https://doi.org/10.29321/MAJ.10.A05027

Farming will never be a success unless the farmer
had more voice in the disposal of
his produce.—P. Morrel.

The
Madras Agricultural Journal
(ORGAN OF THE M. A. S. UNION)

Vol. XIX] OCTOBER 1931 [No. 10

A STATISTICAL STUDY OF PADDY YIELD IN SOUTH KANARA IN RELATION TO WEATHER

By N. K. ADYANTHAYA, B.A. Hons., B.Sc. (Lond.)

Lecturer in Statistics, University of Madras

'We have to contemplate social phenomena as susceptible of prevision, like all other classes within the limits of exactness compatible with their higher complexity.'

—*Auguste Comte.*

The object of this paper is to point out the possibilities of the application of the modern theory of statistics in the fruitful field of Agricultural Meteoroiogy. As an example, a study of yield in relation to weather will be taken up. In this connection it is worth noticing that in 1927, Mr. J. H. Field, the then Director-General of Observatories, while giving his evidence before the Royal Commission on Agriculture in India, stressed the importance of the service which the Meteorological department can render to Indian agriculture by researches on the correlation between weather and crops. How the computed coefficients of correlation between yield and certain meteorological elements enable us to predict the yield in a certain year or period from a knowledge of the conditions of weather in that year or period, will be clear from the treatment of the topic taken up below.

To secure homogeneity in meteorological conditions which is an important factor in a work of this nature, I have taken up the district of South Kanara on the West Coast, where we have a monsoon lasting for nearly three months and a half from about the beginning of June up to nearly the middle of September, and the crops are mostly unirrigated, thereby ensuring that rainfall, and not irrigation, is the factor entering directly into the determination of yield. In Table I are set out the data of the yield of paddy in tons, the area under cultivation, the total rainfall, and the average temperature for the month of June for the twenty-three years from 1907–08 up to 1929–30. With the sole exception of the last, these data were taken from the Season and Crop Reports of the Madras Presidency issued every year by the Department of Agriculture.

TABLE I

Study of Paddy Yield in South Kanara for a period of 23 years.

Year		Total estimated yield in tons	Area under cultivation in acres	Total rainfall in inches	Average June temperature
1907–08	...	277,763	555,279	158·79	80·3
1908–09	...	277,736	555,225	181·43	80·6
1909–10	...	279,729	559,208	135·99	79·7
1910–11	...	307,308	552,908	142·49	77·3
1911–12	...	309,546	556,935	124·19	79·3
1912–13	...	319,292	574,470	162·22	80·1
1913–14	...	314,519	565,881	123·73	80·8
1914–15	...	318,833	573,643	173·86	80·6
1915–16	...	329,352	592,569	122·56	82·3
1916–17	...	390,792	595,493	153·32	78·9
1917–18	...	387,062	602,096	186·69	78·9
1918–19	...	317,000	584,508	89·85	79·7
1919–20	...	399,100	595,919	138·74	80·9
1920–21	...	414,800	595,596	140·45	79·8
1921–22	...	401,300	593,356	156·60	80·9
1922–23	...	407,400	596,398	158·20	79·5
1923–24	...	373,900	581,689	166·00	81·3
1924–25	...	373,800	581,430	174·60	79·0
1925–26	...	373,000	580,245	164·50	80·4
1926–27	...	384,100	579,362	137·80	80·2
1927–28	...	389,050	580,983	150·30	81·0
1928–29	...	381,500	575,429	144·70	79·8
1929–30	...	386,600	577,308	164·00	78·6

It is not easy to deal with the figures as they stand. So indices for the four factors under consideration were constructed, choosing as base the average for the five years from 1912–13 up to 1916–17, and taking it as 100. The revised table is given hereunder.

TABLE II

Year		Index of yield	Index of area under cultivation	Index of rainfall	Index of temperature
1907–08	...	83·01	95·67	107·91	99·75
1908–09	...	83·02	95·66	123·31	100·12
1909–10	...	83·61	95·74	92·42	99·01
1910–11	...	91·85	95·26	96·84	96·02
1911–12	...	92·52	95·96	84·40	98·51
1912–13	...	95·44	98·98	110·25	99·50
1913–14	...	94·01	97·50	84·09	100·37
1914–15	...	95·30	98·83	118·16	100·12
1915–16	...	98·44	102·09	83·30	102·24
1916–17	...	116·81	102·60	104·20	98·01

TABLE II—(continued)

Year	Index of yield	Index of area under cultivation	Index of rainfall	Index of temperature
1917–18 ...	115·69	103·74	126·88	98·01
1918–19 ...	94·75	100·71	61·07	99·01
1919–20 ...	119·29	102·67	94·29	100·50
1920–21 ...	123·98	102·62	95·45	99·13
1921–22 ...	119·95	102·23	106·43	100·50
1922–23 ...	121·77	102·75	107·52	98·76
1923–24 ...	111·76	100·22	112·82	100·99
1924–25 ...	111·73	100·18	118·66	98·13
1925–26 ...	111·49	99·97	111·80	99·88
1926–27 ...	114·81	99·82	93·65	100·62
1927–28 ...	116·29	100·10	102·15	99·13
1928–29 ...	114·03	99·14	98·34	99·63
1929–30 ...	115·56	99·47	111·46	97·64

Now, how are we to correlate these indices ? Can we correlate them as they stand, and does the resulting coefficient of correlation give a measure of the true relationship between the characteristics correlated ? The answer is 'No'. Why ? Because the data we are dealing with, which are really time series, are subject to cyclical and secular influences. In order to measure the relationship between yield and the variations of the weather, we must make allowance for these secular and cyclical changes. We can here follow the method adopted by the Bureau of Statistics of the Department of Agriculture of the United States of America for tackling this difficulty.[1] In their forecasts of the probable yield of the crop at the end of the year from the condition of the crop in any month, they have found by experience that it is not profitable to work directly with the absolute values of the condition and yield on account of the violent fluctuations to which a result on this basis is subject at times ; they have, on the other hand, found that if they took the ratio of the condition figure in any particular year to the average condition figure for the five or three previous years and equated this to the ratio of probable yield to the average yield for the five or three previous years, or symbolically if they took $\frac{C}{\overline{C_5}} = \frac{Y}{\overline{Y_5}}$ where $\overline{C_5}$ and $\overline{Y_5}$ are the average condition and yield for the five previous years, this gave a series from which the effects of cyclical and secular changes were very much mitigated. We are going to adopt the same method here, and instead of correlating the indices directly, we shall correlate the ratio of an index for any year to the average index for three previous years with the ratios of other indices similarly treated. Since our data extend only for a period of twenty-three years, we shall be getting twenty items as against eighteen if we had taken a five-year average, and since there is not much to choose between a three-year and a five-year period average, the index ratios to three-year averages are to be preferred. Denoting yield by Y, area under cultivation by A, the total rainfall by R, and average temperature for the month of June by T, and the average yield for the three

[1] Forecasting the yield and price of cotton, by H. L. Moore.

years previous to the year in question by $\overline{Y_3}$, the average area by $\overline{A_3}$ etc., the data in the new form can be set out as follows :—

TABLE III

Year		$\frac{Y}{\overline{Y_3}}$	$\frac{A}{\overline{A_3}}$	$\frac{R}{\overline{R_3}}$	$\frac{T}{\overline{T_3}}$
1907-08	...	...	...	...	...
1908-09	...	...	...	...	...
1909-10	...	...	...	...	...
1910-11	...	110·4	99·6	89·8	96·4
1911-12	...	107·4	100·4	81·0	100·1
1912-13	...	106·8	103·5	120·9	101·7
1913-14	...	100·8	100·8	86 5	102·4
1914-15	...	101·4	101·4	127·2	100·7
1915-16	...	103·7	103·7	80·0	102 2
1916-17	...	121·8	103·1	109·5	97·1
1917-18	...	111·8	102·5	124·5	97·9
1918-19	...	85·9	98·0	58·3	99·6
1919-20	...	109·4	100·3	96·8	102·2
1920-21	...	112·8	100·2	101·5	100·0
1921-22	...	106·5	100·2	127·3	101·0
1922-23	...	100·6	100·2	108·9	98·7
1923-24	...	91·7	97·7	109·4	101·5
1924-25	...	94·8	98·5	108·9	98·0
1925-26	...	96·9	98·9	98·9	100·6
1926-27	...	102·8	99·7	81·8	101·0
1927-28	...	103·2	100·1	94·6	99·6
1928-29	...	99·9	99·2	95·9	99·8
1929-30	...	100·4	99·8	113·7	97·8

Calculating the coefficients of correlation between the four indices taken in pairs, the coefficients come out as shown below :—

$$r_{\frac{Y}{\overline{Y_3}},\frac{A}{\overline{A_3}}} = \cdot 6688\,; \quad r_{\frac{Y}{\overline{Y_3}},\frac{R}{\overline{R_3}}} = \cdot 3198, \quad r_{\frac{Y}{\overline{Y_3}},\frac{T}{\overline{T_3}}} = -\cdot 2706,$$

$$r_{\frac{A}{\overline{A_3}},\frac{R}{\overline{R_3}}} = \cdot 2901\,; \quad r_{\frac{A}{\overline{A_3}},\frac{T}{\overline{T_3}}} = \cdot 0944, \quad r_{\frac{R}{\overline{R_3}},\frac{T}{\overline{T_3}}} = -\cdot 1482$$

Here $r_{\frac{Y}{\overline{Y_3}},\frac{A}{\overline{A_3}}}$ denotes correlation between ratios $\frac{Y}{\overline{Y_3}}$ and $\frac{A}{\overline{A_3}}$,

$r_{\frac{Y}{\overline{Y_3}},\frac{R}{\overline{R_3}}}$ denotes correlation between $\frac{Y}{\overline{Y_3}}$ and $\frac{R}{\overline{R_3}}$ and so on.

From these values, we see that the correlations between yield and area and yield and rainfall respectively are significant, as we should naturally expect them to be. There is a higher correlation between yield and area than between yield and rainfall, and this is easily explained when we remember that too much of rains may be injurious to crops, while on the other hand, other things being equal, there is nothing to hinder the yield from progressive increase, so long as the area under cultivation also increases. The coefficient of correlation between area and rainfall is probably a bit significant, and so is the correlation between yield and temperature which is negative, thereby showing that a lower June temperature coupled with rains, is conducive to better yield than a high temperature. The standard error of the correlation between area and rainfall is $\frac{\cdot 9147}{\sqrt{20}} = \cdot 20453$ and therefore the ratio of the significance of the observed coefficient of correlation from zero will be nearly 1·4, which shows that the observed coefficient is not quite significant. The same may be said about the correlation between yield and temperature. The other two correlations are not significant.

Now, how will these values help us? While too much reliance may not be placed on a small sample of 20 items, still they enable us to answer some important questions. A question like this may be asked: 'How will you predict the yield in a certain year, if you are given the total rainfall for the year?' Since we know the correlation between yield and rainfall, and also their means and standard deviations, the prediction equation is at once written down in the form

$\frac{Y}{\overline{Y_3}} - 103{\cdot}45 = \cdot 3198 \times \cdot 44923 \left(\frac{R}{\overline{R_3}} - 100{\cdot}76 \right)$. Here, ·44923 is the ratio of the standard deviation of $\frac{Y}{\overline{Y_3}}$ to the standard deviation of $\frac{R}{\overline{R_3}}$. 103·45 and 100·76 are the means of these indices. This equation reduces to $\frac{Y}{\overline{Y_3}}$

$$= \cdot 1437 \frac{R}{\overline{R_3}} + 88{\cdot}97,$$

which is an equation which determines Y from a knowledge of R. We can also make use of the concept of partial correlation, and answer a question like this: 'What effect has the June temperature on the yield, supposing the rainfall to remain constant'? We have simply to find the correlation between $\frac{Y}{\overline{Y_3}}$ and $\frac{T}{\overline{T_3}}$ for constant $\frac{R}{\overline{R_3}}$ or with the usual notation, we have to find $r_{\frac{Y}{\overline{Y_3}}, \frac{T}{\overline{T_3}} \cdot \frac{R}{\overline{R_3}}}$. This is equal to

$$\frac{r_{\frac{Y}{\overline{Y_3}}, \frac{T}{\overline{T_3}}} - r_{\frac{Y}{\overline{Y_3}}, \frac{R}{\overline{R_3}}} \times r_{\frac{T}{\overline{T_3}}, \frac{R}{\overline{R_3}}}}{\sqrt{1 - r^2_{\frac{Y}{\overline{Y_3}}, \frac{R}{\overline{R_3}}}} \sqrt{1 - r^2_{\frac{T}{\overline{T_3}}, \frac{R}{\overline{R_3}}}}} = - \cdot 2382$$

which is negative, and therefore we conclude that a lower June temperature is conducive to a better yield than a higher one, rainfall remaining constant. The correlation, though negative is not quite significant. So we cannot place too much reliance on this result. There is no doubt that there is a certain normal temperature which gives the best yield, other things being equal. Any temperature in excess or in defect of that may be injurious to crops.

Our object is not so much to study the relationship between yield and the individual factors as to combine the various factors which we have taken as affecting the yield, and arrive at a formula from which we are able to predict the probable yield in a certain year from a knowledge of the three factors which we have considered—the area under cultivation, the total rainfall, and the average June temperature. The theory of multiple correlation developed by Professor Karl Pearson and his co-workers enables us to solve this very important problem. We need not here go into the complex theory of multiple correlation, but we may just state some of the results, which may help an investigator engaged on a work of this nature. If x_0 stand for yield, x_1 for area, x_2 for rainfall and x_3 for temperature, the problem of multiple correlation is to find a linear function in x_1, x_2, x_3, that has the maximum correlation with x_0. From this function, the prediction equation is readily written down, and it will be

$$\frac{x_0 - \bar{x}_0}{\sigma_0} = -\frac{R_{01}}{R_{00}}\frac{x_1 - \bar{x}_1}{\sigma_1} - \frac{R_{02}}{R_{00}}\frac{x_2 - \bar{x}_2}{\sigma_2} - \frac{R_{03}}{R_{00}}\frac{x_3 - \bar{x}_3}{\sigma_3}$$

where $\sigma_0, \sigma_1, \sigma_2, \sigma_3$, are the standard deviations of x_0, x_1, x_2, x_3, and R is the symmetrical determinant

$$\begin{vmatrix} 1 & r_{01} & r_{02} & r_{03} \\ r_{10} & 1 & r_{12} & r_{13} \\ r_{20} & r_{21} & 1 & r_{23} \\ r_{30} & r_{31} & r_{32} & 1 \end{vmatrix}$$

and r_{01}, r_{02}, etc., are the correlation between yield and area, yield and rainfall and so on; and R_{00}, R_{01}, R_{02}, etc. are the minors of the determinant R.

The symmetrical determinant to take in this case will be

$$\begin{vmatrix} 1 & \cdot6688 & \cdot3198 & -\cdot2706 \\ \cdot6688 & 1 & \cdot2921 & \cdot0944 \\ \cdot3198 & \cdot2921 & 1 & -\cdot1482 \\ -\cdot2706 & \cdot0944 & -\cdot1482 & 1 \end{vmatrix}$$

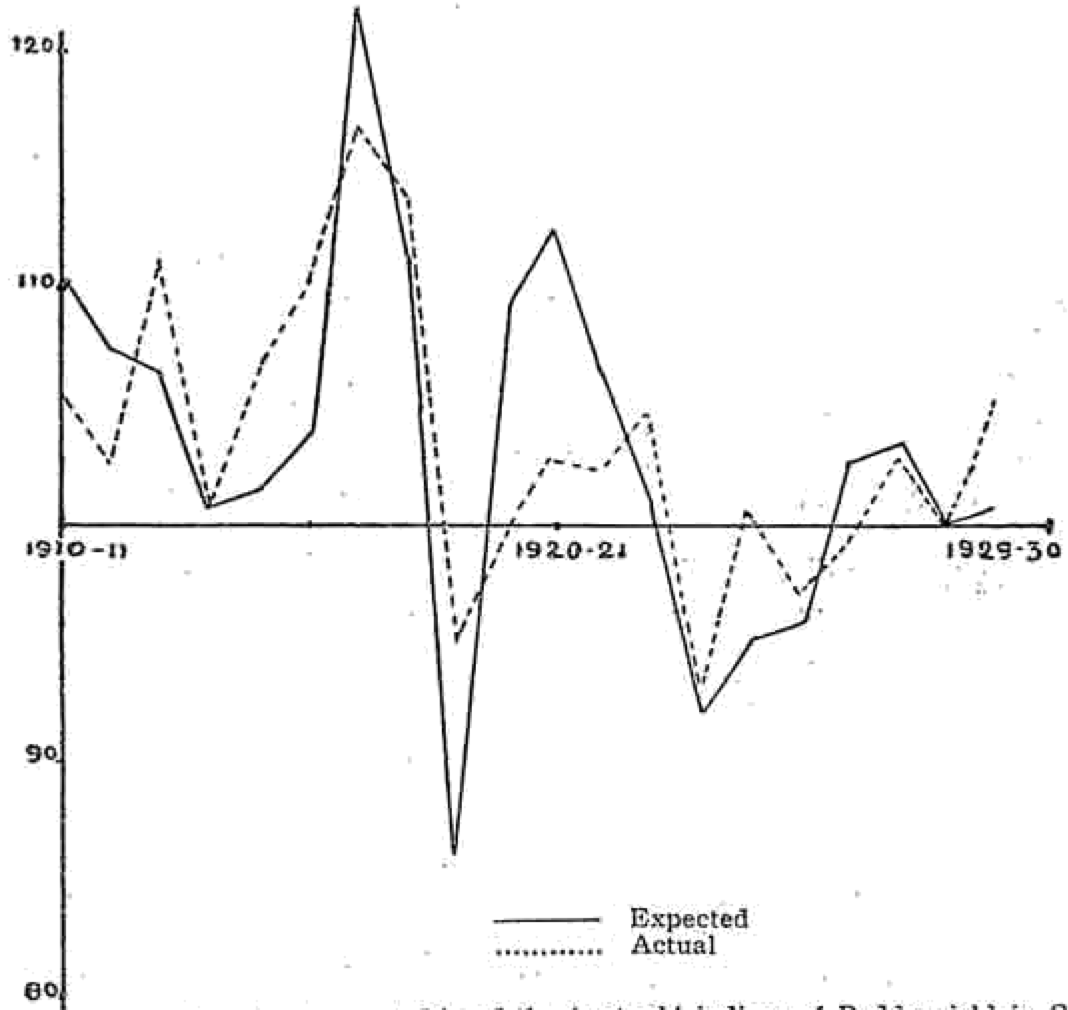
and computing the various minors of this determinant, the prediction equation for the yield x_0 will be found to be

$$x_0 = 3\cdot179\, x_1 + \cdot033\, x_2 - 1\cdot459\, x_3 - 73\cdot25$$

Calculating the values of x_0, for the various values of x_1, x_2, x_3, given in Table III, the expected series of values may be set out against the *actual* as follows :—

Years	Expected	Actual	Difference
1910–11	105·7	110·4	−4·7
1911–12	102·5	107·4	−4·9
1912–13	111·4	106·8	4·6
1913–14	100·7	100·8	−0·1
1914–15	106·4	101·4	5·0
1915–16	109·9	103·7	6·2
1916–17	116·7	121·8	−5·1
1917–18	113·9	111·8	2·1
1918–19	94·9	85·9	9·0
1919–20	99·7	109·4	−9·7
1920–21	102·7	112·8	−10·1
1921–22	102·1	106·5	−4·4
1922–23	104·9	100·6	4·3
1923–24	92·9	91·7	1·2
1924–25	100·5	94·8	5·7
1925–26	97·7	96·9	0·8
1926–27	99·0	102·8	−3·8
1927–28	102·8	103·2	−0·4
1928–29	99·7	99·9	−0·2
1929–30	105·1	100·4	4·7

From the above figures we see that the correspondence between theory and observation is pretty close. The standard deviation of an estimate based on this multiple regression method is $\sigma_0\sqrt{\frac{R}{R_{00}}}$ and this works out as 5·1527. In no case do we observe a departure of the expected from the observed by three times the S.D. The departure is a bit exceptional in only three cases in the three consecutive years, 1918–19, 1919–20 and 1920–21. The departure here comes to nearly twice the S.D. in each case. But when we remember that these were rather exceptional years in that there was a run of drought beginning with year 1918 and running up to the year 1921, we see how hard it is for any equation to give an adequate fit at these three points. 1918 was an exceptionally bad year, and the index of rainfall for that year is 61·07, the lowest in the whole period. 1915–16 was also a pretty bad year from the point of view of rainfall and there also the departure is a bit noticeable. The correspondence between the actual and the expected series can be better appreciated from the graph shown below.



Graph showing the 'expected' and the 'actual' indices of Paddy yield in South Kanara, for the twenty years from 1910-11 to 1929-30. The dotted lines, show the expected series.

Before we close, it is fitting that we should make a statement of the limitations of the method which we have employed, and, if possible, suggest a more suitable method of tackling a problem of this kind, if the nature of the data, and the special circumstances of the problem, admit of a better solution. The method we have adopted is not strictly correct, for the correlations we have found between ratios, the ratios having been derived from progressive three-year averages, may be, after all, spurious, and may not be representative of the presence of real causes in the characteristics correlated. Until about the year 1910, statistical science had no means of satisfactorily correlating time series; this problem puzzled mathematicians for a long time, and a method has been gradually elaborated, and we may say that it holds the field now. That method has been termed 'The Variate Difference Method'. Here the correlation between first differences, between second differences, third differences and so on, of the variates are calculated until the calculated coefficients exhibit signs of stability in sign and magnitude. When this is attained, the resulting coefficient is taken as a measure of the true relationship between the variates. This of course would be the ideal method to use in any practical case, and will no doubt yield very valuable results in a case like the one we considered, where a knowledge of the real underlying causes and the extent of their inter-relationships is essential. But if the series extend only to about twenty years or less, we shall be left with very few items for purposes of correlation, and it would be hazardous to base any arguments or build any theory on the basis of these few items. Moreover, we have to go on taking differences and correlating them until stability in correlation is attained. The series may be nearly exhausted by the time we reach the 6th difference. This was the very consideration which prompted me in the adoption of the alternative method of attack in this particular case. Professor Moore in his paper 'Forecasting the Yield and Price of Cotton' tried the same method, and verified his results by taking the correlation of differences also, and found a very close correspondence between the two. But, if the data available extend for a good number of years, the adoption of 'Variate Difference Correlation Method' would, of course, be ideal.

BIBLIOGRAPHY

ng the Yield and Price of Cotton.

Jacob, S. M. (1910). On the correlations of Areas of Matured Crops and the Rainfall. *Mem. Asiatic Soc. Bengal*, Vol. II.

Hooker, R. H. (1907). The correlation of the Weather and the crops. *Jour. Roy. Stat. Soc.*, Vol. LXX, 1907.

Biometrika. Vol. X, Parts II and III.

THE M.A.S.U. PARLIAMENT

(*Proceedings of a Meeting held on 23rd July 1931*)

Motion before the House :—'That this House is of opinion that the export of indigenous manures from India is not conducive to the welfare of the country.'

Proposer :—Rao Bahadur B. Viswa Nath Garu, F.I.C., Government Agricultural Chemist, Coimbatore.

Opposer :—C. E. Wootton, Esq., B. Sc., Chemist, Manure Works, Messrs. T. Stanes & Co., Ltd., Coimbatore.

Speaker :—Rao Bahadur M. R. Ramaswami Sivan Avl., B.A., DIP. AGRI., Retired Principal, Agricultural College, Coimbatore.

SPEAKER'S OPENING REMARKS

LADIES AND GENTLEMEN,

It is a very great pleasure for me to be at the scene of my labours again, and I am glad that the M.A.S.U. with whose inception I was intimately connected 25 years ago, has added to its activities by these parliamentary ventures. I heard that the first was extraordinarily successful, and from the kind of people who are taking part to-day I am sure this will be equally so. The subject itself affords much room for discussion; so much has to be said for and against, so much has been discussed in Board of Agriculture meetings, Legislative Council discussions and in conferences. And there is plenty of room for expression of different opinions. It is very convenient for me as