nouvelle série de valeurs Y"

$x =$	0	1	2	3	4
$y = Y'' =$	4,02	4,64	5,26	5,87	6,49

à partir desquels on fera une nouvelle série analogue de calculs.

Le tableau des approximations successives montre la convergence des déterminations de a, b et x_{50}, respectivement vers :

3,775 0,757 1,62

La distance effective médiane réelle correspondant à :

$$x_{50} = \frac{5 - 3{,}775}{0{,}757} = 1{,}62$$

est donc :

$$d = 53 - 4\, x_{50} = 46{,}52$$

Tableau des calculs

$(n_1 = n_2 \dots = n_5 = 16)$

Distance X	Distance x_i	r_i	p_i	y_i	Y'_i	y'_i	ω'_i	$\omega'_i\, x_i$	Y''_i
53	0	0	0	$-\infty$	4,75	3,712	0,6223	0	4,02
49	1	9	0,5625	5,16	5,02	5,157	0,6365	0,636	4,64
45	2	9	0,5625	5,16	5,30	5,155	0,6161	1,232	5,26
41	3	12	0,7500	5,67	5,57	5,671	0,5652	1,696	5,87
37	4	16	1,000	∞	5,85	6,561	0,4873	1,949	6,49

x	Y^V_i	y^V_i	ω^V_i	$\omega^V_i\, x_i$	P_i	$16\,P_i$	r_i
0	3,775	3,1896	0,3617	0	0,1103	1,76	0
1	4,532	5,2105	0,5876	0,5876	0,3199	5,12	9
2	5,288	5,1552	0,6177	1,2354	0,6133	9,81	9
3	6,045	5,6036	0,4235	1,2705	0,8520	13,63	12
4	6,802	7,2567	0,1794	0,7176	0,9642	15,43	16

Approximations successives

j	a	b	$x_{50} = \frac{5-a}{b}$: x	$x_{50} = \frac{5-a}{b}$: Pieds
2	4,017	0,619	1,59	46,6
3	3,820	0,731	1,61	46,6
4	3,7791	0,7546	1,618	46,58
5	3,7747	0,7568	1,619	46,52
6	3,7747	0,7569	1,619	46,52

II — MÉTHODE « MINIMUM NORMIT χ^2 ESTIMATE » (BERKSON)

(Estimation par une méthode de χ^2 minimum).

Reprenant une méthode proposée par Taylor (1), Berkson (2) a développé sous le nom de « minimum normit χ^2 estimate » la méthode suivante qui n'exige pas d'itération.

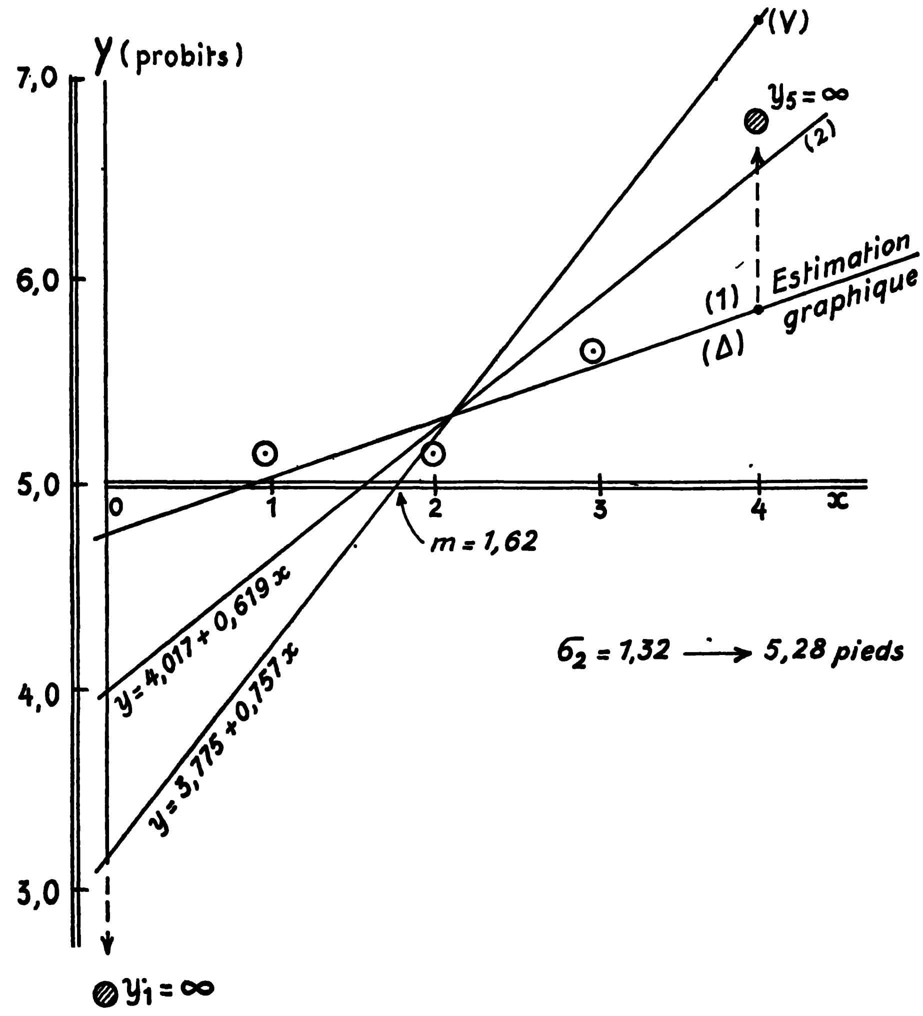
Berkson donne le nom de « normit » à une variable normale réduite : (3).

$$\upsilon = \frac{x - \mu}{\sigma}$$

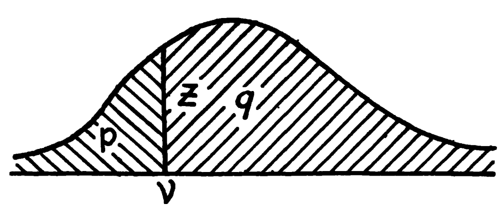
si x est une variable normale et si :

$$p_i = \text{Prob.}\ (x < x_i)$$

(1) TAYLOR — An. Math. Stat. 1953 p. 85.
(2) BERKSON — Journal Am. St. Ass. 1955 p. 529.
(3) Les auteurs anglais emploient aussi l'expression *normal equivalent deviate* au lieu de *normit*.



Graphique 3.



Graphique 4.

On devra avoir

$$\nu_i = \frac{x_i - \mu}{\sigma} = \alpha + \beta x_i$$

$$\alpha = -\frac{\mu}{\sigma} \qquad\qquad \beta = \frac{1}{\sigma}$$

On se propose d'estimer la relation hypothétique

$$\nu = \alpha + \beta x$$

par

$$\nu' = a + bx$$

La quantité, ci-après (normit χ^2) :

$$S = \Sigma\, n_i \frac{Z_i^2}{p_i\, q_i} (\upsilon_i - \upsilon_i')^2$$

est asymptotiquement distribuée comme χ^2 (somme de carrés de variables normales réduites :

$$\left(\frac{\upsilon_i - \upsilon_i'}{Z_i\sqrt{\dfrac{p_i\, q_i}{n_i}}}\right)^2$$

n_i effectif exposé au risque d'intensité x_i

p_i fraction atteinte observée

ν_i « normit » correspondant à p_i

$\nu_i' = a + bx_i$ estimation correspondante

$$Z_i = \frac{1}{\sqrt{2\,\pi}} e^{-\frac{\nu_i'^2}{2}}$$

Les valeurs de a et b, rendant S minimum seront solutions de

$$\frac{\partial S}{\partial a} = \Sigma\, n_i\, \omega_i\, (\nu_i - \nu_i') = o \qquad \nu' = a + bx_i$$

$$\frac{\partial S}{\partial b} = \Sigma\, n_i\, \omega_i\, x_i\, (\upsilon_i - \upsilon_i') = o \qquad \omega_i = \frac{Z_2^2}{p_i\, q_i}$$

d'où

$$\beta \qquad b = \frac{\Sigma\, n_i\, \omega_i\, \upsilon_i\, x_i - \dfrac{\Sigma\, n_i\, \omega_i\, \nu_i\, \Sigma\, n_i\, \omega_i\, \nu_i\, x_i}{\Sigma\, n_i\, \omega_i}}{\Sigma\, n_i\, \omega_i\, x_1^2 - \dfrac{(\Sigma\, n_i\, \omega_i\, x_i)^2}{\Sigma\, n_i\, \omega_i}}$$

$$\alpha \qquad a = \bar{\nu} - b\,\bar{x}$$

$$= \frac{\Sigma\, n_i\, \omega_i\, \nu_i - b\, \Sigma\, n_i\, \omega_i\, x_i}{\Sigma\, n_i\, \omega_i}$$

$$\bar{x} = \frac{\Sigma\, n_i\, \omega_i\, x_i}{\Sigma\, n_i\, \omega_i} \qquad \bar{\nu} = \frac{\Sigma\, n_i\, \omega_i\, \nu_i}{\Sigma\, n_i\, \omega_i}$$

La dose médiane effective étant estimée par :

$$E.\, D_{50} = x_{50} = -\frac{a}{b}$$

L'équation linéaire d'estimation de υ étant

$$\upsilon'_i = a + bx_i = \bar{\upsilon} + b\,(x_i - \bar{x})$$

La précision des 2 estimations est caractérisée par les variances

$$S_v^2 = \frac{1}{\Sigma\, n_i\, \omega_i}$$

$$S_b^2 = \frac{1}{\Sigma\, n\, \omega\, (x - \bar{x})^2}$$

$$S_a^2 = S_{\bar{v}}^2 + \bar{x}^2 S_b^2$$

Dans l'article cité en référence Berkson donne pour faciliter les calculs :

1°) les « normits » υ en fonction de p pour $p = 0$ ... (0,001) ... 1 (υ avec 5 décimales) c'est-à-dire les milliles de la loi normale.

2°) les valeurs de $\omega = \dfrac{Z^2}{p\,q}$ et $\omega\,\upsilon$, avec 4 décimales pour les mêmes valeurs de p.

3°) les « antinormits » c'est-à-dire les valeurs de p en fonction de υ pour

$$\upsilon = 0 \ldots (0{,}001) \ldots 2{,}50 \text{ (p avec 5 décimales)}$$

Pour illustrer sa méthode, Berkson reprend un exemple de Finney dont les éléments essentiels sont reproduits ci-après :

N°	Concentration en mmg/litre D	$D' = \frac{D}{D_1}$	Log D' x	Effectif n	Décès r	$p = \frac{r}{n}$
1	2,6	1	0,000	50	6	0,120
2	3,8	1,46	0,164	48	16	0,333
3	5,1	1,96	0,292	46	24	0,522
4	7,7	2,96	0,471	49	42	0,857
5	10,2	3,92	0,593	50	40	0,880

La méthode Berkson donne :

$$a = -1{,}1345 \pm 0{,}17 \qquad\qquad b = 4{,}1905 \pm 0{,}47$$

soit :

$$\upsilon' = 4{,}1905\, x - 1{,}1345$$

alors que Finney par la méthode des probits trouve $b = 4{,}01$, et que la méthode du maximum de vraisemblance donnerait $b = 4{,}22$.

La méthode proposée par Berkson paraît donc très satisfaisante mais il faut remarquer que pour l'exemple proposé à l'appui de la méthode :

1°) les effectifs soumis à l'expérience sont importants (de l'ordre de 50);

2°) tous les résultats sont directement utilisables : il n'y a pas d'observation

$$p = 0 \quad \text{ou} \quad p = 1$$

(Pour ces valeurs $p = 0$ et $p = 1$ qui ne sont pas directement utilisables dans le calcul de Berkson, ce dernier propose de les remplacer par $\dfrac{1}{2n}$ et $1 - \dfrac{1}{2\,n}$, centres du premier et du dernier intervalle en $\dfrac{1}{n}$).

La méthode de Berkson, appliquée dans ces conditions à l'exemple précédent $n = 16$ avec $p_0 = 0$ et $p_4 = 1$, donnerait $b = 0{,}618$, résultat voisin de celui de la première itération de Finney alors que le résultat des itérations successives de Finney semble converger vers 0,76.

Il semble donc que pour de petits échantillons ne comprenant plus qu'un petit nombre de classes directement utilisables, la méthode de Berkson n'est que très approximativement une méthode de χ^2 minimum et que l'économie de calculs due à l'absence d'itération va à l'encontre de l'efficacité.

Remarques générales diverses

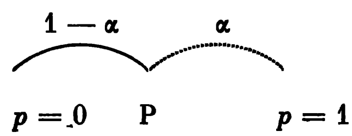
M. Bastien, Ingénieur à l'Institut de Recherches de la Sidérurgie française, ancien élève de l'École d'Application présente, sur la méthode proposée par Berkson, les remarques ci-après :

1°) Le remplacement de $p = 0$ par $\frac{1}{2n}$ $\left(\text{ou } \frac{1}{kn}\right)$ est conventionnel. Il est assez naturel que ce procédé donne des résultats assez satisfaisants pour certaines dispositions des niveaux.

En effet, remplacer $p = 0$ par $\frac{1}{kn}$ revient à se servir de l'extrémité d'un intervalle de confiance de P, à un seuil puisque indépendante de n.

L'extrémité supérieure d'un intervalle de confiance totalement dissymétrique pour P est donné par :

$$(1 - \mathrm{P})^n = \alpha$$



Comme

$$\left(1 - \frac{1}{kn}\right)^n \sim \frac{1}{e^{\frac{1}{k}}}; \text{ on déduit } \alpha \text{ fonction de } k \text{ uniquement.}$$

Pour $k = 2$, $\frac{1}{\sqrt{e}} \sim 0{,}60$. On remplace donc $p = 0$ par l'extrémité supérieure d'un intervalle de confiance (dissymétrique) à 40 %.

Il est à noter qu'un ajustement graphique bien fait utilise aussi d'une façon un peu différente les intervalles de confiance autour des proportions observées, notamment pour les points hors graphique ($p = 0$ ou $p = 1$).

Mais l'emploi de $\frac{1}{2n}$ ou $1 - \frac{1}{2n}$ n'est pas sans inconvénient en raison de son caractère systématique.

Il est évidemment à proscrire dans un certain nombre de cas tels que :

— $n_i = 1$ quel que soit i

— niveau à réponse $p = 0$ ou $p = 1$ très éloigné de l'ensemble des autres.

— niveau à réponse $p = 0$ ou $p = 1$ encadré par des niveaux à réponse $p \neq 0$ ou 1.

D'autre part, Berkson n'envisage pas le cas des n_i petits.

Il semble alors assez aventureux d'utiliser la méthode des moindres carrés avec correction des $p = 0$ ou $p = 1$.

En effet :

— « Normit » p n'est plus guère normalement distribué.

— La présence de nombreux 0 ou 1 amène à des corrections conventionnelles trop nombreuses.

Lorsque les n_i sont grands, il semble assez indifférent d'utiliser une méthode plutôt qu'une autre puisque asymptotiquement elles sont les mêmes.

BIBLIOGRAPHIE

J. Berkson. *A statisticales precise and relatively simple method, of estimating the bio-assay with quantal response, based on the logistic function*, Journal Am. St. Ass. 48 (1953) p. 365-99.

Finney. *Probit Analysis.* Cambridge Univ. Presse 1947.

Berkson. *The minimum normit χ^2 with particular reference to bio-assay.* Journal Am. St. Ass. 50 (1955) p. 329.

E. Morice.

***