When the atoms are represented by hard spheres of diameter a, it is possible to perform the calculations explicitly. As is well known :

$$B = b, \quad C = \tfrac{5}{8}b^2, \quad \text{where } b = 2\pi a^3/3.$$

Born and Green's non-linear integral equation, embodying the superposition-principle, leads to $D = 0 \cdot 2252b^3$, while the linearized formulæ give $D = 0 \cdot 3958b^3$. These values may be compared with the value $D = 0 \cdot 2869b^3$ calculated from first principles by Happel and by Majumdar (see Fowler and Guggenheim[5]).

A fuller account of this work will be published elsewhere.

G. S. RUSHBROOKE
H. I. SCOINS

Clarendon Laboratory,
Oxford.
Oct. 23.

[1] Born, M., and Green, H. S., "A General Kinetic Theory of Liquids", chapters 1 and 2 (Cambridge, 1949).
[2] Kirkwood, J. G., and Boggs, E. M., J. Chem. Phys., 10, 394 (1942).
[3] Mayer, J. E., et al., J. Chem. Phys., 5, 67, 74 ; 6, 87, 101 (1937).
[4] Rodriguez, A. E., Proc. Roy. Soc., A, 196, 73 (1949).
[5] Fowler, R. H., and Guggenheim, A. E., "Statistical Thermodynamics", 289 (Cambridge, 1939).

Simplified Calculation of a Linear Regression

THE "simplified" method of calculating a linear regression put forward by Aldridge, Berry and Davies[1] is the well-known method of orthogonal polynomials which was put on a practical working basis by Fisher[2,3]. Unfortunately, in bringing this procedure again to the attention of research workers, the authors reproduce a misapprehension when they deduce that "whereas it is considered that four points are a minimum for determining the slope of the line, the number of observations should always be even", because observations at the centre point of an odd number of values of the independent variable contribute no information to the linear regression.

If it is known that the relationship is linear, then greatest accuracy for a given amount of labour will be achieved by concentrating all observations at the extremes of the practicable range—equal numbers at each extreme if the cost per observation is equal at all points. If it is required also to check the hypothesis that the regression is linear, then greatest efficiency is given by concentrating the observations at three points.

This may be most easily seen by considering the error variances for the linear and quadratic coefficients when observations are equally distributed between two, three or four points of the independent variable. (The quadratic coefficient is considered in this context only as a test for deviation from linearity.) To retain simple numbers let the extremes of the available range be at $x = -3$ and $+3$; and consider n observations distributed as follows :

x	-3	-1	0	1	3
(a)	$\tfrac{1}{2}n$	—	—	—	$\tfrac{1}{2}n$
(b)	$\tfrac{1}{3}n$	—	$\tfrac{1}{3}n$	—	$\tfrac{1}{3}n$
(c)	$\tfrac{1}{4}n$	$\tfrac{1}{4}n$	—	$\tfrac{1}{4}n$	$\tfrac{1}{4}n$

Let V be the variance of the dependent variate at any single value of x, and fit the orthogonal polynomial

$$y = A + B\,\xi_1 + C\,\xi_2 ;$$

where B is the usual linear regression, and C the quadratic coefficient. (With x symmetrically distributed and mean 0, $\xi_1 = x$, $\xi_2 = x^2 -$ mean of x^2.)

Then the variance of B is $V/\Sigma\xi^2_1$, and of C is $V/\Sigma\xi^2_2$; and the values of the denominators, proportional to the amounts of information on each coefficient, for the above three cases are :

	$\Sigma\xi^2_1$	$\Sigma\xi^2_2$
(a)	$9n$	—
(b)	$6n$	$18n$
(c)	$5n$	$16n$

Thus, although the central observation contributes nothing to B, still the greater spread of two-thirds of the observations in distribution (b) gives 20 per cent more information than does distribution (c), while the concentration of the remaining third gives $12\tfrac{1}{2}$ per cent more information on the deviation from linearity.

The real choice is whether or not it is worth while to vary the proportion of observations in the centre class to gain greater accuracy on one coefficient at the expense of the other, but such variation will not usually be worth while. Greatest accuracy on C is given with distribution $\tfrac{1}{4}n$, $\tfrac{1}{2}n$, $\tfrac{1}{4}n$; but this will increase $\Sigma\xi^2_2$, in the units of the above table, only from $18n$ to $20 \cdot 25n$ at the expense of dropping $\Sigma\xi^2_1$ from $6n$ to $4 \cdot 5n$; on the other hand, if the central frequency be reduced below $\tfrac{1}{4}n$, the evaluation of curvature will be too inaccurate to be of much use as a critical test in cases of doubt, and one might as well 'plunge' for (a).

Yates[4] indicates some other points of view with respect to coefficients determined from three levels. An interesting study of efficient distribution of observations in regression analysis is given by Hotelling[5].

H. FAIRFIELD SMITH
Rubber Research Institute of Malaya,
Kuala Lumpur.
March 13, 1950.

[1] Aldridge, Berry and Davies, Nature, 164, 925 (1949).
[2] Fisher, R. A., "Statistical Methods for Research Workers" (Oliver and Boyd).
[3] Fisher and Yates, "Statistical Tables for Biological Research" (Oliver and Boyd).
[4] Yates, F., Imp. Bur. Soil Sci. Tech. Comm. 35 (1937).
[5] Hotelling, H., Ann. Math. Stat., 12, 20 (1941).

The Permicron: a Unit of Wave Number

IN spite of the great theoretical advantages of recording spectra in terms of frequency or wave number, most data on visible and ultra-violet spectra, more particularly spectra of solutions, are in practice expressed as wave-lengths with the millimicron as a favourite unit. One reason for this preference, apart from habit, is found in the inconvenience of sec.$^{-1}$ and cm.$^{-1}$ as units. Frequencies expressed in sec.$^{-1}$ involve large powers of ten ($\sim 10^{15}$), while the $fresnel$[1] ($= 10^{12}$ sec.$^{-1}$) has never become popular. Wave numbers in cm.$^{-1}$ are of the order of 10^4, and require the use of written or spoken powers of ten if the recorded values are to express the relatively low precision of measurement in spectra of solution. Thus, to say that the absorption maximum of bromine is at 405 mμ implies a reasonable uncertainty of about 10 A. in locating it. To express the value as 24,691 cm.$^{-1}$ or even as 24,700 cm.$^{-1}$ implies a much greater precision than is obtainable experimentally.

The difficulty could be overcome by employing a new unit of wave number, the number of waves $per\ micron$. This unit could be named 'permicron', and written as such or else as μ^{-1} or $/\mu$. Thus the precision of measuring the absorption maximum of