Journal of Steroid Biochemistry, Vol. 11, pp. 1173 to 1180
Pergamon Press Ltd 1979. Printed in Great Britain

THE UNDERLYING STRUCTURE OF THE DIRECT LINEAR PLOT WITH APPLICATION TO THE ANALYSIS OF HORMONE–RECEPTOR INTERACTIONS

N. A. C. CRESSIE* and D. D. KEIGHTLEY†
* School of Mathematical Sciences, Flinders University, Bedford Park, and
† Department of Surgery, Flinders Medical Centre, Bedford Park, South Australia, 5042

(*Received* 25 *September* 1978)

SUMMARY

In assays for hormone receptors, the two parameters (appearing in the Michaelis–Menten equation) to be estimated are the total number of binding sites, and the dissociation constant. This has resulted in at least three estimation methods which convert the problem to straight line fitting. The latest of these, known as the direct linear plot, is the subject of this study. Up to now, it has been used in the absence of the appropriate statistical theory, and hence any attempts at providing standard errors of the estimates, or confidence intervals, have been wrong. The direct linear plot is an attractive estimation procedure since it essentially ignores "wild" points, yet gives virtually identical results to classical least squares methods when the points are "well-behaved". It is this robustness that makes its underlying structure worthy of attention.

INTRODUCTION

Most techniques routinely employed to measure tissue levels of hormone receptors, particularly steroid hormone receptors, involve the incubation of various concentrations of labelled and unlabelled hormone with receptor protein. This provides a series of values for the amount of hormone bound at each incubating concentration of the hormone. A graph may then be drawn from these values, and from this graph estimates can be obtained not only of the number of hormone receptor sites, but also of the dissociation constant for the hormone–receptor interaction. Such a graph is described by the Michaelis–Menten equation:

$$B = \frac{B_{\max}F}{K_D + F}, \qquad (1)$$

where B and F are the bound and free levels of hormone at a particular incubating concentration, often expressed in terms of fmol/mg protein, $B_{\max}$ is the total number of binding sites, and K_D is the dissociation constant for the hormone–receptor interaction.

The values for $B_{\max}$ and K_D may be obtained from a plot of bound hormone (B) against free hormone (F) from the observations $(F_1, B_1), (F_2, B_2), \ldots, (F_n, B_n)$.

Two further ways of writing equation (1) are

$$(B/F) = (B_{\max}/K_D) - (B/K_D), \qquad (2)$$

a transformation introduced by Scatchard[8], and

$$(1/B) = (1/B_{\max}) + (K_D/B_{\max})(1/F), \qquad (3)$$

which is attributed to Lineweaver and Burk[7]. Clearly, fitting a straight line to n data points and estimating the slope and intercept of the straight line are tasks better suited to routine laboratory work, than fitting and estimation for the hyperbola given by equation (1). The plot of points $(B_1, B_1/F_1), (B_2, B_2/F_2), \ldots, (B_n, B_n/F_n)$ is known as a *Scatchard plot*, this being the most common plot used in steroid–receptor assay. The plot of points $(1/F_1, 1/B_1), (1/F_2, 1/B_2), \ldots, (1/F_n, 1/B_n)$ is known as a *double reciprocal plot*. Their respective connections with equations (2) and (3) should be obvious.

A plot introduced by Cornish-Bowden and Eisenthal[2], namely the direct linear plot, is a third technique for those who wish to exploit the comparative ease of drawing a straight line. Subsequently Eisenthal and Cornish-Bowden[3] gave statistical implications for the direct linear plot in an enzyme kinetic context, while Woosley and Muldoon[13] did a similar thing for a steroid–protein interaction.

The purpose of this paper is to look critically at these techniques of data representation, to observe that certain implications of the direct linear plot have been missed, and to give solutions to hitherto unanswered statistical problems associated with these plots. To be more specific, consider the data of Table 1 obtained from an estrogen receptor assay of human uterine cytosol fraction. The data is given in the convenient form of B, B/F; its Scatchard plot is presented in Fig. 1. Notice that the observation $(B_{12}, B_{12}/F_{12})$ in the bottom right hand corner looks very much like an outlier: i.e. an observation which is to be given little weight or credence. The usual way to estimate α and β in, $B/F = \alpha + \beta B$ is via the *method of least squares*. This results in, $\alpha = 1.247$, $\beta = -0.022$. So $B_{\max} = \alpha/(-\beta) = 57.98$, $K_D = 1/(-\beta) = 46.43$.

Least squares is an excellent method of estimation when only the dependent variable contains error, and when the error does not result in outlying points. If

Table 1. Bound and Bound/Free values obtained from an estrogen receptor assay of human uterine cytosol fraction; one outlier present. Values for B are expressed as fmol [^{3}H]-estradiol-17β bound/mg cytosol protein, while B/F is the ratio of the bound to free hormone

B	17.7	18.5	28.5	29.4	35.9	39.1	39.6	39.9	40.3	42.2	44.2	66.1
B/F	1.015	1.110	0.684	0.718	0.343	0.162	0.164	0.397	0.237	0.250	0.144	0.232

Table 2. Entries show: estimate $\pm$ $\hat{se}$(estimate), for three different estimation procedures, using the data of Table 1. B_{max} is expressed as fmol/mg cytosol protein, while K_D is in pmol/l

	$B^*_{max} \pm \hat{se}(B^*_{max})$	$K^*_D \pm \hat{se}(K^*_D)$
Least squares on Scatchard plot	57.893 $\pm$ 22.175	46.429 $\pm$ 10.785
Least squares on Scatchard plot; outlier removed	47.069 $\pm$ 6.722	27.270 $\pm$ 2.245
Direct linear plot	47.972 $\pm$ 2.165	32.324 $\pm$ 8.292

Table 3. Bound and Bound/Free values obtained from a second estrogen receptor assay of human uterine cytosol fraction; no outliers present; B and B/F are expressed as in Table 1

B	14.3	14.6	22.3	23.6	28.8	31.5	33.4	34.3	35.7	35.8	37.5	38.1
B/F	0.399	0.408	0.285	0.306	0.167	0.185	0.125	0.128	0.076	0.098	0.081	0.105

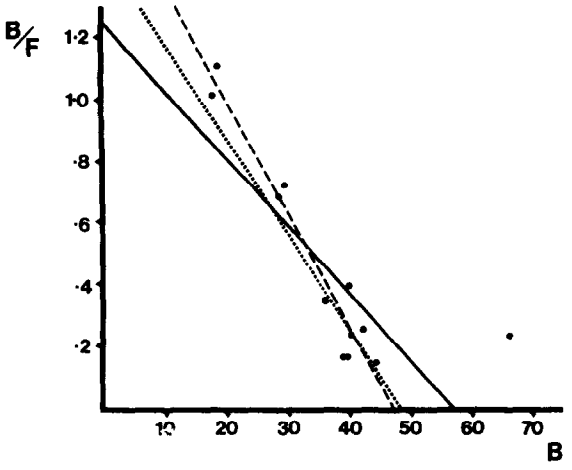

Fig. 1. Scatchard plot of data obtained from an estrogen receptor assay of human uterine cytosol fraction. On the ordinate is the ratio of bound (B) to free (F) hormone found at particular incubating concentrations of the hormone, while on the abscissa is the amount of hormone bound, expressed as fmol [^{3}H]-estradiol bound/mg cytosol protein. The values for the points are found in Table 1. The lines on the graph represent the least squares line on all 12 points (——) ($B/F = 1.247 - 0.022$ B), the least squares line with the outlier removed (– – –) ($B/F = 1.726 - 0.037$ B), and the direct linear plot (· · · · ·) ($B/F = 1.484 - 0.031$ B).

one data point is completely atypical (for example, due to a temporary aberration in one of the measuring devices), then it grossly affects the estimates. We really should remove the "wild" point and estimate B_{max} and K_D from the remainder. This involves the thorny problem of identifying the outlier. Alternatively, we could take the approach of looking for a *different* estimation procedure, one that is affected in only a small way by wild points. These are known as *robust* estimation procedures; see Huber[5]. This was indeed the motivation of Cornish-Bowden and Eisenthal[2] when they introduced the direct linear plot. Returning to Fig. 1, the full line represents the least squares line, the dashed line is the least squares line with the outlying point (66.1, 0.232) removed, and the dotted line is obtained from the direct linear plot. Note how close the last two lines are to each other. The direct linear plot gives estimates,

$$B_{max} = 47.97, \quad K_D = 32.32,$$

which are to be compared to the now reliable least squares estimates based on all but the outlier:

$$B_{max} = 47.07, \quad K_d = 27.27.$$

Of course, statistical estimation does not stop at just giving estimates; it must also provide us with standard errors of these estimates. Up to now, those who have written on the direct-linear-plot procedure have failed to adequately consider this equally important half of estimation theory. We will show in section 2 that the direct linear plot, although presented by Cornish-Bowden and Eisenthal[2] as a graphical technique, gives estimates that could equally be derived from a robust regression procedure known in the statistics literature for some time (Theil[12]). This hitherto unnoticed underlying structure not only allows a simple computational procedure to be set up, but more importantly, gives us the framework within which we can provide standard errors and so complete the estimation commitment. The subsequent parts of section 2 give the necessary theory. Section 3 presents illustrative examples, and finally, discussion and conclusions are given in section 4.

2. THE DIRECT LINEAR PLOT

2(a) *The underlying structure*

The equation describing the direct linear plot is a rearrangement of the Scatchard equation:

$$B_{max} = B + (B/F)K_D. \tag{4}$$

For each observation (F_i, B_i), a straight line can be drawn in the parameter space (i.e. K_D, B_{max} space), with slope B_i/F_i and intercept B_i. Consider the intersection point of the two lines defined by the observations (F_i, B_i) and (F_j, B_j). Woosley and Muldoon[13] show that it is given by

$$K_D = (B_j - B_i)/\{(B_i/F_i) - (B_j/F_j)\} \tag{5}$$

$$B_{max} = (F_i - F_j)/\{(F_i/B_i) - (F_j/B_j)\} \tag{6}$$

For n data points, there are $\frac{1}{2}n(n-1)$ such values for K_D and B_{max}, obtained by counting the number of pairs of distinct lines (i.e. the number of (i, j) such that $1 \leq i < j \leq n$). The direct linear plot estimate for K_D is then the median of the $\frac{1}{2}n(n-1)$ values for K_D. Similarly, the direct linear plot estimate for B_{max} is the median of the $\frac{1}{2}n(n-1)$ values for B_{max}.

However there remain problems associated with the use of the direct linear plot. Firstly the name "direct linear plot" is inappropriate, since in practice it is very time consuming to draw n straight lines in the parameter space, let alone to study the $\frac{1}{2}n(n-1)$ intersections. For $n = 10$ data points, there are 10 straight lines to draw, and 45 intersections to look at. Rather, one would directly calculate the values given by Eqs (5) and (6), and extract the medians from these values. To determine if the estimates are valid, one should see how Eq. (1) (or even (2) or (3)) fits the data; it is at *this* stage that a plot is required.

Secondly, the interpretation of the parameter estimates given by Eisenthal and Cornish-Bowden[3] and Woosley and Muldoon[13] in terms of medians of intersection points has in fact directed attention away from their real nature. In fitting a straight line of the form $y = a + bx$ to the points (x_1, y_1), (x_2, y_2), ..., (x_n, y_n), we naturally ask how we can choose a and b to provide a suitable fit to the points? The following estimator for b was introduced by Theil[12], and more generally investigated by Sen[9]. Firstly we compute,

$$s_{ij} = (y_i - y_j)/(x_i - x_j), \quad \text{for} \quad i < j.$$

That is, for any two points, s_{ij} is the slope of the line joining them. Theil proposed that

$$b = \text{median of the } \tfrac{1}{2}n(n-1)\ s_{ij} \text{ values.}$$

In the Scatchard plot $(B_i, B_i/F_i)$, Theil's value for the slope b in the plot of,

$$B/F = a + bB$$

becomes:

$$b = \text{median of } s_{ij} \text{ values,}$$

where now,

$$s_{ij} = \{(B_i/F_i) - (B_j/F_j)\}/(B_i - B_j).$$

Thus from equation (2),

$$-\frac{1}{K_D} = \text{median of } \{s_{ij}\}$$

$$= \frac{1}{\text{median of } \{1/s_{ij}\}}.$$

Therefore,

$$K_D = \text{median of } \{-1/s_{ij}\}$$

$$= \text{median of } \{(B_j - B_i)/[(B_i/F_i) - (B_j/F_j)]\}.$$

A comparison with Eq. (5) shows this to be the direct linear plot estimate. *In reality then, this estimate is immediately derivable from Theil's robust estimate of slope obtained from the Scatchard plot.*

A similar result is found for the estimate of B_{max} given by Eq. (6), as we now demonstrate. Another linear plot rarely used in the analysis of hormone–receptor interactions comes from a further way of re-writing equation (1):

$$F/B = (K_d/B_{max}) + (1/B_{max})F \tag{7}$$

Theil's estimate of the slope is given by

$$1/B_{max} = \text{median of } \{[(F_i/B_i) - (F_j/B_j)]/(F_i - F_j)\},$$

and hence,

$$B_{max} = \text{median of } \{(F_i - F_j)/[(F_i/B_i) - (F_j/B_j)]\}$$

A comparison with Eq. (6) confirms that *the direct linear plot estimate of B_{max} is immediately derivable from the robust slope-estimate of the* (rarely used) *plot of F/B against F.*

2(b) Straight line regression

The problem, very simply stated, is to estimate parameters a and b that give a "good" straight line fit,

$$y = a + bx$$

to data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. If statistical techniques are to be used, thus allowing reference to such powerful notions as unbiasedness, standard errors and confidence intervals, then a clear statement should be made about the statistical assumptions involved. The first assumption is that the x_i values are observed without experimental error; for this reason the x-variable is often called the control variable, or explanatory variable. The y-variable, however, is assumed to be observed with error. In statistical terms it is called a random variable.

The error, denoted here as e, accounts for the difference between the observed points and the (unknown) straight line relationship between y and x. Thus,

$$Y = \alpha + \beta x + e, \tag{8}$$

where we use Y instead of y to denote that it is a *random variable.* Therefore the data is assumed to satisfy:

$$Y_i = \alpha + \beta x_i + e_i \quad (i = 1, \ldots, n)$$

At some point, one has to specify the nature of the randomness, and this is usually done by probability statements:

$$Pr\{Y_i \leqslant y\} = G_i(y - \alpha - \beta x_i) \quad (i = 1, \ldots, n), \tag{9}$$

where the cumulative distribution functions $G_1, G_2, \ldots, G_n$ will be specified and determined by the experimenter's knowledge of the types of errors which might occur. Usually the n experiments are performed under identical conditions so that one can assume that $G_1 = G_2 = \ldots = G_n = G$. Most often the form of G can be assumed to be that of a normal distribution, and under this assumption the least square estimates of α and β, obtained by minimising:

$$\sum_{i=1}^{n} (Y_i - \alpha - \beta x_i)^2, \tag{10}$$

with respect to α and β, are the minimum variance unbiased estimates (Sen[9]). Typically though, as Huber[5] has pointed out, such a normal distribution assumption is not realistic, and the types of errors that should be allowed for are those which produce larger (positive or negative) values than the normal distribution. Unfortunately, in these cases, the least squares estimates of Eq. (10) perform extremely poorly. This had led to a recent development in statistics, known as robustness; which is the search for statistical procedures which remain near optimal for slight departures from normality. Moreover, when one realises that the form of the distribution function G is in reality often unknown, robust procedures are even more deserving of attention.

Our aim in this section is to provide suitable robust estimates for α and β, which remain valid for a broad class of G. We follow closely the work of Sen[10].

In the general case, suppose for each x_i (assume without loss of generality that $x_1 \leqslant x_2 \leqslant \ldots x_n$) there are n_i observations $Y_{i1}, \ldots, Y_{in_i}$ such that,

$$Y_{ij} = \alpha + \beta x_i + e_{ij} \quad (j = 1, \ldots, n_i;\ i = 1, \ldots, k),$$

and

$$n = \sum_{i=1}^{k} n_i$$

is the total number of observations on the y-variable. Sometimes, for example, all experiments are performed in duplicate, and then $n_i = 2$, and $n = 2k$. Now define the W-values by:

$$W_{ij,rs} = (Y_{js} - Y_{ir})/(x_j - x_i)$$

$$(1 \leqslant r \leqslant n_i;\ 1 \leqslant s \leqslant n_j) \tag{11}$$

for $1 \leqslant i < j \leqslant k$. Also, let

$$N = \sum_{1 \leqslant i < j \leqslant k} n_i n_j$$

be the total number of W-values. If the N values in Eq. (11) are arranged in ascending order of magnitude and denoted by,

$$W_{(1)} \leqslant W_{(2)} \leqslant \cdots \leqslant W_{(N)},$$

then the desired estimate of β in Eq. (8) is

$$\beta^* = \begin{cases} W_{(m+1)} & \text{if } N = 2m + 1 \\ \frac{1}{2}\{W_{(m)} + W_{(m+1)}\} & \text{if } N = 2m \end{cases}, \quad (12)$$

where m is a non-negative integer. Now this is precisely the median of the N W-values given by Eq. (11), and is exactly the estimate upon which the direct linear plot estimates are based (see section 2(a) above). Looking at the problem in this slightly more general context means that we have avoided the possibility of dividing by $(x_j - x_i)$ when $x_j = x_i$, which would result in an undefined W-value.

Now, by subtracting the slope effect, the intercept α is estimated. Define,

$$Y^*_{ir} = Y_{ir} - \beta^* x_i \quad (1 \leqslant r \leqslant n_i), \quad (13)$$

for $1 \leqslant i \leqslant k$. For the n Y^*-values in Eq. (13), the $N^* = \frac{1}{2}n(n+1)$. mid-ranges are defined by:

$$V_{ij,rs} = \tfrac{1}{2}\{Y^*_{ir} + Y^*_{js}\} \quad (1 \leqslant r \leqslant n_i;\ 1 \leqslant s \leqslant n_j),$$

for $1 \leqslant i < j \leqslant k$, and:

$$V_{ii,rs} = \tfrac{1}{2}\{Y^*_{ir} + Y^*_{is}\} \quad (1 \leqslant r \leqslant s \leqslant n_i),$$

for $1 \leqslant i \leqslant k$. Note that the V-values have been defined in such a way that there is no doubling-up on a value. For example, $V_{21,rs}$ is not allowed since it is taken care of by $V_{12,rs}$, just as $V_{ii,21}$ is taken care of by $V_{ii,12}$. Then the ordered V-values become:

$$V_{(1)} \leqslant V_{(2)} \leqslant \ldots \leqslant V_{(N^*)}, \quad (14)$$

and the desired estimate of α in equation (8) is

$$\alpha^* = \begin{cases} V_{(m^*+1)} & \text{if } N^* = 2m^* + 1 \\ \frac{1}{2}\{V_{(m^*)} + V_{(m^*+1)}\} & \text{if } N^* = 2m^* \end{cases} \quad (15)$$

where m^* is a non-negative integer. That is, the median of the paired averages of the Y^*-values is used (Adichie [1]).

As an example, assume the simple case of $n_i = 1$ for $i = 1, \ldots, k$, and hence $n = k$ (this means that all the x-values are distinct; i.e. it is not possible that $x_i = x_j$ for some $i \neq j$). We revert to using Y_1, Y_2, $\ldots$, Y_n, and the W-values become:

$$(Y_j - Y_i)/(x_j - x_i) \quad (1 \leqslant i < j \leqslant n).$$

The median of these $\frac{1}{2}n(n-1)$ values is the estimate β^* of (12). Define,

$$Y^*_i = Y_i - \beta^* x_i.$$

Then the V-values become,

$$V_{ij} = \tfrac{1}{2}\{Y^*_i + Y^*_j\} \quad (1 \leqslant i \leqslant j \leqslant n).$$

The median of these $\frac{1}{2}n(n+1)$ values is the estimate α^* of (15). Suppose we work with the data of Table 1 using B as the x-variable and B/F as the y-variable (i.e. a Scatchard plot). Then $n_i = 1$; $i = 1, \ldots, 12$, and $n = k = 12$. The W-values of (11) are:

$$W_{1,2} = (1.110 - 1.015)/(18.5 - 17.7) = 0.119$$

$$W_{1,3} = (0.684 - 1.015)/(28.5 - 17.7) = -0.030$$

$$W_{1,4} = \text{etc.}$$

When these values are ordered, and the median obtained, we get $\beta^* = -0.031$. The Y^*-values of (13) are:

$$Y^*_1 = 1.015 - (-0.031)(17.7) = 1.564$$

$$Y^*_2 = 1.110 - (-0.031)(18.5) = 1.684$$

$$Y^*_3 = \text{etc.}$$

and hence we can compute the V-values:

$$V_{1,1} = \tfrac{1}{2}(1.564 + 1.565) = 1.564$$

$$V_{1,2} = \tfrac{1}{2}(1.564 + 1.684) = 1.624$$

$$V_{1,3} = \text{etc.}$$

When these are ordered and the median chosen, we get $\alpha^* = 1.556$.

Fig. 2 shows the robust straight line regressions:

$$B/F = 1.556 + (-0.031)B,$$

superimposed over the data points.

Having defined the estimates β^* and α^*, there remains the necessity of finding their statistical properties; e.g. standard errors, confidence intervals, etc. The next part of this section does this, followed by a final part which corrects the standard error estimates used by Woosley and Muldoon[13] in the direct linear plot.

2(c) *Statistical properties*

Although Woosley and Muldoon[13] and Eisenthal and Cornish-Bowden[3] use statistical reasoning

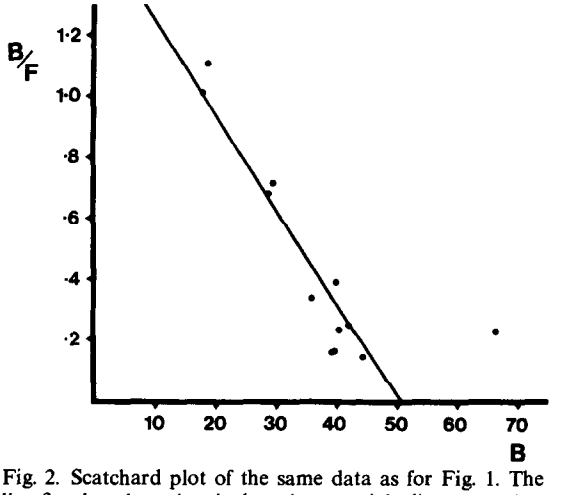

Fig. 2. Scatchard plot of the same data as for Fig. 1. The line fitted to the points is the robust straight line regression where equation is $B/F = 1.556 - 0.031$ B.

to derive direct-linear-plot estimates, there is a need for the estimates' statistical properties. Suppose we let,

$$T_n^2 = \sum_{i=1}^{k} n_i(x_i - \bar{x}_n)^2, \quad \text{where } \bar{x}_n = \sum_{i=1}^{k} n_i x_i/n$$

$$A_n^2 = \frac{1}{12}\left\{n(n^2-1) - \sum_{i=1}^{k} n_i(n_i^2-1)\right\}$$

$$\rho_n = \sum_{i=1}^{k} \{n_i m_{i-1} + n_i(n_i+1)/2 - n_i(n+1)/2\} \times \{x_i - \bar{x}_n\}/(T_n A_n),$$

where $m_i = n_1 + n_2 + \cdots + n_i$ $(i = 1, 2, \ldots, n)$ and $m_0 = 0$. Then Sen[9] has shown the following:

Property 1. If (i) ρ_n is strictly positive, and (ii) $T_n \to \infty$ as $n \to \infty$, then

$$\rho_n T_n(\beta^* - \beta)$$

has asymptotically a normal distribution with zero mean, and variance given by

$$\sigma^2(g) = \left\{12\left(\int_{-\infty}^{\infty} g^2(x)\,dx\right)^2\right\}^{-1}, \tag{17}$$

where $g(x)$ is the probability density function of the error term e:

$$g(x) = G'(x).$$

The following conclusions may be drawn:

(a) β^* is consistent. That is, for large n, β^* approaches the true (but unknown) value β. Sen[9] has also shown that the distribution of β^* is symmetrical about the true parameter β.

(b) β^* has the asymptotic standard error,

$$se(\beta^*) = \left\{\sqrt{12}\rho_n T_n \int_{-\infty}^{\infty} g^2(x)\,dx\right\}^{-1}. \tag{18}$$

(c) Confidence intervals (asymptotic) can be constructed for the unknown parameter β:

$$(\beta^* - Z_{\epsilon/2}\cdot se(\beta^*), \quad \beta^* + Z_{\epsilon/2}\cdot se(\beta^*)). \tag{19}$$

This is the $100(1-\epsilon)\%$ confidence interval for β where $Z_{\epsilon/2}$ is found from the normal tables and is defined by that value which solves:

$$\epsilon/2 = \int_{Z_{\epsilon/2}}^{\infty} (2\pi)^{-1/2} e^{-x^2/2}\,dx;$$

e.g. $\epsilon = 0.05$ gives $Z_{\epsilon/2} = 1.96$.

Property 2. Let α^* be the estimate of α, given by Eq. (15). According to Adichie[1] and Sen and Puri[11], if (i) ρ_n is strictly positive, (ii) $0 < \lim n^{-1}T_n^2 < \infty$, (iii) $|\lim \bar{x}_n| < \infty$, and (iv) $\lim\{\max(x_j - \bar{x}_n)^2/T_n^2\} = 0$ then

$$n^{1/2}(\alpha^* - \alpha),$$

has asymptotically a normal distribution with zero mean and variance,

$$\tau^2(g) = \left\{12\left(\int_{-\infty}^{\infty} g^2(x)dx\right)^2\right\}^{-1}\{1 + \lim[\bar{x}_n^2 n/(T_n^2\rho_n^2)]\}. \tag{20}$$

The following conclusions may be drawn:

(a) α^* is consistent

(b) α^* has the (asymptotic) standard error,

$$se(\alpha^*) = \left\{\sqrt{12}\rho_n T_n \int_{-\infty}^{\infty} g^2(x)\,dx\right\}^{-1}\left\{\frac{\rho_n^2 T_n^2}{n} + \bar{x}_n^2\right\}^{1/2}. \tag{21}$$

(c) The $100(1-\epsilon)\%$ confidence interval for α is given by

$$(\alpha^* - Z_{\epsilon/2}\cdot se(\alpha^*), \ \alpha^* + Z_{\epsilon/2}\cdot se(\alpha^*)). \tag{22}$$

Expressions (18), (19), (21), and (22) require knowledge of $\int_{-\infty}^{\infty} g^2(x)\,dx$. Now the probability density function of the errors is often assumed to be normal distribution with mean zero and scale parameter δ. that is,

$$g(x) = (2\pi\delta)^{-1/2} e^{-x^2/2\delta^2}.$$

Here

$$\int_{-\infty}^{\infty} g^2(x)\,dx = (2\sqrt{\pi}\,\delta)^{-1},$$

and δ is usually estimated by,

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} (Y_{ij} - \hat{\alpha} - \hat{\beta}x_i)^2/(n-2).$$

Recall however, our reticence in assuming normality, since the data is subject to outliers. We need therefore an estimate of $\int_{-\infty}^{\infty} g^2(x)\,dx$, which makes no assumptions about the (parametric) form of g; such an estimate is given by Sen[9]. Define,

$$L_n \equiv \frac{1}{18}\left\{n(n-1)(2n+5) - \sum_{j=1}^{k} n_j(n_j-1)(2n_j+5)\right\},$$

and recall that

$$N = \sum_{i \le i < j \le k} n_i n_j.$$

Then put

$$M_1 \equiv N/2 - (1.349)(L_n)^{1/2}/2,$$
$$M_2 \equiv N/2 + (1.349)(L_n)^{1/2}/2.$$

Now the W-values defined by (11) can be considered as a sample, which when ordered, give an interquartile range:

$$Q \equiv W_{([M_2+1])} - W_{([M_1])}, \tag{23}$$

where $[x] \equiv$ largest integer $\leqslant x$.

Property 3. Under the conditions of Property 1,

$$\sqrt{3}\rho_n T_n Q/(1.349),$$

is asymptotically the constant,

$$\int_{-\infty}^{\infty} g^2(x)\,dx.$$

The following conclusions may be drawn:

(a) From (18), $se(\beta^*)$ is estimated by:

$$\hat{se}(\beta^*) = \tfrac{1}{2}Q/(1.349), \tag{24}$$

where Q is given by (23). Then an (asymptotic) $100(1-\epsilon)\%$ confidence interval for β^* is got by substituting (24) into (19).

(b) From (21), $se(\alpha^*)$ is estimated by:

$$\hat{se}(\alpha^*) = (\tfrac{1}{2}Q/1.349)\left\{\frac{\rho_n^2 T_n^2}{n} + \bar{x}_n^2\right\}^{1/2}. \qquad (25)$$

Then an (asymptotic) $100(1-\epsilon)\%$ confidence interval for α^* is got by substituting (25) into (22).

For the data presented in Table 1, $Q = 0.0214$, $\rho_n^2 = 0.8332$, $T_n^2 = 1801.5565$, $\bar{x}_n = 36.7833$, and hence:

$$\hat{se}(\beta^*) = 0.0079 \quad \text{and} \quad se(\alpha^*) = 0.3051.$$

Combining this with the estimates $\beta^* = -0.0309$, and $\alpha^* = 1.5556$, gives a 95% ($Z_{\epsilon/2} = 1.96$ in (19)) confidence interval for β:

$$(-0.0465, -0.0154),$$

and a 95% confidence interval for α:

$$(0.9576, 2.1535).$$

2(d) Standard errors for the direct linear plot

As we have seen in section 2a, the *direct linear plot estimate for* K_D is:

$$K_D^* = -1/\beta^*,$$

where β^* is the robust estimate of slope in the Scatchard plot given by (2). Also *the direct linear plot estimate for* B_{max} is:

$$B_{\max}^* = 1/\beta^{**},$$

where β^{**} is the robust estimate of slope in the plot of F/B against F, given by (7).

In order to get approximate standard errors for K_D^* and B_{max}^*, we therefore need a formula which relates $se(K_D^*)$ to $se(\beta^*)$. We refer the reader to Frishman[4], from which we conclude that,

$$se(K_D^*) = se(\beta^*)/\{E(\beta^*)\}^2, se(B_{\max}^*) = se(\beta^{**})/\{E(\beta^{**})\}^2$$

Hence,

$$\hat{se}(K_D^*) = \hat{se}(\beta^*)/\{\beta^*\}^2, \quad \hat{se}(B_{\max}^*) = \hat{se}(\beta^{**})/\{\beta^{**}\}^2, \qquad (26)$$

$$K_D^* = -1/\beta^*, \quad B_{\max}^* = 1/\beta^{**}.$$

Equations (26) summarize what one needs to know about the estimation of K_D and B_{max}. For the data of Table 1, we have already seen that $\beta^* = -0.0309$, $\hat{se}(\beta^*) = 0.0079$, and hence $K_D^* = -1/(-0.0309) = 32.324$, together with $\hat{se}(K_D^*) = (0.0079)/(-0.0309)^2 = 8.2915$. Similarly, by converting Table 1 from $(B, B/F)$ points, to $(F, F/B)$ points, we can go through exactly the same procedure of sections 2(b) 2(c) to find:

$$\beta^{**} = 0.0208, \ \hat{se}(\beta^{**}) = 0.0009,$$

and hence $B_{\max}^* = 47.9719$, together with $\hat{se}(B_{\max}^*) = 2.1653$.

Unfortunately, the formula for calculation of standard errors given by Woosley and Muldoon[13], P.627, is wrong. By using that formula they implicity assume that their direct linear plot estimates are least squares estimates.

3. EXAMPLES

In the course of presenting sections 1 and 2 we used the data of Table 1, which were obtained from an estrogen receptor assay of human uterine cytosol fraction. The nature of the assay is outlined by Keightley, Tilley and Cant[6]. The results are brought together here, since they illustrate the robustness of the direct linear plot. They will be given in the form: estimate $\pm$ $\hat{se}$(estimate), for the Scatchard plot on all points (see Eq. (2)), for the Scatchard plot on all but the outlying point $(B_{12}, B_{12}/F_{12})$, and for the direct linear plot (see section 2(a)). Table 2 contains the Figs.

The second example works with the data of Table 3, which were obtained from an estrogen receptor assay of a second human uterine cytosol fraction. Figure 3 illustrates the 12 points plotted, with the least squares line, and the direct-linear-plot line superimposed.

The startling thing about the points $(B_1, B_1/F_1), \ldots, (B_{12}, B_{12}/F_{12})$ for this example, is that there are

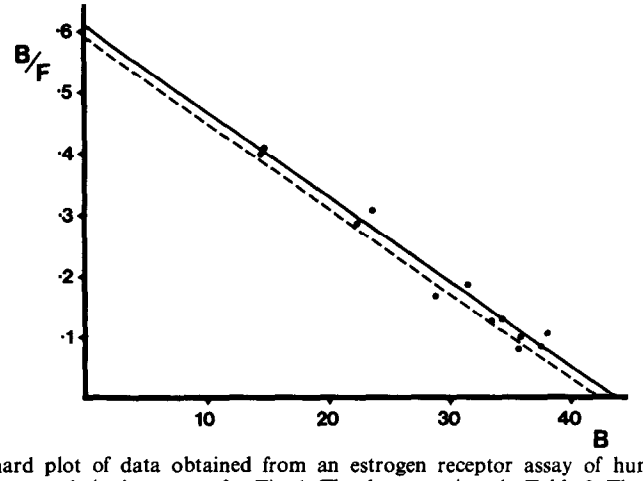

Fig. 3. Scatchard plot of data obtained from an estrogen receptor assay of human uterine cytosol fraction. Ordinate and abscissa are as for Fig. 1. The data are given in Table 3. The equation describing the least squares regression analysis line (——) is $B/F = 0.607 - 0.014\,B$, while that describing the direct linear plot (– – –) is $B/F = 0.588 - 0.014\,B$.

Table 4. Entries show: estimate $\pm$ $\hat{se}$(estimate), for two different estimation procedures, using the data of Table 3; B_{max} and K_D are expressed as in Table 2.

	$B^*_{max} \pm \hat{se}(B^*_{max})$	$K^*_D \pm \hat{se}(K^*_D)$
Least squares on Scatchard plot	43.174 $\pm$ 4.083	71.176 $\pm$ 3.992
Direct Linear plot	42.651 $\pm$ 1.114	72.572 $\pm$ 5.723

no obvious outliers. We therefore expect least squares to give excellent estimates of K_D and B_{max}. The crucial question then is: do the direct-linear-plot estimates do almost as well? The answer is clearly yes, when we look at the numerical results presented in Table 4.

4. CONCLUSIONS

Assay results from hormone receptors are most commonly expressed as a Scatchard plot, with a straight line fitted to the points via least squares regression analysis. Often with such assays, one or two points lie some distance from a line drawn through the remainder; i.e. an inspection by eye might choose certain points to be obviously aberrant. Generally they represent an error in experimental technique or measurement. If a least squares linear regression analysis is performed on all points of the plot, the values obtained for the number of binding sites and for the dissociation constant could be strikingly different from those obtained if the outlying points are ignored; the first example of section 3 illustrates this point. While some outlying points justifiably can be ignored, often this task is rather arbitrary. In this paper, the use of a robust technique of linear regression analysis is examined. This does not require agonizing decisions about whether to reject or not, since it *automatically* gives little weight or credence to unusually "wild" points. Further, it still has validity when *both* the dependent and independent variables contain error ([13]). A final requirement of the robust analysis is that if all the points are "well-behaved", then its resulting estimates should be almost identical to the estimates from the least squares analysis; the second example of section 3 confirms this.

This paper supports a recent suggestion that an analysis known as the direct linear plot, could be useful in analysing results from hormone receptor assays (Woosley and Muldoon[13]). The underlying structure of the direct linear plot is exposed here, enabling a thorough investigation of the estimates. These are more robust than standard least squares estimates, in that they are very much less affected by outlying points; this in turn flows on to estimates of the error.

We suggest that in analysing assay results from hormone receptors, the use of least squares in the Scatchard plot should be examined more closely. Certainly the robust method of linear regression represents a marked improvement, and should be adopted as routine (perhaps alongside a least squares analysis), particularly when outlying points are suspected. Finally we note that transformations from the Michaelis-Menten equation to scales *other* than those used by Scatchard (e.g. see Eq. (7)), have proved to be more appropriate in our analyses of hormone receptor data. This will be investigated elsewhere.

Acknowledgement—The authors would like to thank Ray McLean for computational assistance.

REFERENCES

1. Adichie J.: Estimates of regression parameters based on rank tests. *Ann. Math. Statist.* **38** (1967) 894–904.
2. Cornish-Bowden A. and Eisenthal R.: Statistical considerations in the estimation of enzyme kinetic parameters by the direct linear plot and other methods. *Biochem. J.* **139** (1974) 721–730.
3. Eisenthal R. and Cornish-Bowden A.: A new graphical procedure for estimating enzyme kinetic parameters. *Biochem. J.* **139** (1974) 715–720.
4. Frishman F.: On the arithmetic means and variances of products and ratios of random variables. In *Statistical Distributions in Scientific Work* (Edited by G. Patil, S. Kotz and J. Ord). Reidel, Dordrecht NL, Vol. 1 (1975) pp. 401–406.
5. Huber P.: Robust statistics: a review. *Ann. Math. Statist.* **43** (1972) 1041–1067.
6. Keightley D., Tilley W. and Cant E.: Effect of snap-freezing, dithiothreitol and storage on estimations of estrogen receptor sites. *Clin. chim. Acta* **88** (1978) 337–343.
7. Lineweaver H. and Burk D.: The determination of enzyme dissociation constants. *J. Am. Chem. Soc.* **56** (1934) 658–666.
8. Scatchard G.: The attractions of proteins for small molecules and ions. *Ann. N.Y. Acad. Sci.* **51** (1948) 660–672.
9. Sen P.: Estimates of the regression coefficient based on Kendall's tau. *J. Am. Statist. Assoc.* **63** (1968) 1379–1389.
10. Sen P.: Robust statistical procedures in problems of linear regression with special reference to quantitative bio-assays, I. *Rev. Int. Statist. Inst.* **39** (1971) 21–38.
11. Sen P. and Puri M.: On robust non-parametric estimation in some multivariate linear models. In *Multivariate Analysis—II* (Edited by P. Krishnaiah). Academic Press, New York (1969) pp. 33–54.
12. Theil H.: A rank-invariant method of linear and polynomial regression analysis. *Proc. Kon. Ned. Akad. v. Wetensch. A*, **53** (1950) 386–392.
13. Woosley J. and Muldoon T.: Comparison of the accuracy of the Scatchard, Lineweaver–Burk and direct linear plots for the analysis of steroid–protein interactions. *J. steroid Biochem.* **8** (1977) 625–629.