37

III. *The Influence of Rainfall on the Yield of Wheat at Rothamsted.*

By R. A. Fisher, *M.A., Head of the Statistical Department, Rothamsted Experimental Station, Harpenden, Fellow of Gonville and Caius College.*

Communicated by Sir John Russell, *F.R.S.*

(Received September 26,—Read November 8, 1923.)

Contents.

1. *General Problem of Evaluating the Effects of Weather on Crops.*

At the present time very little can be claimed to be known as to the effects of weather upon farm crops. The obscurity of the subject, in spite of its immense importance to a great national industry, may be ascribed partly to the inherent complexity of the problem which it presents, and more especially to the lack of quantitative data obtained either under experimental or under industrial conditions, by the study of which accurate knowledge alone can be acquired. Of the industrial applications of such knowledge it is unnecessary to speak here in detail. It is sufficient to indicate that the present system of Life Insurance, which safeguards the economic stability of many thousands of families, and occupies the activities of many of the greatest financial corporations, was made possible by the studies of statistics of human mortality by the mathematicians of the eighteenth and nineteenth centuries, and that on the basis of adequate knowledge similar economic stability with its attendant security of capital should be within the reach of the industrial farmer.

Click here to view part 2 of paper

The inherent complexity of the relationships which it is sought to elucidate, between the yields of farm crops, and the previous weather which largely controls those yields, arises primarily from the complexity of the problem of specifying the weather itself. Meteorologists have, however, gradually devised a number of instrumental observations, which although far from specifying the total environment of the growing plant, as understood by the plant physiologist, do nevertheless give a sufficiently detailed account of the general environmental conditions of the growing crop, in so far as these vary from season to season. It is probable indeed that almost all the weather influences to which crop variations are due could be expressed in terms of the instrumental observations of a modern meteorological station. The actual difficulty of calculating the crop variations from given instrumental records is, however, immense ; only an attack of the most preliminary kind upon the general problem can be attempted in this paper. The complete aim of agricultural meteorology should, however, be emphasised, for it is only by its substantial achievement that other causes of crop variation can be freed from much obscurity.

A valuable example of an investigation in which this aim was clearly held in view is to be found in A. WALTER'S study in 1910 of the effects of weather upon the Sugar crop in Mauritius (17). WALTER was able to give a very complete account of the meteorological causes affecting the yield of sugar, to such an extent that this author considers that under uniform estate management and methods of cultivation the yield may be predicted from meteorological data only with a standard error of the order of 1 per cent. The number of harvests available for this enquiry seems, however, to have been only 13, and more recent investigations in statistical theory make it now possible to realise, more clearly than was at the time possible, some of the dangers which arise from paucity of crop data.

2. *Effects of paucity of Crop Data.*

The biometrical investigations for which the method of multiple correlation was developed, differ from such agricultural studies as the present in two main particulars. In the first place the number of individuals measured was taken to be large, of the order of 1,000, or at least of some hundreds, and the special problems of distribution which arise in small samples have only more recently begun to receive attention. In the second place the number of measurements taken of each individual, or the number of *variates*, was generally small, and in all cases far smaller than the number of meteorological elements which may plausibly be regarded as affecting the crop. The sequence of weather to which crop variations may be ascribed extends over a year, or even more, and consists in this country of a series of abrupt, relatively violent and transient spells, each of which has its influence on the crop. If we wished to analyse the sequence no more closely than by monthly averages, we should still have 12 values for rainfall, and 12 more for maximum and minimum temperature, dew point, grass minimum, solar maximum and soil temperatures, nor would it be unreasonable to include some such measure of insolation as is given by " Hours of Bright Sunshine," and averages for the direction and force

of the wind. The number of meteorological elements might be made to exceed even the longest series of crop records available, for the Rothamsted wheat records provide, after necessary deductions, only 60 yields. Consequently, if the computer were provided not with yield data at all, but with an equal number of values composed at random, he would still be able to express them with perfect accuracy in terms of the weather records, for the number of unknowns available would exceed the number of equations for them to satisfy.

It is of more practical importance that even when we have selected a number of meteorological variables which is less than the number of crops recorded, a strong semblance of dependence may be produced, even when fictitious data, unrelated to the weather, are substituted for the true crops. For let there be n values of the dependent variate y and n each of the q meteorological variates $x_1, x_2, \ldots, x_q$. If now we obtain a partial regression equation

$$Y = \bar{y} + S(c_k(x_k - \bar{x}_k)),$$

showing the apparent dependence of y upon $x_1, \ldots, x_q$, then the correlation between y and Y, namely R, will be the multiple correlation of y with $x_1, \ldots, x_q$, and will be greater than the correlation of y with any other linear function of $x_1, \ldots, x_q$. Since R is necessarily positive (or zero), as was noticed by HOOKER (10, p. 7), there will be generally an appearance of correlation produced in this way, even if no such relation really exists. This effect becomes particularly marked if the number in the sample is small.

When there is no real correlation it is possible to determine *a priori* the distribution of the multiple correlation of any quantity y with any q variates $x_1, x_2, \ldots, x_q$. There are many ways of approaching the solution, of which perhaps the most enlightening is to regard the deviations $y - \bar{y}$ as components determining the direction of a radius vector OP in n dimensions. Then q other directions are specified by the q meteorological variates, and any linear regression formula is represented by some line through the origin in the space of q dimensions, specified by these q directions. The multiple correlation, R, is then the cosine of the angle which the line OP makes with the space of q dimensions. If the variate y is unrelated to the variates $x_1, \ldots, x_q$, the line OP may be regarded as drawn at random through O, in the space of n dimensions; using this fact it may be shown without difficulty that the chance that R^2 falls into the elementary range dR^2 is

$$df = \frac{\frac{n-3}{2}!}{\frac{n-q-3}{2}!\,\frac{q-2}{2}!}(R^2)^{\frac{q-2}{2}}(1-R^2)^{\frac{n-q-3}{2}}d(R^2). \quad \ldots\ldots (I)$$

It will be noted that the distribution is independent of any correlations which may exist between the meteorological variates, provided no one of them can be exactly

calculated from the others. In the case $q = 1$, when only one independent variate is used, the formula reduces to

$$df = \frac{2.\frac{n-3}{2}!}{\frac{n-4}{2}!\sqrt{\pi}}(1-R^2)^{\frac{n-4}{2}}dR,$$

which is known (FISHER (4), 1915) to be the distribution of the correlation (taken positive) derived from a sample of n, between two uncorrelated variates. It may also be seen that when $q = n-1$, R is necessarily equal to $+1$.

The mean value of R^2 is

$$\overline{R^2} = \frac{q}{n-1},$$

and since

$$S(Y-\bar{y})^2 = R^2 S(y-\bar{y})^2,$$

we may say that in the absence of correlation, the variance of y (the square of its standard deviation) will be distributed on the average equally between the q degrees of freedom of the regression formula, and the $(n-q-1)$ degrees of freedom in which y departs from the regression formula. This view of the matter enables us to follow the above conclusions to some extent into the more general case in which y is really in some degree correlated with $x_1, \ldots, x_q$. For if y be imagined as made up of two parts, $y_0 + y'$, such that y_0 is wholly determined as a linear function by the independent variates and y' is wholly uncorrelated with them, then y_0 and y' must be mutually uncorrelated, so that each will contribute a certain percentage to the variance of y. Let the fraction contributed by y_0 be A, then the regression formula found by correlating a sample of y' with the independent variates will differ from that found by correlating y, merely by y_0 which is by hypothesis a linear function of $x_1, \ldots, x_q$. Consequently the average proportion of the variance in the regression formula for y will be

$$\overline{R^2} = A + \frac{q}{n-1}(1-A),$$

or

$$1-\overline{R^2} = (1-A)\frac{n-q-1}{n-1}.$$

If the sample be increased indefinitely $(n \to \infty)$ then the limiting value of R^2 is A. An estimate of A from a finite sample, free from the positive bias of R^2, is obtained by taking

$$1-A_1 = \frac{n-1}{n-q-1}(1-R^2). \quad \ldots\ldots\ldots \quad (II)$$

A complete discussion of the errors of random sampling in multiple regression requires a knowledge of the frequency distribution of R, or of A_1, calculated by equation (II), in the general case when A is not zero; equation I gives this distribution only when

$A = 0$. It provides a means of testing if an observed value of R differs significantly from zero, but not for testing in general, whether one value of R^2 is significantly greater than another.

From equation (I) it appears that the probability that R, obtained from a random sample of n observations, should exceed any specified value, is, if q is even,

$$P = (1 - R^2)^{\frac{1}{2}(n-q-1)} \left\{ 1 + \frac{n-q-1}{2} R^2 + \frac{(n-q-1)(n-q+1)}{2.4} R^4 + \ldots \frac{(n-q-1)\ldots(n-5)}{2.4\ldots(q-2)} R^{q-2} \right\},$$

and, if q is odd,

$$P = \frac{2}{\sqrt{\pi}} \frac{\frac{n-q-2}{2}!}{\frac{n-q-3}{2}!} \int_{\sqrt{\frac{R^2}{1-R^2}}}^{\infty} \frac{dz}{(1+z^2)^{\frac{1}{2}(n-q)}} + \frac{2}{\sqrt{\pi}} \frac{\frac{n-q-2}{2}!}{\frac{n-q-3}{2}!} R (1 - R^2)^{\frac{1}{2}(n-q-1)} \left\{ 1 + \frac{n-q}{3} R^2 + \ldots \frac{(n-q)(n-q+2)\ldots(n-5)}{3.5\ldots(q-2)} R^{q-3} \right\}.$$

The analogy of these formulæ with those giving the probability, P, that χ^2, the Pearsonian test of goodness of fit, should exceed any specified value is obvious. It will be noticed that in the second formula the probability integral of the normal curve is replaced by that of the Type VII curves which has been tabulated by "STUDENT" (16, 1917). When q is small, very few of the terms of the series are involved, thus for $q = 6$, or 7, the series terminates with the term in R^4.

The effect of increasing q, the number of variates used, is to increase somewhat rapidly the probability that R should exceed any specified value, even for an independent variate wholly uncorrelated with those used to predict it; this increase may be exemplified in the following table, where we have taken $n = 13$, $q = 4$, 6, 8 and calculated the chance of R exceeding 0·5, 0·6, 0·7, 0·8 and 0·9.

TABLE I.—$n = 13$.

q	R = 0·5	0·6	0·7	0·8	0·9
4	0·6328	0·4094	0·2002	0·0598	0·0055
6	0·8965	0·7491	0·5187	0·2509	0·0505
8	0·9844	0·9402	0·8248	0·5906	0·2424

It will be seen that the chance of obtaining a value $R > 0{\cdot}9$ is only one in 180 for $q = 4$; it rises to one in twenty for $q = 6$, and about one in four for $q = 8$. For $q = 8$, therefore, an observed multiple correlation $0{\cdot}9$ cannot be regarded as significant, that is as convincing evidence that the "crop" is in any way influenced by the meteorological variates. The only value, in fact, in the above table which could be regarded as definitely significant is the value $0{\cdot}9$ for $q = 4$, while for $q = 6$ this is suggestive only of real influence.

A still more insidious source of illusory high correlations lies in the fact that the particular variates chosen for correlation with the crop figures, are often chosen *because* they appear in fact to be associated with the crop. It is a common practice, as a preliminary to the study of weather correlations, to search for the so-called critical periods by such methods as the following. The years are arranged in order of crop yield, and a number of meteorological values are plotted on a chart in this order; those meteorological values which show an apparent trend upwards or downwards, are then picked out, and used to construct a weather formula by which the crop may be predicted. Such methods are especially deceptive when the series of years is short (about 20 or 30), for if such a process were carried out thoroughly we should not have merely the random sampling distribution of R, which as we have seen is capable of yielding sufficiently high values from uncorrelated material, but we should be choosing that set of q variates, out of a larger number p, which gave the highest value of R.

The effect of such a process, if dummy data were substituted for the actual crop records, could be accurately foretold by the solution of the following problem. A line OP is drawn through a point O, at random in n dimensions, p directions (where p may exceed n) are also chosen at random, and of them q ($q < n$) are selected so that the space of q dimensions through these makes the least possible angle, θ, with OP. Find the sampling distribution of θ (or, of $R = \cos\theta$).

I can put forward no general solution of this problem; the case $p = q$ has been solved above. The additional advantage conferred by a choice of variates may be clearly seen from the case $q = 1$. For

$$\int_0^R \frac{2 \cdot \frac{n-3}{2}!}{\sqrt{\pi} \cdot \frac{n-4}{2}!} (1 - R^2)^{\frac{n-4}{2}} dR$$

represents the chance that for a single uncorrelated variate chosen at random, the correlation in a sample shall not exceed R; hence it follows that the chance that for none of p uncorrelated variates chosen at random, does the correlation exceed R, must be

$$\left\{ \int_0^R \frac{2 \cdot \frac{n-3}{2}!}{\sqrt{\pi} \cdot \frac{n-4}{2}!} (1 - R^2)^{\frac{n-4}{2}} dR \right\}^p;$$

and this is the chance that the highest correlation of p shall not exceed R.

If a standard of significance be chosen such that it will be exceeded in random samples once in 20 times, the effect of having a choice of p variates will increase the chance of exceeding any assigned rare value nearly p times. Thus for $n = 13$, 0·6 will not be judged significant, even for a single variate correlated if p exceed 2, 0·7 ceases to be significant if $p > 7$, and 0·8 when p exceeds about 54. When we consider that p may be very great, since not only is the number of meteorological elements available for selection large, but also since many writers allow themselves to use complicated functions of the instrumental data, involving adjustable weights, special conventions of sign, and allowances for periods judged to be unusual in their effects, so that these artificial meteorological data may be calculated in an enormous variety of ways, it is clear that the conclusions we have drawn as to the dangers of applying multiple correlation formulæ to small samples, are very much to be emphasised when a choice is made as to what meteorological elements to correlate.

In view of the foregoing facts it would seem worth while to lay down the following conditions for arriving at unprejudiced results :—

(i) The meteorological variates to be employed must be chosen without reference to the actual crop record.

(ii) If multiple variates are to be used allowance must be made for the positive bias of R^2.

(iii) Relationships of a complicated character should be sought only when long series of crop data are available.

3. *The Analysis of the Season.*

By the Season is meant the whole sequence of weather directly or indirectly influencing the crop from its inception to the time the produce is weighed. In studying the influence of the rain on the wheat crop, we have chosen a period of a year ending on the 31st of August of the year in which the crop was harvested. To obviate the irregularity of the calendar, this year was taken to be of 366 days, so that it commenced on either the 31st of August or the 1st of September prior to the sowing of the seed. Taking into consideration a single element, namely Rain, only, this sequence is in our climate sufficiently complex. The general distribution of rain through the year could, indeed, be roughly represented by dividing the period up into say 6 sections and recording the total rainfall in each period. If the rainfall were approximately evenly distributed in successive days or weeks, this method might indeed give a fairly accurate picture of the sequence of rainfall. But notoriously the rain is in fact often concentrated in short spells, with rainless periods intervening, consequently if March and April formed one section, a great part of the rain ascribed to that section might fall early in March in one year, and late in April in another, and it would be impossible to regard such falls as equivalent

in their effects on the crop. Whereas a late April fall might well be nearly equivalent to one occurring early in May.

This consideration suggests that a finer subdivision of the year is necessary; that months or even weeks should be treated separately. From the agricultural and meteorological point of view there is everything to be said for such subdivision; it raises, however, new mathematical difficulties; for whereas the evaluation of determinants of 6 rows and columns, or the solution of simultaneous linear equations in 6 unknowns is sufficiently rapid, disproportionate labour is involved in increasing the 6 to 12, and if 52 unknowns were attempted the labour involved would become fantastic.

A consideration which has been still more influential in framing our method is that even if we were to calculate the 52 partial regression coefficients showing the average effect in bushels per acre of an extra inch of rainfall, for each week of the year, such a calculation would leave out of consideration the all-important fact that this effect may be expected to change *continuously* during the year. The true partial regression coefficients for neighbouring periods must be relatively alike. The method of partial correlation as hitherto developed, takes no account of the serial character of our weather variates; after calculating the partial regression coefficients, we should still be far from the facts, if we did not smooth the series so obtained by a continuous curve, which should average out the independent errors in the values obtained for the successive weeks.

Disregarding, then, both the arithmetical and the statistical difficulties, which a direct attack on the problem would encounter, we may recognise that whereas with q subdivisions of the year, the linear regression equations of the wheat crop upon the rainfall would be of the form

$$\bar{w} = c + a_1 r_1 + a_2 r_2 + \ldots + a_q r_q$$

where r_1, r_2, ..., r_q are the quantities of rain in the several intervals of time, and a_1, ... a_q are the regression coefficients, so if infinitely small subdivisions of time were taken, we should replace the linear regression function by a *regression integral* of the form

$$\bar{w} = c + \int_0^T a r \, dt, \qquad \text{(III)}$$

where $r\,dt$ is the rain falling in the element of time dt; the integral is taken over the whole period concerned, and a is a *continuous* function of the time t, which it is our object to evaluate from the statistical data.

It will be seen that by proceeding to the limit all artificiality has been eliminated from the quantity a, which now represents an objective physical quantity, namely the average benefit to the crop in bushels per acre per inch of rain, falling in the time-element considered. This, of course, is more than even a daily record of rain can tell us, but owing to the relatively slow changes in the functions a, we shall find it sufficient to divide the 366 day year into 61 equal periods of 6 days each.

It should be noted that corresponding to the quadratic terms of a regression formula the independent variates of which form a continuous series, *i.e.*, to

$$\mathrm{S}\,(b_{pq}r_p r_q),$$

we have the double regression integral

$$\int_0^{\mathrm{T}}\int_0^{\mathrm{T}} b\,(t,\ t')\,r_t r_{t'}\,dt\,dt',$$

where b is a continuous function of the two epochs t and t', just as a is a continuous function of t only.

The concept of the regression integral, involving a regression function varying continuously with the time, is not only of service in displaying in a precise mathematical form the nature of the relationship which is to be investigated, but when the functions a and b vary with the time relatively slowly, it suggests a statistical procedure by which values of these functions may be obtained from the data. That the values must change relatively slowly is apparent from the fact that the state of advancement of the crop, as indicated either by the time of harvest, or by phenological observations on common weeds, often varies relatively to the Calendar date by as much as a fortnight.

If, then, T_0, T_1, T_2 ... be a series of orthogonal functions of the time, such that

$$\int_0^{\mathrm{T}} \mathrm{T}_r\mathrm{T}_s\,dt = 0 \qquad (r \neq s)$$

$$\int_0^{\mathrm{T}} \mathrm{T}_r^2\,dt = 1,$$

we may express the rate of rainfall at any epoch, in the form

$$r = \rho_0\mathrm{T}_0 + \rho_1\mathrm{T}_1 + \rho_2\mathrm{T}_2 + \dots,$$

where

$$\rho_s = \int_0^{\mathrm{T}} r\mathrm{T}_s\,dt;$$

also we may express the regression function in the form

$$a = \alpha_0\mathrm{T}_0 + \alpha_1\mathrm{T}_1 + \alpha_2\mathrm{T}_2 + \dots, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(IV)}$$

where

$$\alpha_s = \int_0^{\mathrm{T}} a\mathrm{T}_s\,dt,$$

noting in the latter case that relatively few terms of the expansion will be required, if a is a slowly-varying quantity.

Then we shall have from equation (III)

$$\bar{w} = c + \int_0^T ar\,dt = c + \alpha_0\rho_0 + \alpha_1\rho_1 + \alpha_2\rho_2 + \ldots .$$

Now the values of ρ_0, ρ_1, $\rho_2, \ldots$, may be obtained for each year from the rain record, and by correlating them with the crop data, we can obtain the values of α_0, α_1, α_2, ... , as partial regression coefficients of the crop on the coefficients of the rainfall distribution ; then from (IV) we can evaluate the expansion of a to as many terms as we have obtained. It is easy to verify that in the same manner the double regression integral may be expressed as a quadratic regression formula for the crop in terms of the rainfall distribution values.

The advantage of this indirect method of attack lies not only in the facts that for a fine division of the year, the arithmetical difficulties of a direct attack are insurmountable, and that no existing crop record is sufficiently long to justify the calculation from it of 50 or 60 independent regression coefficients, for the probable error of such determinations depends on the excess of the number of observations over that of the coefficients evaluated (9), but also in the increased accuracy attained through utilising our knowledge that the regression function must in fact vary relatively slowly. This advantage is analogous to that obtained by the author in collaboration with Miss Mackenzie, who found that using the correlations of weekly rainfall between different stations in Great Britain from a record of about 40 years, it is possible to estimate the correlation, for any one week of the year, with an accuracy which, for any single week, would have required records for over 1,000 years ; this was due to the fact that the sequence of correlations could be well represented by a harmonic curve of only the second order, so that use could be made of some 2,000 actual simultaneous observations of weekly rainfall to eliminate errors of random sampling (8. Fisher and Mackenzie, 1922). In the present investigation the accuracy is limited by the number of crops recorded, but the method employed enables full use to be made of the meteorological data.

In the practical application of the method of the regression integral, it is not necessary that the rainfall record should be strictly continuous ; indeed daily observations are themselves too numerous to be conveniently handled. The rain was therefore divided up into 61 periods of 6 days each, the integrations being replaced by summations over 61 terms. This introduces a slight modification into the form of the orthogonal functions of the time, for it is necessary for exact work that

$$\sum_1^{61} (T_r T_s) \qquad r \neq s$$

should vanish. The most convenient orthogonal functions to use are those developed by Esscher (3) in respect to mortality, and independently by the present author, in eliminating the slow changes observable in the wheat yields of the experimental plots

(5, 1921). If the time is measured from the middle point of a series of n terms, in units equal to the interval between successive terms, we have

$$T_0 = \frac{1}{\sqrt{n}},$$

$$T_1 = \sqrt{\frac{12}{n(n^2-1)}}\,t,$$

$$T_2 = \sqrt{\frac{180}{n(n^2-1)(n^2-4)}}\,(t^2 - n_2),$$

$$T_3 = \sqrt{\frac{2{,}800}{n(n^2-1)(n^2-4)(n^2-9)}}\left(t^3 - \frac{n_4}{n_2}t\right),$$

$$T_4 = \sqrt{\frac{44{,}100}{n(n^2-1)(n^2-4)(n^2-9)(n^2-16)}}\left(t^4 - \frac{n_6 - n_2n_4}{n_4 - n_2^2}t^2 + \frac{n_2n_6 - n_4^2}{n_4 - n_2^2}\right),$$

$$T_5 = \sqrt{\frac{698{,}544}{n(n^2-1)(n^2-4)(n^2-9)(n^2-16)(n^2-25)}}\left(t^5 - \frac{n_2n_8 - n_4n_6}{n_2n_6 - n_4^2}t^3 + \frac{n_4n_8 - n_6^2}{n_2n_6 - n_4^2}t\right),$$

where n_2, n_4, n_6 and n_8 stand for the mean values of t^2, t^4, t^6 and t^8. So that

$$n_2 = \tfrac{1}{12}(n^2 - 1),$$

$$\frac{n_4}{n_2} = \tfrac{1}{20}(3n^2 - 7),$$

$$\frac{n_6 - n_2n_4}{n_4 - n_2^2} = \tfrac{1}{14}(3n^2 - 13),$$

$$\frac{n_2n_8 - n_4^2}{n_4 - n_2^2} = \tfrac{3}{560}(n^2 - 1)(n^2 - 9),$$

$$\frac{n_2n_8 - n_4n_6}{n_2n_6 - n_4^2} = \tfrac{5}{18}(n^2 - 7),$$

$$\frac{n_4n_8 - n_6^2}{n_2n_6 - n_4^2} = \tfrac{1}{1008}(15n^4 - 230n^2 + 407).$$

These functions are in reality very convenient to handle, the practical arithmetic involved will be explained in Section (5) under the analysis of the rain data. If n be increased the functions tend to take the form of Legendre Polynomials.

4. *The Correlation of Residuals.*

When it is desired to study the correlation of variables such as annual figures, in which progressive changes are observable, it is necessary in order to obtain results which can be relied upon, and compared, in the same manner as those obtained from homogeneous material, to eliminate in some way the influence of the progressive changes. Material of this kind is very abundant and of the utmost importance; the bulk of official statistics,

vital, economic, epidemiological, meteorological and agricultural, may be said to be awaiting a method of reduction which shall deal adequately with the difficulties which this type of data presents. My investigations have not led me to any satisfactory or complete solution of the problem, but since systematic methods have been developed and strongly advocated, it would seem worth while at the present time to put forward a group of mutually connected considerations, which shed new light upon various aspects of the problem, and bring clearly into view the sources of error to which the more obvious methods of approach are exposed.

It has been observed (4, FISHER, 1915) that if the individual values of a sample be taken as the co-ordinates of a point in generalised space, the mean standard deviation and coefficient of correlation of a sample are capable of a very beautiful geometrical interpretation. For if we assign any value $\bar{x}$ as the mean of n observations $x_1, x_2, \ldots, x_n$, the equation

$$\mathrm{S}(x) = n\bar{x}$$

represents a plane section of the distribution, placed at right angles to the line

$$x_1 = x_2 = \ldots x_n$$

and meeting it at the point

$$x_1 = x_2 = \ldots = x_n = \bar{x}.$$

If this point is taken as origin, we have

$$\bar{x} = 0,$$

and the point representing any sample has rectangular co-ordinates

$$x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_n - \bar{x},$$

so that its distance from the new origin is the square root of

$$n\mu^2 = \mathrm{S}(x - \bar{x})^2$$

where μ is the standard deviation of the sample.

The mean of the observations thus specifies the origin, and the standard deviation specifies the length of the radius vector from the origin to the sample point; the correlational properties of the sample depend on the direction of this radius vector. For if a corresponding construction be made for the values of a second variate

$$y_1, y_2, \cdots, y_n,$$

and each of the axes of the y-space be brought into coincidence with the corresponding axes of the x-space, it has been shown, and it is easy to see, that the correlation coefficient between x and y is the cosine of the angle between the two *radii vectores*.

The geometrical interpretation of the correlation coefficient provides the simplest possible proof of a theorem of some importance in the present connection, namely, that: *If $x_1', x_2', \ldots x_n'$ be the co-ordinates of the first sample-point with respect to any rectilinear*

orthogonal axes through the same origin and y_1', y_2', ... y_n' the corresponding functions of $y_1, y_2 \ldots y_n$ then the correlation coefficient between x' and y' is equal to the correlation coefficient between x and y.

The system of orthogonal functions of the chief practical importance are those which arise in the elimination of progressive changes by means of a polynomial function of the time. These are the same functions as are used in analysing the sequence of weather in each particular season; it is worth noting that the above theorem is true of any orthogonal system.

In an unchanging series of independent values, *i.e.*, one in which the frequency of occurrence of any value is independent of the time, all the quantities x', except x_1', will be distributed in symmetrical curves about the mean value zero. If the values of x are normally distributed about the mean m, with standard deviation σ, then all the values of x' will be normally distributed with standard deviation σ, x_1', about a mean $m\sqrt{n}$, and the remainder about zero. This follows at once from the fact that the distribution of the points representing the sample will be a globular normal cluster, and consequently the distribution of their projections on a line drawn in any direction will be invariable. Even when the values of x are not normally distributed, those of x' will in general be much more nearly so, owing to the properties of compound distributions which have often been discussed.

The simplest type of changing series is one in which the several values are distributed independently and in similar distributions, but about a changing mean. It is with a view to eliminating slow changes in the mean, that most methods of obtaining residuals from smoothed values are employed. It is usually the case that slow changes in the mean value may be sufficiently represented by a polynomial whose degree is small compared to the number of observations in the series. In such cases the first few values of x' will be abnormally great (positive or negative), while the remainder will be normally and independently distributed. The whole group of values of x' is heterogeneous, and the correlation of two such series will have no easy interpretation; its value may be governed by the values of a few exceptional terms, and therefore its probable error will be excessive. If, however, we reject the exceptional terms and calculate the correlation from the normally distributed remainder, we shall be correlating homogeneous material and the coefficient obtained will have a simple interpretation and its usual precision. Thus if the first p values of x' suffice to specify the slow changes in the mean value, the correlation required is that obtained by ignoring the first p values of x', and y', or in other words putting them equal to zero.

This process is only an extension of the universal practice of eliminating the mean values of the variates in calculating the correlation coefficient; in working with an unchanging series we habitually ignore the values of x'_1 and y'_1. In the more general case we take

$$(n-p)\,\mu_1{}^2 = \sum_{p+1}^{n}(x'^2), \qquad (n-p)\,\mu_2{}^2 = \sum_{p+1}^{n}(y'^2), \qquad (n-p)\,\mu_1\mu_2 r = \sum_{p+1}^{n}(x'y').$$

From another point of view this may be regarded as the result of finding the partial correlation between x and y when the first $(p-1)$ powers of t, regarded as a system of correlated variates, are eliminated.

The value of r obtained in this way is, as we have shown, identical with that obtained from the mean squares and mean products of the residuals of the two series after fitting polynomials of degree $(p-1)$. The values of such residuals are indeed heterogeneous in respect of variability, and mutually connected by p equations; but since the values of x' (and equally of y') are independent and equally variable, the value of r obtained from such residuals will be distributed in random samples, exactly as is that obtained from an ordinary correlation of $n-p+1$ independent homogeneous values. The frequency distribution of r in such cases has been treated in detail (6, FISHER, '*Metron*,' 1921).

The elimination from data of obvious heterogeneity is, of course, a different process from a process which leaves the material necessarily homogeneous. The latter process is perhaps unobtainable, but it would be desirable, in cases where complications are expected, to have a means of testing whether the heterogeneity has been sufficiently eliminated. The elimination of the slowest changes is easy, since the earliest terms of the polynomial are quickly calculated. The size of the coefficients affords a criterion of rejection; thus for a record of the dressed grain from two plots on Broadbalk for 67 years I obtained (5, 1921).

Dunged plot 2B.			Plot 11 (no potash).		
Degree.	x'/μ,	P.	Degree.	x'/μ.	P.
1	−0·82	0·41	1	−4·78	0·0000018
2	−2·38	0·017	2	+0·51	0·61
3	−2·50	0·012	3	+0·20	0·84
4	−2·06	0·039	4	−0·96	0·34
5	+4·02	0·000058	5	+1·72	0·086

This first column shows the degree of the polynomial term; the second, the value of x' in terms of the standard deviation as estimated from the residuals after a curve of the 5th degree had been eliminated; the third column gives the probability of such a term occurring by chance.

Since the standard deviation is derived effectively from the sum of 61 independent squares, the exact form of the distribution of x/μ, which is STUDENT'S Type VII curve, for $n = 62$, is taken to be equivalent to the normal distribution. On the dunged plot the first term gives no grounds for elimination; no significant deterioration having occurred on this plot, and it is immaterial whether or not such a term were retained in the correlations. The remaining four terms all show the existence of a slow change

in the yield. These four terms contribute to the variance over half as much as the remaining 61 terms put together. It should be noted that while little can be lost in rejecting the first term from the correlation, nothing can be gained by retaining it, for though its smallness renders it innocuous, its claim to rank as part of the homogeneous material is overturned by the size of the following terms. On plot 11 the deterioration due to exhaustion of potash is very great, and no other term except perhaps the 5th is large enough to excite suspicion. Our grounds for eliminating the first five terms in this case are that all the plots show similar and almost proportional slow changes, and that this is doubtless the case with Plot 11, where, however, the great variability of the annual yields has masked the significance of the coefficients.

Certainty as to the homogeneity of the remaining values of x' could, I think, only be obtained by a complete calculation of their values ; in the absence of tables of the coefficients this would be very laborious. The whole set should be normally distributed about zero, and heterogeneity would be indicated by β_2 significantly exceeding its normal value, 3. In the neighbourhood of the normal curve β_2 has been shown to be the most efficient statistic for such a test (7, 1922) ; a method of calculating β_2 for the x' distribution from the actual values of x would therefore be a valuable resource, especially if any way could be found of calculating it without the separate evaluation of the values of x'. It is believed that little heterogeneity remains in the series of wheat yields from Broadbalk after fitting polynomials up to the 5th degree.

The occurrence of series of values exhibiting slow changes is so widespread that it is not surprising that a number of different methods have been prepared for dealing with the difficulty, and that some difference of opinion should exist as to the disadvantages of each. We shall show that all the methods of elucidation which have been at all fully * developed fall into one class, and that the main drawback of this class of methods is reduced to a minimum by the use of polynomials fitted to the whole series.

Slow changes may be eliminated in several ways :—

(i) by fitting a polynomial or other curve to the whole of the data ;

(ii) by the use of " smooth " values obtained by compounding a number of neighbouring terms ;

(iii) by repeatedly differencing the observed series.

In all these cases the resulting residuals designed for use as correlation variates may be successfully freed of any slow progressive trend which vitiates the original series ; but in all cases this is achieved by entangling together to some extent the successive values. This effect is inevitable, for we must judge of the smooth value from which our residual is measured by the values at neighbouring epochs.

It might appear that simple differencing should be placed in a separate class, since here no smooth value objectively appears. There is no essential difference however ; for if $2r$ is any even integer, we can construct a function,

$$v = u - (-)^r k\delta^{2r}u,$$

* For elucidation, read elimination.

which is in effect the smooth value, the deviations from which are proportional to the $(2r)^{\text{th}}$ differences; to make $\overline{v^2}$ a minimum for a series of equally variable quantities in random order we require

$$k = \frac{(2r!)^2}{(r!)^2\, 4r!}.$$

Thus the smooth value corresponding to 6th difference is

$$\begin{aligned} v_0 &= u_0 + \tfrac{5}{231}\,\delta^6 u_0 \\ &= \tfrac{1}{231}\{131u_0 + 75\,(u_1 + u_{-1}) - 30\,(u_2 + u_{-2}) + 5\,(u_3 + u_{-3})\}. \end{aligned}$$

The formula so obtained is identical with that given by SHEPPARD (15, p. 31), for 7 point smoothing, using a polynomial of the 5th degree. The variate difference method is therefore only an extreme form of the use of SHEPPARD'S smoothing formula—the extreme in which the number of terms is a minimum for given degree of the slow change eliminated—while the other extreme is represented by the process of fitting a polynomial of the required degree to the whole of the series.

In treating these three processes as special forms of SHEPPARD'S smoothing process one distinction must be made. In fitting a polynomial to the whole of the series we wish to use not only the residual of the middle term, but the whole series of residuals. So in using smooth values from (say) a 15 point formula, the first and last seven residuals may be obtained from the curves fitted to the first and last sets of 15 points. In the applications which have been made of the Variate Difference Method, only the residuals of the middle terms have been used. The number in the series has been diminished by one with each differencing; if however we wish to add the missing terms, this is easily done by means of binomial coefficients; for example if

$$a,\ b,\ c$$

are the sixth differences of a series, the three missing residuals prior to a are as follows

$$-\frac{a}{20},\ +\frac{6}{20}a,\ -\frac{15}{20}a,\ a,\ b,\ c,\ \ldots\ .$$

The effects of such processes in entangling the neighbouring terms may best be seen by considering the effect of applying them to an unchanging series of equally variable quantities. If u stand for such a quantity then, for example, the sixth difference may be written

$$v_0 = -20u_0 + 15\,(u_1 + u_{-1}) - 6\,(u_2 + u_{-2}) + (u_3 + u_{-3}),$$

hence evidently

$$\begin{aligned} \overline{v^2} &= +924\,\overline{u^2}, \\ \overline{v_p v_{p+1}} &= -792\,\overline{u^2}, \\ \overline{v_p v_{p+2}} &= +495\,\overline{u^2}, \end{aligned}$$

and so on, the numerical coefficients being those of the expansion $(1 + x)^{12}$. Consequently the correlations between neighbouring values of v will be

$$\begin{aligned} r_1 &= -\tfrac{6}{7} &&= -0{\cdot}8571, \\ r_2 &= \tfrac{5}{8}\cdot\tfrac{6}{7} &&= +0{\cdot}5357, \\ r_3 &= -\tfrac{4}{9}\cdot\tfrac{5}{8}\cdot\tfrac{6}{7} &&= -0{\cdot}2381, \\ r_4 &= &&= +0{\cdot}0714. \\ r_5 &= &&= -0{\cdot}0130, \\ r_6 &= &&= +0{\cdot}0011. \end{aligned}$$

If such high correlations as these are produced in an originally uncorrelated series, it is clear that they cannot be used without drastic correction in an examination into such correlations between neighbouring terms as exist in the original series. Equally large are the effects when two separate series are correlated. If we take, for example, a second series u', and obtain

$$v' = \delta^6 u',$$

then if the original two series were correlated so that

$$\overline{u_p u'_{p+k}} = \rho_k \sqrt{\overline{u^2} \cdot \overline{u'^2}},$$

it is easy to see that the correlations obtained from the 6th differences will be expressible in terms of the ρ series, in the form

$$r_k = \tfrac{1}{924}\,\delta^{12}\rho_k.$$

By such a method therefore we shall not obtain the true values ρ_k, but quantities proportional to the 12th difference of the series. Only if all values of ρ, except one, vanish, and we correlate the corresponding values of u and u', will we obtain an estimate of the correlation unaffected by gross inaccuracy. This constitutes a fatal objection to the applications which have been made of the variate difference method in its original form.

The same source of error still persists in more moderate degree when smoothing formulæ involving more terms are used. Let us take for example SHEPPARD'S formula for fitting a polynomial of the 5th degree to sets of 15 points. The smoothed middle point is

$$(1 - \tfrac{50}{11}\delta^6 + \tfrac{75}{13}\delta^8 + \tfrac{36}{13}\delta^{10} + \tfrac{10}{17}\delta^{12} + \tfrac{1}{3}\tfrac{5}{7}\tfrac{5}{3}\delta^{14})\,u\,;$$

whence we have the residual

$$\begin{aligned} v = \frac{1}{11\,.\,13\,.\,17\,.\,19}\{&35126u_0 - 10125(u_1+u_{-1}) - 7500(u_2+u_{-2}) - 3755(u_3+u_{-3}) \\ &+ 165(u_4+u_{-4}) - 2937(u_5+u_{-5}) + 2860(u_6+u_{-6}) - 2145(u_7+u_{-7})\}. \end{aligned}$$

The correlations between neighbouring terms of an originally uncorrelated series now become

$$\begin{aligned} r_1 &= -0{\cdot}3075 & r_8 &= -0{\cdot}0132 \\ r_2 &= -0{\cdot}2370 & r_9 &= +0{\cdot}0072 \\ r_3 &= -0{\cdot}1119 & r_{10} &= +0{\cdot}0158 \\ r_4 &= +0{\cdot}0355 & r_{11} &= +0{\cdot}0099 \\ r_5 &= +0{\cdot}1440 & r_{12} &= -0{\cdot}0027 \\ r_6 &= +0{\cdot}1211 & r_{13} &= -0{\cdot}0076 \\ r_7 &= -0{\cdot}1565 & r_{14} &= +0{\cdot}0028. \end{aligned}$$

The correlations are now much more moderate. Their alterations are less violent, and since their sum is compelled to be $-0{\cdot}5$, their actual values are permitted to be considerably smaller. Such values are, however, sufficiently large to show that great inaccuracy will be introduced if we ascertain the mutual correlations of members of a series from those of the residuals from smooth values obtained from 15 adjacent points.

It is evident also that, apart from the question of the correlation of successive values, the correlations between two series will be vitiated in the same manner as by the variate difference method, though to a less degree; the correlations obtained may be expressed in terms of high differences of the true correlations, for if we write

$$\phi(\delta^2) = \tfrac{50}{11}\delta^6 + \tfrac{75}{13}\delta^8 + \tfrac{36}{13}\delta^{10} + \tfrac{10}{17}\delta^{12} + \tfrac{15}{323}\delta^{14},$$

and

$$\phi^2(\delta^2) \text{ for } \{\phi(\delta^2)\}^2,$$

then we have

$$r_k = \frac{\phi^2(\delta^2)\rho_k}{\phi^2(\delta^2)\rho_0}.$$

The general problem of eliminating the cross correlations from a pair of series showing slow changes is extremely complex; in the simplest case each value of the one series is correlated with only two adjacent values of the second series. To this case probably belongs the relation of wheat crop to weather, for the crop is admittedly much affected by the weather in the harvest year, and may to a much less extent be influenced, through the condition of the seed, or of the soil, by the weather of the previous year. It is fortunate, therefore, that one of the examples, upon which the largest amount of work has been expended, is probably also of the same type, and will serve to indicate the order of magnitude of the errors to be expected, through neglecting the cross correlation of the crop with the weather of the previous year.

The two series of infantile death rates in the first and second years of life, are probably connected principally, if not wholly, by the fact that the mortality, in any one year, of children in the second year of life, refers to nearly the same group of children as the mortality in the previous year of children in the first year of life; and by the second fact that mortality in the two age groups during the same year will be conditioned by the

same meteorological and epidemiological conditions. If the first effect is the main object of study, the latter will appear as a cross correlation introducing errors into our estimate of the first effect, according to the method of estimation employed.

In 1915 (1) ELDERTON and PEARSON found, using the variate difference method, the value $-0\cdot688$ for the correlation between the mortalities of the same group of children (males) in the two years. In 1923 the same authors using SHEPPARD'S 15-point smoothing formula find the value $-0\cdot463$. The discrepancy is very great and suggests that the neglect of the correlation between the mortalities of the two groups of children in the same year has produced a large negative bias, which is more pronounced in the earlier estimate. Fitting a polynomial to the whole series further diminishes, though it does not eliminate, the error; for the polynomials of the 4th, 5th and 6th degrees we find the values $-0\cdot308$, $-0\cdot311$, $-0\cdot377$.

Now, assuming that only two correlations are really operative, these values may be corrected by calculating in a similar manner the apparent correlations in mortality for the two groups of children in the same year. These values are $+0\cdot4706$, $+0\cdot5456$ and $+0\cdot5217$; taking, then, the correlation between successive residuals of a series of n terms fitted by a polynomial of degree r to be $-(r+1)/(n-1)$, we have the equations

$$\text{4th degree}\begin{cases}\rho_0 - 0\cdot0877\rho_1 = -0\cdot3083\\ -0\cdot0877\rho_0 + \rho_1 = +0\cdot4706\end{cases}\qquad \rho_0 = -0\cdot2691;$$

$$\text{5th degree}\begin{cases}\rho_0 - 0\cdot1053\rho_1 = -0\cdot3109\\ -0\cdot1053\rho_0 + \rho_1 = +0\cdot5456\end{cases}\qquad \rho_0 = -0\cdot2563;$$

$$\text{6th degree}\begin{cases}\rho_0 - 0\cdot1228\rho_1 = -0\cdot3768\\ -0\cdot1228\rho_0 + \rho_1 = +0\cdot5217\end{cases}\qquad \rho_0 = -0\cdot3175.$$

The values of ρ_0 obtained from these equations are unbiassed estimates of the correlation required; they agree in indicating a correlation about $-0\cdot3$. This value may be confirmed from the figures given by PEARSON and ELDERTON for their 15-point smooth curve, which lead to the equations

$$\begin{cases}\rho_0 - 0\cdot3075\rho_1 = -0\cdot4548\\ -0\cdot3075\,\rho_0 + \rho_1 = +0\cdot6521\end{cases}\qquad \rho_0 = -0\cdot2808.$$

The concordance of these results indicates that whereas the variate difference method has exaggerated the value of this correlation to the extent of more than doubling it, the correlation of residuals from the 15-point smooth curve has reduced the error to about 50 per cent., and is moreover capable of correction provided only a few important correlations are present. The method of polynomial fitting has introduced errors of about 20 per cent. only, and the correction applied to it may be expected to be for this reason all the more precise.

In calculating the effects of weather upon the crop, it is probable, therefore, that the

effects of weather previous to the harvest year considered may be ignored. HOOKER (10, 1907) has indeed indicated that some real effect may be ascribable to previous weather, but his correlations in the case of wheat are low, and as we have seen, if the polynomials fitted to the whole series are utilised to eliminate the slow changes, the spurious correlations introduced by this cause must be much further diminished. In the case of the infantile death rates when the correlation ignored was much greater than that which we sought to evaluate, the value obtained was raised only from $-0{\cdot}26$ to $-0{\cdot}31$; in the present case we may be certain that the ignored effects are much smaller than those evaluated, and their effect, if any, on the values obtained must be extremely small.

5. *The Rothamsted Rain and Wheat Data.*

The Rothamsted wheat data employed in this investigation are derived from Broadbalk field, which has grown wheat under experimental conditions since 1844. Thirteen plots have been continuously under uniform treatment since 1852. This series is unique in its length, which as we have seen is a most important consideration for statistical purposes. A second feature of great value is that the yields are derived from measured plots by actual weighing, and are not derived from estimates based on visual observation as are the county averages published by the Ministry of Agriculture. An account of the variations in yield of dressed grain of these 13 plots has already been published (5, FISHER, 1921). It was found that the variations observable could be divided into three groups, ascribable to three separate causes. (i) On many of the plots a progressive diminution is observable owing to the exhaustion of the soil in certain of the essential plant nutrients; (ii) on all the plots slow changes in yield have taken place, which may probably be ascribed to variations in the weed infestation of the field as a whole; this variation, unlike that ascribable to other causes, is in all the plots approximately proportional to the mean yield. These two causes of variation may be eliminated by fitting to each series a polynomial of the 5th degree. The remaining variation, consisting of the deviations of the actual yield from the smooth average value given by the polynomial, is ascribable primarily to variations of the season under which the crop grew to maturity. Conventionally this season has been taken to be the weather for 366 days, ending on August 31st of the harvest year. The first two causes of variation have already been discussed; it is the purpose of the present paper to consider the fluctuations in yield from year to year in relation to the rainfall record.

Table II gives the manurial treatment of the 13 plots considered together with the mean yields obtained from them, and the mean rate of deterioration. In Table III are shown the deviations in each year from the corresponding polynomial value; it will be observed that certain years have been omitted as not available for correlation with rainfall. The rainfall record commenced in February, 1853, and is therefore complete for the crop harvested in 1854 and subsequently. In 1889 and 1890 the field was partially fallowed by halves, and the crops in 1890 and 1891 following the fallow were

beneficially affected; the same is true of the harvests of 1905, 1906 and also of 1915. It has been possible to use the value for 1916, since in this case the two halves of the field were harvested separately, and the half fallowed in 1915 could be rejected. With these omissions it is believed that the series of crop yields from this field are as homogeneous a series as it would be possible to obtain; it will be seen that 60 years are available for correlation with rainfall.

TABLE II.

Plot.	Manure per acre.						Mean (Bushels per acre).	Mean annual diminution (Bushels per acre).
	Sulphate of potash.	Sulphate of soda.	Sulphate of magnesia.	Super-phosphate.	Sulphate of ammonia.	Chloride of ammonia.		
	lb.	lb.	lb.	lb.	lb.	lb.		
2B	Dung	14 tons	—	—	—	—	34·549	0·031
3 & 4	No Manure		—	—	—	—	12·269	0·097
5	200	100	100	392	—	—	14·180	0·090
6	200	100	100	392	100	100	22·581	0·141
7	200	100	100	392	200	200	31·367	0·144
8	200	100	100	392	300	300	35·694	0·092
10	—	—	—	—	200	200	19·504	0·157
11	—	—	—	392	200	200	22·046	0·219
12	—	366½	—	392	200	200	28·319	0·181
13	200	—	—	392	200	200	30·209	0·123
14	—	—	280	392	200	200	27·765	0·231
17 *	200	100	100	392	—	—	14·510	0·092
18 *	—	—	—	—	200	200	29·006	0·114

* Alternate.

In February, 1853, readings of rainfall were commenced with a large gauge, 0·001 acre, built for the purpose; the readings of this large gauge have been consistently higher than those of the 5-in. and 8-in. gauges which are at present placed beside it. It may be concluded that the large gauge gives a better estimate of the amount of rain falling on the field. Daily readings are available for the whole period, the rain up to 9 a.m. being ascribed to the previous day. The rainfall was divided as explained in Section 3 into 61 periods of 6 days each, and each set of 61 values was then analysed by calculating the coefficients of the polynomials up to the 5th degree. Thus the amount and distribution of rain in each season was represented by a series of 6 numbers; such a representation of a complicated sequence of events might seem to be insufficient, but it has been shown in Section 3, that only such coefficients are required as correspond to the regression function; since the latter varies relatively slowly, little would be gained by following in more detail the rapid fluctuations of the weather.

The computation of the rainfall coefficients involved a great deal of labour. The method

TABLE III.

—	2B.	3 & 4.	5.	6.	7.	8.	10.	11	12.	13.	14.	17 and 18.	
												Minerals.	Ammon. Sulphate.
1854	+ 9·54	+ 6·86	+ 5·97	+ 7·62	+11·95	+13·98	+12·79	+15·83	+13·16	+12·98	+12·44	+ 6·43	+13·23
	+ 0·45	+ 1·33	− 0·42	− 0·06	− 2·93	− 5·65	− 1·14	− 7·67	− 2·61	− 2·80	− 2·62	− 0·67	+ 1·04
1856	+ 0·15	− 1·59	+ 0·54	− 1·34	− 0·75	− 0·03	− 0·07	+ 0·91	− 2·44	− 3·45	− 0·99	− 1·66	− 2·00
	+ 3·81	+ 3·89	+ 4·63	+ 5·69	+ 6·03	+ 7·82	+ 5·25	+ 8·10	+ 6·90	+ 6·98	+ 6·74	+ 6·26	+ 7·02
1858	+ 0·47	+ 1·14	− 0·08	− 1·05	− 0·52	+ 0·51	− 1·53	+ 0·63	+ 0·20	+ 0·38	+ 0·93	+ 1·39	− 0·26
	− 2·45	+ 1·45	+ 1·73	+ 0·01	− 5·04	− 7·39	− 4·70	− 4·10	− 2·84	− 2·62	− 2·85	+ 0·08	− 1·21
1860	− 6·52	− 3·44	− 2·64	− 7·61	−11·87	−10·85	− 9·93	− 8·96	−10·19	−10·56	−10·08	− 4·46	− 8·53
	− 3·79	− 5·04	− 2·74	− 1·55	− 4·26	− 6·94	−12·12	− 6·74	− 4·32	− 2·89	− 3·52	− 1·20	− 0·60
1862	+ 0·26	− 0·06	+ 0·08	− 0·32	− 2·70	− 2·20	− 2·20	− 3·74	− 2·99	− 4·21	− 5·30	− 0·92	− 5·37
	+ 6·48	+ 2·86	+ 2·50	+11·80	+15·86	+14·56	+15·69	+15·67	+17·88	+17·10	+18·00	+ 2·37	+13·66
1864	+ 3·20	+ 1·41	+ 0·21	+ 4·15	+ 8·88	+ 9·38	+ 9·05	+ 6·98	+ 9·23	+ 7·63	+ 6·28	− 0·56	+ 4·10
	+ 1·12	− 0·79	− 1·94	− 1·20	+ 4·29	+ 3·84	+ 3·36	− 1·47	+ 0·07	+ 2·26	+ 2·12	− 0·54	− 0·09
1866	− 2·53	− 1·50	− 2·34	− 4·79	+ 5·01	− 6·82	+ 3·57	− 0·05	− 5·31	− 9·35	− 5·63	− 4·18	− 4·54
	− 6·82	− 4·66	− 5·82	− 8·67	−11·75	− 7·72	− 4·37	− 5·23	− 8·14	− 9·51	−10·00	− 5·39	− 6·79
1868	+ 8·25	+ 3·87	+ 3·01	+ 4·73	+ 6·92	+ 9·20	+ 3·78	+ 6·95	+ 8·13	+ 6·68	+ 9·78	+ 3·21	+ 7·58
	+ 5·51	+ 1·83	+ 1·61	− 1·26	− 3·58	− 1·75	− 2·08	− 3·63	− 3·58	− 4·49	− 3·40	+ 1·29	− 6·38
1870	+ 4·45	+ 3·11	+ 5·00	+ 8·26	+ 9·41	+ 9·48	+ 1·32	+ 0·13	+ 5·29	+ 5·93	+ 5·38	+ 4·78	+ 5·46
	+ 7·55	− 1·99	− 1·44	− 4·60	− 8·10	− 7·36	−10·48	−13·51	− 8·08	− 0·20	− 5·34	+ 2·27	− 0·02
1872	+ 1·43	− 0·44	− 0·20	− 0·55	+ 0·01	+ 1·13	− 1·59	+ 3·36	+ 0·87	+ 0·17	+ 1·38	− 0·24	− 2·18
	− 3·80	+ 0·78	− 0·01	− 4·74	− 7·18	− 6·40	+ 0·72	− 4·01	− 4·78	− 5·75	− 4·25	− 1·03	− 7·50
1874	+ 8·97	+ 0·96	+ 0·52	+ 4·50	+10·76	+ 7·04	+ 7·29	+10·09	+12·38	+ 8·38	+ 8·76	+ 1·54	+ 5·63
	− 1·24	− 2·59	− 3·09	− 3·74	− 2·48	− 3·21	− 4·85	− 4·38	− 1·31	− 1·16	− 1·12	− 0·47	− 1·89
1876	− 6·22	− 2·03	− 1·70	− 4·22	− 4·69	− 3·30	− 4·87	− 7·59	− 6·97	− 2·85	− 5·00	− 1·55	− 0·94
	− 6·05	− 1·17	− 0·49	− 5·34	− 8·20	− 8·04	− 0·12	− 3·71	− 8·11	− 9·54	− 8·27	− 1·86	−14·80
1878	− 2·13	+ 1·92	+ 2·55	+ 2·80	+ 3·15	+ 5·39	+10·99	+ 8·27	+ 3·66	+ 1·81	+ 5·45	+ 3·47	+ 1·54
	−14·69	− 5·95	− 6·53	− 9·55	−12·00	−12·25	−13·13	− 9·94	−11·34	−11·58	−10·29	− 8·65	− 7·26
1880	+ 7·27	+ 1·54	+ 5·43	+ 7·15	+ 5·97	+ 2·30	− 5·25	+ 4·84	+ 4·03	+ 5·34	+ 4·50	+ 3·12	+ 4·99
	− 1·35	+ 2·45	+ 0·30	+ 1·12	− 2·17	− 2·58	+ 1·76	+ 0·65	− 1·56	+ 0·48	+ 0·91	+ 1·20	+ 3·93
1882	+ 0·58	+ 0·09	− 0·07	+ 2·22	+ 6·54	+ 3·44	+ 7·69	+ 9·58	+ 9·28	+ 4·25	+ 8·10	+ 3·41	+ 2·69
	+ 2·44	+ 2·56	+ 2·96	+ 6·37	+ 6·36	+ 7·83	+ 0·71	+ 5·68	+ 5·24	+ 5·92	+ 6·37	+ 3·17	− 0·46
1884	− 1·00	+ 1·83	+ 2·52	+ 4·44	+ 8·16	+ 9·02	+ 8·80	+11·35	+10·08	+ 4·58	+ 9·26	+ 0·43	+ 4·80
	+ 5·90	+ 2·63	+ 2·01	− 0·13	+ 0·74	+ 1·76	+ 7·17	+ 1·64	+ 1·46	− 1·56	− 0·23	− 0·22	+ 3·57
1886	+ 1·53	− 2·44	− 2·15	+ 0·43	+ 3·84	+ 6·83	− 4·49	− 3·51	+ 0·28	+ 8·25	+ 3·68	− 0·02	+ 7·95
	− 0·98	+ 2·48	+ 0·85	+ 0·26	− 2·47	− 1·61	+ 4·24	+ 1·14	+ 3·59	− 3·16	+ 1·27	− 3·05	+ 0·48
1888	+ 1·54	− 1·91	− 1·93	+ 0·06	+ 3·32	− 1·36	− 5·87	− 9·59	− 3·14	+ 3·48	− 1·43	− 0·93	+ 1·33
	+ 3·33	+ 0·18	+ 0·99	− 0·96	− 2·46	− 1·77	− 5·71	− 4·84	− 2·95	− 4·28	− 3·26	− 4·11	− 7·69
1890	—	—	—	—	—	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—	—	—	—	—	—
1892	− 5·61	− 3·41	− 4·46	− 2·84	− 2·57	− 0·74	+ 7·03	− 5·42	− 4·00	− 2·56	− 4·31	− 2·95	− 3·03
	− 5·18	− 2·28	− 0·71	− 5·79	−14·55	−17·45	−10·31	−13·48	−16·97	−15·29	−15·64	− 3·18	−11·87
1894	+ 5·72	+ 4·80	+ 7·33	+12·71	+13·29	+ 9·41	+11·30	+17·87	+18·61	+15·72	+15·77	+11·57	+ 4·72
	+ 3·86	− 2·50	− 0·04	− 4·22	− 2·96	+ 0·09	− 8·76	− 5·19	− 5·57	− 3·49	− 6·41	− 3·26	− 3·52
1896	+ 3·86	+ 3·39	+ 5·04	+ 4·55	+ 2·03	+ 4·15	+ 3·30	+ 3·05	+ 5·06	− 4·58	+ 1·11	+ 1·10	+ 3·25
	− 2·88	− 3·74	− 2·34	− 5·81	− 6·49	− 3·20	− 2·12	− 4·84	− 7·13	− 4·72	− 8·55	− 4·98	− 1·95
1898	− 1·99	− 0·43	− 2·16	− 4·21	− 6·51	−10·83	+ 1·65	− 3·70	− 4·12	− 5·66	− 3·50	− 0·93	− 6·27
	+ 2·78	− 0·75	− 2·18	− 5·89	− 3·28	− 0·95	+ 3·81	+ 0·92	− 0·10	− 4·27	+ 0·88	− 2·01	− 5·29
1900	− 6·06	− 0·28	− 0·04	− 4·96	− 4·30	+ 4·16	+ 0·77	− 1·77	− 4·00	− 2·72	− 3·66	− 3·78	− 2·75
	+ 0·85	− 0·51	− 0·21	− 1·44	− 3·95	+ 2·88	+ 1·76	+ 0·69	− 1·77	− 2·47	− 3·87	+ 2·40	− 2·86
1902	+ 3·41	+ 1·36	+ 1·05	+ 3·15	+ 5·50	+ 6·17	+ 5·00	+ 3·70	+ 5·20	+ 7·70	+ 6·67	+ 5·12	+ 5·85
	− 7·60	− 4·09	− 5·47	− 4·65	− 5·29	− 2·66	− 4·98	− 3·23	−10·39	− 4·71	− 8·08	− 8·75	− 4·70
1904	−14·09	− 7·00	− 6·83	−10·16	−11·51	−12·87	− 7·81	− 5·98	−11·42	− 9·67	− 7·14	−10·67	− 8·99
	—	—	—	—	—	—	—	—	—	—	—	—	—
1906	—	—	—	—	—	—	—	—	—	—	—	—	—
	+ 2·91	− 0·82	− 1·17	+ 4·78	+ 5·77	− 0·34	+10·45	+15·10	+16·01	+ 5·98	+14·58	− 1·98	+ 3·59
1908	+ 7·76	+ 2·90	+ 3·96	+ 3·76	+ 6·53	+13·54	+ 5·04	+ 4·51	+ 7·95	+ 8·00	+ 4·93	+ 2·47	+ 6·73
	+ 2·90	− 0·01	− 1·48	− 0·02	+ 3·21	− 0·56	− 5·89	− 9·89	− 4·63	+ 0·62	− 4·36	− 2·77	+ 4·39
1910	− 0·70	− 1·26	− 1·46	+ 0·93	+ 0·87	− 4·02	− 1·71	+ 4·21	+ 2·80	− 1·01	+ 3·97	− 1·41	+ 1·76
	+ 7·92	+ 3·92	+ 3·65	+ 1·16	+ 1·76	+ 5·81	+ 7·31	+ 4·67	+ 4·17	+ 4·16	+ 4·86	+ 2·82	+ 3·61
1912	− 9·77	− 3·83	− 5·25	−13·03	−16·12	−18·99	−12·68	−12·33	−17·11	−18·62	−15·95	− 4·73	−15·91
	− 5·02	− 3·10	− 3·64	− 3·89	− 6·30	− 3·26	− 5·15	− 3·77	− 4·00	− 5·79	− 2·95	− 3·37	− 2·76
1914	+ 3·65	− 2·51	− 0·62	+ 1·91	+ 0·96	+ 0·23	− 2·40	− 0·49	+ 1·50	− 4·16	− 3·85	− 5·11	− 5·19
	—	—	—	—	—	—	—	—	—	—	—	—	—
1916	− 6·70	− 3·70	− 2·98	− 4·60	− 2·22	−10·26	+ 4·11	− 0·32	− 3·55	− 2·92	− 3·81	− 5·22	− 0·94
	−10·99	− 1·48	− 1·31	+ 2·50	+ 0·17	+ 4·03	− 1·60	− 3·42	− 3·16	+ 4·39	− 1·84	+ 2·82	+ 2·12
1918	+ 6·60	− 0·03	− 2·11	− 3·19	− 1·48	+ 2·05	− 0·68	− 0·93	+ 1·32	− 0·74	+ 1·52	+ 2·11	+ 0·39

employed was that of successive summation, introduced by G. F. HARDY. By this method from a column of numbers $x_1, x_2, \ldots, x_n$ we obtain in succession

$$S_1 = x_1 + x_2 + x_3 + \ldots + x_n = S(x)$$

$$S_2 = nx_1 + (n-1)x_2 \ldots + 2x_{n-1} + x_n = S\{(n+1-r)x_r\}$$

$$S_3 = \frac{n(n+1)}{2}x_1 + \frac{(n-1)n}{2}x_2 + \ldots + 3x_{n-1} + x_n = S\left\{\frac{(n+1-r)(n+2-r)}{1.2}x_r\right\},$$

and so on. The successive summations carried as far as S_6 lead to very large numbers, but the work is straightforward and easily checked. It is not necessary to carry out the summation of the column in one piece. It may with advantage be broken in the middle ; for example in the majority of the rainfall analyses the first 29 terms were summed as above, giving

$$z_1 = \overset{29}{\underset{1}{S}}(x_r)$$

$$z_2 = \overset{29}{\underset{1}{S}}\{(30-r)x_r\}$$

$$z_3 = \overset{29}{\underset{1}{S}}\left\{\frac{(30-r)(31-r)}{1.2}x_r\right\} \text{ and so on ;}$$

while the remaining 32 values were summed backwards from the bottom of the column, dropping one term at the end of each summation, and so giving

$$z_1' = \overset{61}{\underset{30}{S}}(x_r)$$

$$z_2' = \overset{61}{\underset{31}{S}}\{(r-30)x_r\} = -\overset{61}{\underset{30}{S}}\{(30-r)x_r\}$$

$$z_3' = \overset{61}{\underset{32}{S}}\left\{\frac{(r-30)(r-31)}{1.2}x_r\right\} = \overset{61}{\underset{30}{S}}\left\{\frac{(30-r)(31-r)}{1.2}x_r\right\} \text{ and so on.}$$

Taking now alternately sums and differences, we obtain

$$S_1' = z_1 + z_1' = \overset{61}{\underset{1}{S}}(x_r)$$

$$S_2' = z_2 - z_2' = \overset{61}{\underset{1}{S}}\{(30-r)x_r\}$$

$$S_3' = z_3 + z_3' = \overset{61}{\underset{1}{S}}\left\{\frac{(30-r)(31-r)}{1.2}x_r\right\} \text{ and so on ;}$$

from which the final sums of the whole column may be obtained from the equations

$$S_1 = S_1'$$

$$S_2 = S_2' + 32S_1'$$

$$S_3 = S_3' + 32S_2' + \frac{32.33}{1.2}S_1'.$$

The series $S_1, S_2, S_3, \ldots\ldots$, by whichever method it is obtained, is divided by the series of divisors

$$61, \quad \frac{61 \cdot 62}{1 \cdot 2}, \quad \frac{61 \cdot 62 \cdot 63}{1 \cdot 2 \cdot 3} \quad \text{and so on,}$$

yielding a series of numbers of similar magnitude

$$a = \frac{1}{61} S(x_r)$$

$$b = \frac{1 \cdot 2}{61 \cdot 62} S\{(n+1-r)\, x_r\}$$

$$c = \frac{1 \cdot 2 \cdot 3}{61 \cdot 62 \cdot 63} S\left\{\frac{(n+1-r)(n+2-r)}{1 \cdot 2} x_r\right\}$$

with similar equations for d, e, and f. These may be regarded as weighted means of the series $x_1, \ldots, x_n$; it will be observed that the weights involve r, and therefore the time, up to the 5th degree. The series is therefore equivalent for the purpose of evaluating the polynomial coefficients to the first five moments of rainfall as distributed in time. The values at which we have arrived are, however, more convenient for computing purposes than the moments would be, for the quantities required by the method of Section 3 are

$$\rho_0 = S(xT_0) = \frac{1}{\sqrt{n}} S(x) = \sqrt{n}\, a$$

$$\rho_1 = S(xT_1) = \sqrt{\frac{12}{n(n^2-1)}}\, S(tx) = \sqrt{\frac{3n(n+1)}{n-1}}\,(a-b)$$

$$\rho_2 = S(xT_2) = \sqrt{\frac{180}{n(n^2-1)(n^2-4)}}\, S\left\{\left(t^2 - \frac{n^2-1}{12}\right) x\right\}$$

$$= \sqrt{\frac{5n(n+1)(n+2)}{(n-1)(n-2)}}\,(a - 3b + 2c) \text{ and so on.}$$

The factors under the square root are not necessary for the correlational work, so that actually it is only necessary to calculate for each year

$$\left.\begin{aligned} a' &= a \\ b' &= a - b \\ c' &= a - 3b + 2c \\ d' &= a - 6b + 10c - 5d \\ e' &= a - 10b + 30c - 35d + 14e \\ f' &= a - 15b + 70c - 140d + 126e - 42f \end{aligned}\right\} \quad \ldots\ldots \quad (V)$$

These are the independent variates in terms of which the wheat yield is to be expressed

in the form of a regression equation. The regression coefficients of the yield upon these will have to be divided by factors of the form

$$\sqrt{\frac{(2s+1)\,.\,n(n+1)\,\ldots\,(n+s)}{(n-1)\,\ldots\,(n-s)}}$$

in order to give the coefficients α_s of Section 3, but as in the expansion of the regression function, α_s is multiplied by

$$\sqrt{\frac{(2s+1)\,.\,((2s)\,!\,)^2}{(s\,!)^4\,n(n^2-1)\,\ldots\,(n^2-s^2)}}$$

we actually multiply the regression coefficients by

$$\frac{(2s)\,!}{(s\,!)^2\,n\,(n+1)\,\ldots\,(n+s)}$$

in order to obtain the coefficients of the polynomial,

$$t^s + \ldots,$$

in the expansion of the regression function.

6. *The Rain Distribution Values.*

The values actually obtained for the quantities a', ..., f', for the 65 periods ending August 31st, 1854 to 1918, are given in Table IV; for tabulation they have been multiplied by 1,000; thus the value 347 in the second column shows that the average rainfall for the first period was 347 thousandths of an inch every 6 days. The figures in the third column measure the average rate at which rainfall was increasing or decreasing during the period, as indicated by a straight line fitted to the recorded values; c' measures the parabolic term in the rainfall sequence, and so on taking more and more complex features of the distribution into account.

These values are themselves of very great interest, since the incidence of rainfall has not previously been analysed in this way.

The individual peculiarities of the successive seasons are brought out clearly, and it is possible to examine the sequence of years by adequate statistical methods. We may first enquire whether the observed sequence accords with the view that each season is an independent product of random causes under constant climatic conditions, or whether on the other hand the sequence indicates progressive changes in the quantity and distribution of the rainfall.

The series of 65 values of a' was therefore analysed by summation in the same manner as the 61 values of the rainfall record of each year had been treated; as an illustration

TABLE IV.

Harvest Year.	a'	b'	c'	d'	e'	f'	Harvest Year.	a'	b'	c'	d'	e'	f'
1854	347	− 14	+ 33	− 16	− 21	− 1	1887	387	− 82	− 6	+ 29	− 18	+14
1855	398	+ 90	+ 41	+ 8	− 35	− 22	1888	500	+ 48	+ 59	− 20	− 10	−13
1856	487	− 18	+ 12	+ 1	− 27	+ 35	1889	498	+ 72	− 1	− 11	− 36	−10
1857	404	− 9	+ 19	+ 15	− 9	− 1	1890	450	+ 20	+ 35	− 1	− 15	−12
1858	441	− 69	+ 69	− 38	− 12	+ 15	1891	383	+ 65	+ 35	+ 22	0	+28
1859	412	+ 64	+ 3	− 14	− 4	− 5	1892	487	− 29	+ 70	+ 45	− 35	+29
1860	598	+ 39	+ 35	− 12	− 12	+ 11	1893	398	− 38	+ 37	+ 26	− 16	− 4
1861	381	− 8	+ 3	− 17	− 24	+ 1	1894	488	− 7	+ 24	+ 30	− 8	+16
1862	451	+ 22	− 13	− 13	− 21	+ 18	1895	474	− 3	+ 66	+ 51	− 42	−11
1863	414	− 21	+ 36	+ 19	− 7	+ 14	1896	399	− 27	+ 21	+ 32	− 4	+13
1864	331	− 56	+ 16	− 24	+ 3	+ 3	1897	618	−131	+ 77	− 32	+ 55	− 5
1865	454	+ 43	+ 61	− 30	+ 29	− 11	1898	323	− 11	+ 9	− 20	+ 4	−17
1866	568	− 2	+ 3	+ 38	− 45	+ 8	1899	405	− 20	− 42	+ 27	− 17	+ 5
1867	513	− 14	+ 21	− 27	+ 7	− 17	1900	514	− 47	+ 27	+ 35	+ 4	+ 2
1868	346	− 11	+ 14	+ 26	+ 42	− 11	1901	406	− 10	− 7	+ 31	+ 2	+11
1869	426	− 44	− 37	+ 10	− 3	− 3	1902	382	+ 30	+ 18	+ 12	− 5	+ 2
1870	354	− 65	+ 10	+ 21	+ 9	− 11	1903	520	+106	+ 28	− 12	− 8	−16
1871	451	− 7	+ 16	− 23	− 32	− 16	1904	517	− 75	+ 30	+ 32	+ 2	+ 7
1872	475	+ 8	+ 13	− 23	+ 5	− 6	1905	428	+ 35	+ 58	0	+ 17	+11
1873	503	− 57	+ 6	+ 64	− 31	+ 2	1906	389	− 37	− 13	+ 11	− 9	−15
1874	357	− 12	+ 30	− 10	− 3	− 6	1907	481	− 44	+ 13	+ 5	− 47	+24
1875	526	− 11	+ 42	− 33	− 26	− 42	1908	494	− 5	+ 6	+ 24	− 2	+38
1876	524	− 80	+ 29	+ 18	− 8	+ 42	1909	419	+ 52	+ 26	− 26	− 5	− 7
1877	648	− 76	+ 9	+ 22	+ 23	− 38	1910	507	− 16	+ 16	+ 23	− 11	+ 5
1878	532	+ 38	+ 10	+ 20	+ 1	+ 35	1911	465	− 43	− 38	+ 25	− 35	+11
1879	674	+105	+ 37	+ 26	− 35	− 3	1912	672	+ 18	+ 33	+109	− 9	− 2
1880	350	+ 39	+ 25	− 29	− 7	− 24	1913	451	− 54	− 35	+ 20	− 2	+18
1881	603	− 78	+110	+ 35	+ 41	+ 16	1914	419	− 42	− 14	− 10	+ 8	−57
1882	530	− 11	− 8	+ 5	− 28	+ 15	1915	604	− 14	− 36	+ 65	− 29	−55
1883	573	− 81	+ 6	+ 25	− 39	− 14	1916	582	− 17	− 26	+ 51	+ 15	+14
1884	422	− 53	+ 35	+ 1	0	+ 8	1917	577	+ 54	+ 82	+ 58	− 25	− 2
1885	434	− 20	− 37	− 11	+ 9	0	1918	449	− 18	+ 15	− 22	− 11	+23
1886	508	− 83	+ 33	− 30	− 13	0							

of the method the numerical values are given in Tables V and VI; before summation each a' was reduced by 400, the true value of the mean being inserted in the fourth column. The second column gives the successive sums, S_1 to S_6; from these the third column is derived by dividing by factors of the form

$$\frac{65 \cdot 66 \ldots (65 + r)}{r!}, \qquad r = 1, \ldots, 5.$$

Thus the third column gives the values of a to f for the series a'. The fourth column gives the corresponding values a' to f' for the series a', obtained by equations (V); while the fifth column, obtained by multiplying by factors of the form

$$\sqrt{\frac{(2r+1) \cdot 65 \ldots (65 + r)}{64 \ldots (65 - r)}}.$$

are the actual values of the first five transformed co-ordinates spoken of in Section 4, as x_2' to x_6'.

TABLE V.—Analysis of sequence of values of a'.

—	S_1 to S_6.	a to f.	a' to f'.	x_2' to x_6'.
1	4,521	69·55385	469·55385	
2	128,341	59·832634	+9·72122	+137·85
3	2,603,055	54·3378562	−1·26834	− 23·95
4	40,090,879	49·22841040	+7·79456	+182·35
5	500,901,859	44·570118750	+2·35049	+ 66·31
6	5,313,933,165	40·528492120	−2·62489	− 88·43

In an unchanging series the values x_2', x_3' ... vary about zero in an approximately normal distribution, the standard deviation of which may be obtained from that of the original series; for

$$\sum_{2}^{n} (x_r'^2)$$

is the sum of the squares of the deviations of the original series from their mean. Slow changes in the original series will be indicated by high positive or negative values in x_2', x_3', ..., and if such slow changes are suspected, it will be better to estimate the variance due to random causes from

$$\sum_{7}^{n} (x_r'^2),$$

from which the first five values have been omitted. From the sums of the squares of the deviations we may thus obtain a series of values each obtained from the last by deducting the square of the corresponding value x_r'; from each such sum may be obtained an estimate of the standard deviation due to random causes, by dividing by the number of squares concerned (degrees of freedom), and taking the square root. Such estimates will be equivalent to those derived from the residuals left after polynomials of the first to the fifth degree have been successively fitted; but the labour of calculating the polynominal values is avoided.

TABLE VI.

Degrees of freedom.	Sum of squares.	Mean square.	Standard deviation.
64	465,105	7267·3	85·25
63	446,103	7081·0	84·15
62	445,529	7186·0	84·77
61	412,278	6758·7	82·21
60	407,881	6798·0	82·45
59	400,061	6780·7	82·35