INVESTIGATIONS ON THE OPTIMUM PRE-EXPERIMENTAL PERIOD IN FIELD EXPERIMENTATION ON PERENNIAL CROPS

By T. P. Abraham and G. A. Kulkarni

Institute of Agricultural Research Statistics, New Delhi

It is the general experience that in field experiments on tree crops the experimental error is considerably large compared to annual crop experiments. The data examined at the Institute of Agricultural Research Statistics on some of the crops like mango, arecanut and coconut showed a coefficient of variation as high as 60 to 80 per cent. with single tree plots. Such large variation is partly due to the inherent biological differences among the plant material. In view of the large area occupied by individual trees, it is generally difficult to increase the number of replications from more than 4 or 5. Therefore, other methods such as use of calibrating variates have to be adopted for increasing the accuracy of these experiments. Pre-experimental yields collected for a few years prior to the application of treatments, where possible, serve as a useful calibrating variate. Such observations can be used either for efficient grouping of plots and blocks or for covariance adjustments of the experimental fields. Often it will be advantageous to make use of certain qualitative characters, such as geographical contiguity, disposition of land, etc., as basis for formation of plots and blocks and to use quantitative data on calibrating variates for purpose of covariance adjustments.

In using pre-experimental yield for covariance adjustments it is important to know the minimum period for which the pre-experimental data should be recorded. This, in turn, depends on the pattern of correlation between yields in different years. In the present paper, an investigation has been made on the optimum pre-experimental period with special reference to coconut crop.

Let $y_1, \ldots, y_m$ be the yields of 'm' pre-experimental years immediately preceding the experimental years and $y_{m+1}, \ldots, y_{m+n}$ the yields of 'n' experimental years of a plot. The variance of the mean yield of the experimental period after adjustment for any linear combination

$$\sum_{i=1}^{m} a_i y_i$$

of pre-experimental yields is given by

$$\frac{1}{n^2} V\left(\sum_{j=1}^{n} y_{m+j}\right)(1 - R^2) \tag{1}$$

where R^2 is the correlation of the experimental yield total on the calibrating variate

$$\sum_{i=1}^{m} a_i y_i.$$

The variance of the mean yield for the $m + n$ years is

$$\frac{V(y_1 + \ldots + y_{m+n})}{(m + n)^2}.$$

Therefore the average efficiency of covariance is given by

$$\frac{V\left(\sum_{i=1}^{m+n} y_i\right)}{V\left(\sum_{j=1}^{n} y_{m+j}\right)(1 - R^2)} \cdot \frac{n^2}{(m + n)^2} \tag{2}$$

ignoring the factor

$$\left(1 + \frac{t_{xx}}{E_{xx}}\right)$$

where t_{xx} is the treatment mean square for calibrating variate and E_{xx} is the error sum of squares for the same variate. Since pre-experimental yields are taken before the application of treatments

$$\text{Exp.} \left\{\frac{tr.\ M.S.}{Error\ M.S.}\right\} = 1$$

and so the expectation of

$$1 + \frac{t_{xx}}{E_{xx}},$$

is equal to

$$1 + \frac{1}{v},$$

where v is the error degrees of freedom. Therefore, if the error degree of freedom is large, as will be the case normally, we may ignore this factor and the formula given at (2) will be applicable. It will be observed that the efficiency will depend on the variance of yields in any year and the covariance between y_i and y_j, the values of m, n and

R^2. We may reasonably assume that the experimental error is homogeneous from year to year, *i.e.*,

$$V(y_i) = \sigma^2.$$

We shall now study the changes in efficiency with variations in m, n and R^2 for some of the plausible models of covariance between y_i and y_j.

Let the correlation between y_i and y_j be denoted by ρ_{ij}. We may assume that $\rho_{ij} = f(s)$ where $s = j - i$, $j > i$.

It is reasonable to assume that $f(s)$ is a monotonic decreasing function in 's' with a lower bound $\geqslant -1$. We shall investigate two special cases of $f(s)$ which are likely to be typical of the general situation.

Case I.—$f(s)$ = a constant say $= \rho$.

If the covariate is the total of the pre-experimental yields it can be easily shown that (2) above will be equal to

$$\frac{n}{m+n} \frac{1 + (m+n-1)\rho}{1 + (n-1)\rho} \frac{1}{1 - R^2}$$

where

$$R^2 = \frac{mn\rho^2}{[1 + (m-1)\rho]\,[1 + (n-1)\rho]}\,.$$

This reduces to

$$E = \frac{n}{m+n} \frac{1 + (m-1)\rho}{1 - \rho}. \tag{3}$$

If instead of total pre-experimental yield we take multiple covariance on the 'm' pre-experimental yields, R^2 is given by the usual formula for multiple correlation coefficient in terms of total correlations as

$$1 - \frac{\begin{vmatrix} 1 & a & a & a & \dots & a & a \\ a & 1 & \rho & \rho & \dots & \rho & \rho \\ a & \rho & 1 & \rho & \dots & \rho & \rho \\ a & \rho & \rho & 1 & \dots & \rho & \rho \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ a & \rho & \rho & \rho & \dots & \rho & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho & \rho & \dots & \rho & \rho \\ \rho & 1 & \rho & \dots & \rho & \rho \\ \rho & \rho & 1 & \dots & \rho & \rho \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \rho & \rho & \rho & \dots & \rho & 1 \end{vmatrix}}$$

where

$$a = \frac{\rho\sqrt{n}}{[1 + (n-1)\rho]^{\frac{1}{2}}}.$$

the determinant on the numerator is of order $m + 1$ and that in the denominator of order m. Simplifying we get

$$R^2 = \frac{mn\rho^2}{[1 + (m-1)\rho][1 + (n-1)\rho]}$$

which on substitution in (2) leads to (3). That the multiple correlation of mean experimental yield with the pre-experimental yields taken as separate independent variables is equal to the correlation if mean experimental yield on the total of the pre-experimental yield can be readily seen from the fact that with the assumption cov. $(y_i, y_j) =$ const. the regression equation should be of the form $\lambda y_1 + \dots + \lambda y_m$. From considerations of symmetry, therefore, the regression equation is merely a constant times the total pre-experimental yield.

From (3) it is seen that covariance will be efficient only if $\rho > 1/(n + 1)$. The optimum value of $m = m_0$ got by maximising (3) subject to $m + n = k =$ const. is given by

$$m_0 = \frac{k}{2} + \frac{1}{2} - \frac{1}{2\rho}.$$

With moderate correlation of 0·3 to 0·6 about 4 years is indicated as optimum for a total period of 10 years. However, the change in efficiency of covariance adjustment with more than 2 or 3 pre-experimental years data is seen from (3) to be small.

Case II.—$f(s) = \alpha + \beta\rho^s \quad \rho < 1$

In this case α will be the limiting correlation. Such a correlation function will be obtained if we assume that the yield of a tree in any particular year t will, after removing the common trend, consist of two independent components x_t and z_t with zero expectations and

$$x_t = \lambda x_{t-1} + e_t$$

where e_t is an independent process with zero mean, *i.e.*, x_t is a simple Markov process. x_t will arise because of the genetic peculiarities of the particular tree and its positional effect of a permanent nature as well as physiological factors relating the yields of successive years. z_t arises on account of the sources of error which vary independently from tree to tree and year to year.

$$V\left(\sum_{i=1}^{n} y_{m+i}\right)$$

and R^2 in (1) above can be calculated as given below:

Let y' denote $y - E(y)$, then

$$\text{Cov.}\left(\sum_{i=1}^{n} y'_{m+i}, \sum_{j=1}^{m} y_j'\right) = E\left(\sum_{i=1}^{n} y'_{m+i}\right)\left(\sum_{j=1}^{m} y_j'\right) = E\sum_{i=1}^{m}\sum_{j=m+1}^{m+n} y_i' y_j'$$

$$= \sigma^2 \Sigma\Sigma(\alpha + \beta\rho^{j-i})$$

$$= \sigma^2\left[mn\alpha + \beta\rho\,\frac{(1-\rho^m)(1-\rho^n)}{(1-\rho)^2}\right]$$

$$= \sigma^2 C \text{ (say)}$$

$$V\left(\sum_{i=1}^{m} y_i'\right) = E\left(\sum_{i=1}^{m} y_i'\right)^2$$

$$= \sum_{i=1}^{n} E(y_i'^2) + 2\sum_{i<i'}\sum E(y_i' y_{i'}')$$

$$= \sigma^2\left[m + m(m-1)\alpha + \frac{2(m-1)\beta\rho}{1-\rho} - \frac{2\beta\rho^2(1-\rho^{m-1})}{(1-\rho)^2}\right]$$

$$= 6^2 B \text{ (say)}.$$

Similarly,

$$V\left(\sum_{i=1}^{n} y'_{m+i}\right) = \sigma^2\left[n + n(n-1)\alpha + \frac{2(n-1)\beta\rho}{1-\rho} - \frac{2\beta\rho^2(1-\rho^{n-1})}{(1-\rho)^2}\right]$$

$$= \sigma^2 A \text{ (say.)}$$

Therefore (2) becomes

$$\frac{\sigma^2}{n^2}\left[\frac{BA - C^2}{B}\right] \tag{4}$$

There is no simple expression giving explicitly the value of minimising (4) subject to $m + n = \text{const.} = k$.

The variances can, however, be worked for particular values of m, n and ρ and the relative efficiencies of covariance adjustments can be studied. This has been done in Section 2 for data on coconut crop. If the multiple covariance is carried out, the value of R^2 will be given by

$$1 - \frac{\begin{vmatrix} 1 & b_1 & b_2 & b_3 & \dots & b_m \\ b_1 & 1 & c_1 & c_2 & \dots & c_{m-1} \\ b_2 & c_1 & 1 & c_1 & \dots & c_{m-2} \\ b_3 & c_2 & c_1 & 1 & \dots & c_{m-3} \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & & & \cdot \\ b_{m-1} & c_{m-2} & & & \dots & c_1 \\ b_m & c_{m-1} & & & \dots & 1 \end{vmatrix}}{\begin{vmatrix} 1 & c_1 & \dots & c_{m-1} \\ c_1 & 1 & & c_{m-2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ c_{m-1} & c_{m-2} & \dots & 1 \end{vmatrix}} \tag{5}$$

where

$$b_i = \frac{n\alpha + \beta\rho^i \dfrac{(1-\rho^n)}{1-\rho}}{\sqrt{A}}$$

$$c_j = \alpha + \beta\rho^j.$$

These determinants are not easily evaluated and can be calculated for the particular data.

APPLICATION TO EXPERIMENTATION ON COCONUT CROP

Material for study.—The yield data on 183 trees from 1937 to 1957 were available from Agricultural Research Station, Pilicode, Kerala. The trees were of uniform age and were planted in 1908. In 1942, a randomised block experiment was laid out in experimental area with four cultivation treatments and five replications. These cultivation treatments were found to have no effect on the yield. In subsequent investigations, the correlations were calculated within plots so as to eliminate the treatments and block effects, if any. Since the trees exhibited remarkable biennial swing, the yields of two successive years were combined in all the calculations to eliminate the effect of biennial swing. The data of 1937 were left out to maintain the even number of years. Thus the data of 20 years divided into 10 periods of two years each were available for the study.

The correlation coefficient between consecutive periods, separated by one period, etc., were calculated. The average values of the correlation coefficients obtained are shown in Table I.

TABLE I

Average coefficients of correlation between yields separated by different periods

No. of separating period(s)	Coefficient of correlation	Estimated
1	0·7477	0·7965
2	0·6851	0·6811
3	0·6817	0·6099
4	0·5330	0·5660
5	0·5586	0·5389
6	0·4919	0·5221

The asymptotic regression, $\alpha + \beta\rho^s$, was found to fit closely the data. The estimated values of the correlations from the curve are given in Table I. The curve was fitted by a method given by Patterson (1956). The relative variance of the mean yield for experimental period with the different number of pre-experimental periods is given in Table II,

TABLE II

Relative variances

Pre-experimental period	Experimental periods (n) 1	2	3	4	5	6
0	1·0000	0·8983	0·8387	0·7952	0·7612	0·7337
1	0·3656	0·3526	0·3546	0·3552	0·3537	0·3505
2	0·3925	0·3651	0·3577	0·3518	0·3457	0·3392
3	0·4228	0·3831	0·3674	0·3557	0·3454	0·3339
4	0·4467	0·3974	0·3752	0·3590	0·3453	0·3333

These were calculated for the values of $\alpha = 0 \cdot 4952$, $\beta = 0 \cdot 4884$ and $\rho = 0 \cdot 6169$ given above and the formula given at (4).

The efficiency of covariance can be judged by comparing the variance for any number of experimental periods with the variance for the same total number of years, but having different pre-experimental and experimental periods. For instance, the efficiency of a single period of covariance and three experimental periods is a little over 200 per cent. compared to four experimental periods without covariance.

From Table II, it is clear that covariance results in a substantial gain in information. However, there is little gain by increasing the pre-experimental period to more than one period.

The full advantage of pre-experimental data can be taken only by using multiple covariance on the individual pre-experimental periods as has been already mentioned earlier. It is interesting, therefore, to study the increase in efficiency that can be obtained by using individual periods as calibrating variates compared to a single calibrating variate consisting of the total yield for all the pre-experimental years. Using the formula (5), the multiple correlation coefficients were calculated taking 7 experimental periods and pre-experimental periods ranging from 1 to 4. The correlations obtained are given in Table III.

The increase in the correlation coefficients by taking multiple covariance is very small compared to taking the total of the pre-experimental yields as the covariance. Therefore, in covariance analysis

TABLE III

Correlation between	R	Correlation between	r
(1) y_n and y_m	0·715	y_n and y_m	0·715
2) y_n and $[a_{m-1}y_{m-1}+a_m y_m]$	0·733	y_n and $[y_{m-1} + y_m]$	0·728
(3) y_n and $[a_{m-2}y_{m-2} + a_{m-1}y_{m-1}+a_m y_m]$	0·746	y_n and $[y_{m-2}+y_{m-1}+y_m]$	0·735
(4) y_n and $[a_{m-3}y_{m-3}+a_{m-2}y_{m-2} + a_{m-1}y_{m-1}+a_m y_m]$	0·756	y_n and $[y_{m-3}+y_{m-2}+y_{m-1} + y_m]$	0·739

Note.—The a's are the coefficients in the regression equation $Y_n = \Sigma a_i y_i$.

it is sufficient to use the total for adjusting the experimental yields. This will save considerable time in computations.

Pearce and Brown (1960) working with apple crop and Chapas (1961) working with oil palm also found from empirical data that about two years pre-experimental data collected immediately preceding the experiment are sufficient to obtain maximum efficiency from covariance analysis.

SUMMARY

The optimum number of pre-experimental years required to collect the data before start of the experiment so as to use them for reduction in experimental error by covariance has been investigated for some correlation models for the correlation between any two years yields. Two patterns of correlation between any two years yields were considered. These are (i) constant correlation between any two years yields and (ii) that the correlation between any two years yields decreases as the number of years separating the two years increases, *i.e.*, a model of form $R_{ij} = \alpha + \beta\rho^{j-i}$ where R_{ij} is the correlation coefficient between the i-th and j-th years' data. Examination of the yield data of coconut trees for 20 years showed that the correlations fit very closely the second of the above models. It was found that about two experimental periods data immediately prior to the experimental period are sufficient for covariance analysis. This procedure resulted in more than 100 per cent. gain in information.

ACKNOWLEDGEMENTS

The authors feel grateful to Dr. G. R. Seth, Statistical Adviser, Indian Council of Agricultural Research, for his keen interest and encouragement.

REFERENCES

1. Shrikhande, V. J. .. *J. Ind. Soc. Agri. Stat.*, 1957, **9**, 82–99.
2. Patterson, H. D. .. *Biometrics*, 1956, **12**, 323–29.
3. Pearce, S. C. and Brown, A. H. F. *J. Hort.*, 1960, **35**, 56–65.
4. Chapas, L. C. .. *Emp. J. Exp. Agri.*, 1961, **29**, 212–24.