SOME SUGGESTIONS FOR MINIMIZING ERRORS FROM MISHAPS IN FIELD EXPERIMENTS WITH PERENNIALS

By V. ABEYWARDENA

Biometrician

Coconut Research Institute, Ceylon.

Field experiments with perennials are very often hampered by several trees dying through natural causes while the experiment is in progress. As a result of such mishaps, experimental plots may be rendered non-comparable within the limits of the experimental error. This arises not only due to the differences in plant density resulting from these mishaps but also due to possible interactions between plant density and plant growth, which latter is the factor most commonly missed in the statistical evaluation of results.

From a review of available literature pertaining to experiments with perennials and also from the author's recent association with some experiments and the statistical analyses of their results at certain crop research institutes of Ceylon, it appears that the method popularly adopted in order to correct for these missing plants in the plots, is to calculate a mean value per surviving plant in each plot and use same as the plot variate for further statistical analyses.

This approach which we have been used to adopt almost instinctively, carries with it the implicit assumption that any gaps occurring in the plots as a result of these mishaps or the replacement of such gaps with relatively younger plants, are of no consequence to those plants that survive. Such an assumption, if accepted without verification, cannot be considered reasonable in view of the fact that the phenomenon of competition in plant communities — especially perennials — is not in the least uncommon, Hutchinson and Panse (1935) refers to the presence of competition in cotton and wheat; Christides (1939) in cotton; Amos and Hoblyn (1928) in Hops; Taylor (1951) in strawberries; Sharpe and Blackmon (1951) in peccan trees. Pearce (1953) virtually admits the loss of competition due to gaps in the case of most tree crops, when he deals in detail on some methods of statistical control of such situations. The author (1961 and 1962) has observed that this feature of competition is present in young coconut plantations too. He has also found it to be highly suggestive from his limited experience of experimental data on rubber trees, wherein, incidentally, mishaps are a very common feature.

Even if this assumption of the negligibility of the competitive feature were found reasonable through verification — in fact, it can be so under certain conditions — there is yet a further bias always arising from the use of a mean plot value purely as a result of the experimental error. This latter, of course, does not lead to any serious errors unless the experimental error is appreciably large.

Therefore there is sufficient ground to fear that the use of a mean value per surviving plant as the plot variate without sufficient verification as to its validity, is very seldom, if ever justified under plantation conditions.

However the fact remains that the mean value approach is still popular among some Research workers. This is quite understandable because it so happens that more often Research workers who are aware of the biological interactions in a plant community are not necessarily conversant with the statistical approaches and on the other hand, statisticians who are aware of the statistical approaches are not necessarily conscious of the biological interactions. For the benefit of those who still adopt this approach, it is intended, in this paper to discuss some hypothetical situations wherein (1) the effect, on neighbouring plants, of gaps or of younger plants which have filled such gaps, and (2) the experimental error itself, can invalidate the use of a mean value per surviving plant as the plot variate. The "covariance analysis" approach in some form or other, is shown to be the correct one to be adopted in all such situations.

Some hypothetical situations which invalidate the use of the mean value per surviving plant

Let us consider a series of plots constituting an experiment. Prior to the application of treatments, these are considered identical in respect of area, plant number, and soil fertility etc., subject of course to the experimental error, the mathematical distribution of which is known to follow the Guassian Curve of Errors, provided the causes of variation are many and no one of them alone accounts for more than about 30 per cent of the variation (Lush 1954).

When some plants die (through natural causes) in certain plots, the situation changes and another source of variation is introduced — namely plant density. It is this source of variation that we are accustomed to eliminate by using the mean girth per surviving plant (if girth happens to be the character measured) as the variate for further statistical analysis. The implicit assumption behind this approach, as insinuated earlier, is that the total tree girth in a plot is an exact linear function of the number of plants in a plot. That is, for example, the total tree girth in a particular plot with four surviving plants will be exactly double what it would be if it had only two surviving plants; and the total girth in a plot with twelve plants surviving, will be four times what it would be if it had only three plants, so on and so forth. In short the assumption is that, barring the treatment effects, the mean girth per plot is a constant for the plots whatever their plant density — subject, of course, only to the random experimental error of known mathematical distribution.

Let us examine to what extent we are valid in this assumption and what its implications are on our interpretations.

Errors involved in this assumption, as mentioned earlier, can arise from either of two sources or both — namely (*a*) errors arising from advantages or disadvantages accruing to the survivors due to gaps in the plots as a result of lost plants or their replacements by relatively younger plants and (*b*) errors due to experimental error. These are considered separately below.

Errors arising from gaps and replacements

Depending on the circumstances, errors arising from this source can take one of three different patterns, which we shall discuss below under situations I, II and III.

Situation I:—

For convenience, let us assume that the experimental error is nil and any two plots are originally identical in all respects. Under these circumstances, if a certain number of plants die in the two plots leaving two plants in one and six in the other, one can reasonably expect that the two-plant plot, considered on an individual plant basis, will show better growth than the six-plant plot because in the case of the former, the loss of competition arising as a result of lost plants is relatively more. More often than not one can expect the plot with more survivors, to show a lower mean growth level than that with less survivors. On this basis, the curve of the total girth (y) of a plot against the number of surviving plants (x) in the plot will reasonably be of an "increasing — decreasing" (or diminishing returns) type (Fig. 1).

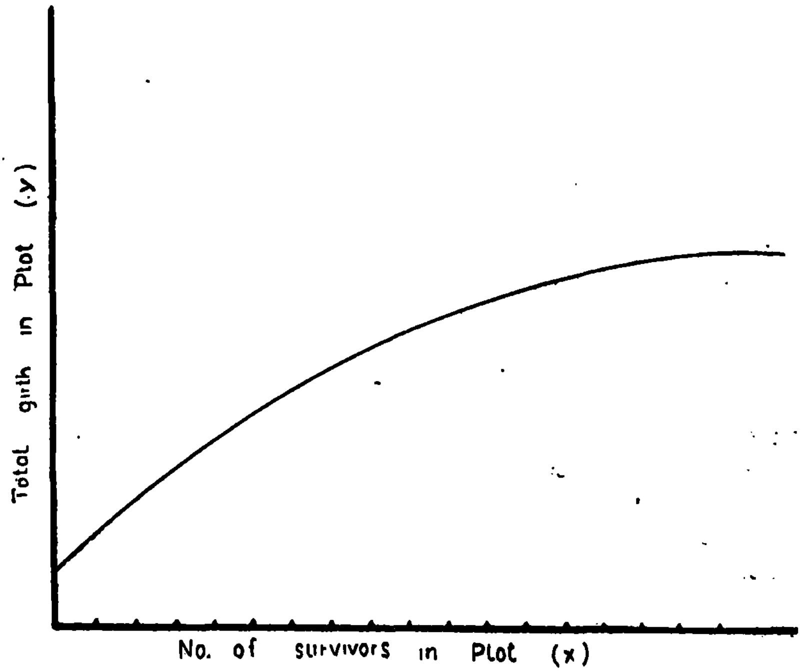

Fig. 1. The relationship between total girth and number of surviving plants (Diminishing returns)

The total girth in a plot invariably increases as the number of trees increases; but the increase with each additional tree becomes less and less.

The situation is similar when such gaps are replaced by relatively younger plants, except that the advantage accruing to the plots with fewer survivors is less — the loss of competition being less.

The erroneous trends indicated by the use of a mean value per surviving plant in the above situation is explained by the numerical example given below.

Example:—

TABLE I. Mean value versus value obtained through covariance technique

Plot No.	Col. (1) No. of survivo.	Col. (2) Total Girths	Col. (3) Expe. treat. effc./plot	Col. (4) Mean Girth	Col. (5) Patt. on B. of M. Vq.	Col. (6) Patt. on B. of regr.
1	6	78.49	0.0	13.1	13.1	10.4
2	7	84.50	0.4	12.1	12.5	10.8
3	8	89.89	0.1	11.2	11.3	10.5
4	9	94.70	0.8	10.5	11.3	11.2
5	10	98.89	0.5	9.9	10.4	10.9
6	11	102.50	0.3	9.3	9.6	10.7
7	12	105.50	0.9	8.8	9.7	11.3

Suppose in an experiment there were seven plots allotted to seven treatments, with the number of surviving plants varying from (say) six to twelve; and the total tree girth of the surviving plants in each plot are as given in column (2) Table 1, indicating a diminishing returns relationship between total girth and plant density as suggested earlier with zero as the error. Suppose it were known that each of the seven treatments could bring about an increase in girth per plant of such absolute amounts as given in column (3).

Then if we adopt the mean approach to compare the treatments, we can expect the final pattern of the treated plots to be as given in column (5) where the value for a particular treatment is merely the mean girth plus the corresponding treatment effect.

On the other hand, we have an alternative approach wherein we can determine the curvilinear relationship between total girth and plant stand independent of the treatments. In this particular (errorless) case the relationship is given by

$$Y = 29.92 + 9.894\,X - 0.2997\,X^2$$

where Y is the total girth and X the plant stand.

If we adjust the total girth for the plant stand on the basis of this curvilinear regression, we get the treatment pattern as given in column (6).

It is clear that the latter approach gives the true picture of the treatment effects. The difference between any two values in column (6) gives the exact difference between the corresponding treatment effects given in column (3). Whereas if we based our judgement on mean girths given in column (5), our interpretation would have been wrong.

It is also observed that if the relationship is of the diminishing returns type indicated here, the mean girth approach overestimates the plots with fewer surviving plants and underestimates those with more surviving plants.

Situation II:—

A completely different situation may arise due to this same phenomenon. A situation may arise that some plots are originally very poor or contain some harmful factor and some are rich, in which case relatively more plants may die in the poorer plots, leaving only a couple of plants. The pattern of diminishing returns may not be noticed here. In the poorer plots, not only more plants die but the survivors will also be poor and *vice versa* for the richer plots. Therefore the relationship of total girth to plant stand will be of the increasing returns type (Fig. 2).

A similar trend can also be expected under a quite different set of circumstances. Conditions may have been such that shade from the plant canopy might have been conducive to better growth. In such a set up, if some plants die in the experimental plots (say) as a result of a gale, then, one can expect a plot with more survivors to perform better on an individual plant basis than one with less survivors.

The curve of the total girth (y) against the number of surviving plants (x), under such circumstances, may show increasing returns (Fig. 2).

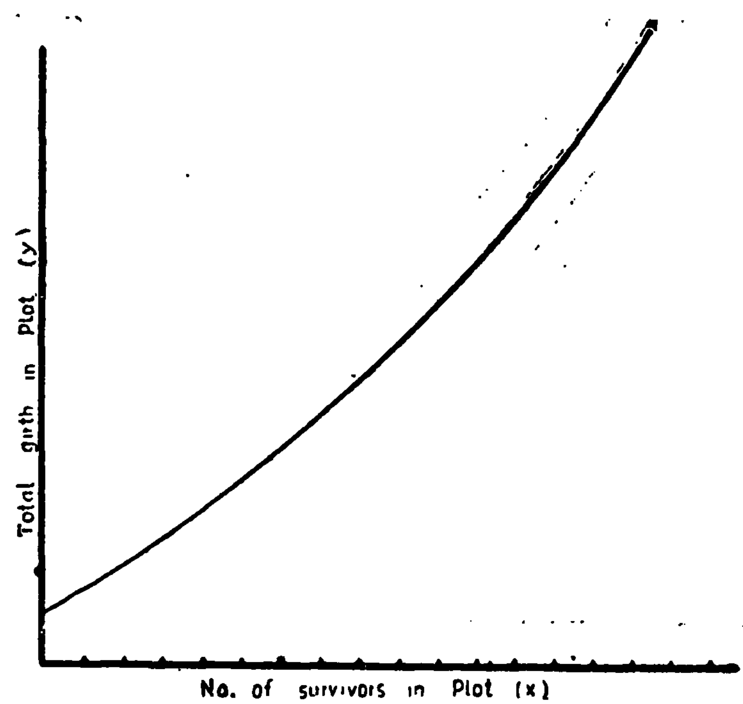

Fig. 2. The relationship between total girth and number of surviving plants (Increasing returns)

Here too the correct treatment pattern is obtained only through the curvilinear regression approach. It will be noted that in this case, the mean girth approach underestimates the plots with fewer survivors and overestimates those with more.

Situation III:—

A more complex situation may also arise — especially when there is extreme variability in the experimental block. Here one may notice a combination of both situations I and II, giving rise to an S-shaped curve for total girth and plant stand (Fig. 3), wherein the poorer plots will show increasing returns and richer plots will show diminishing returns.

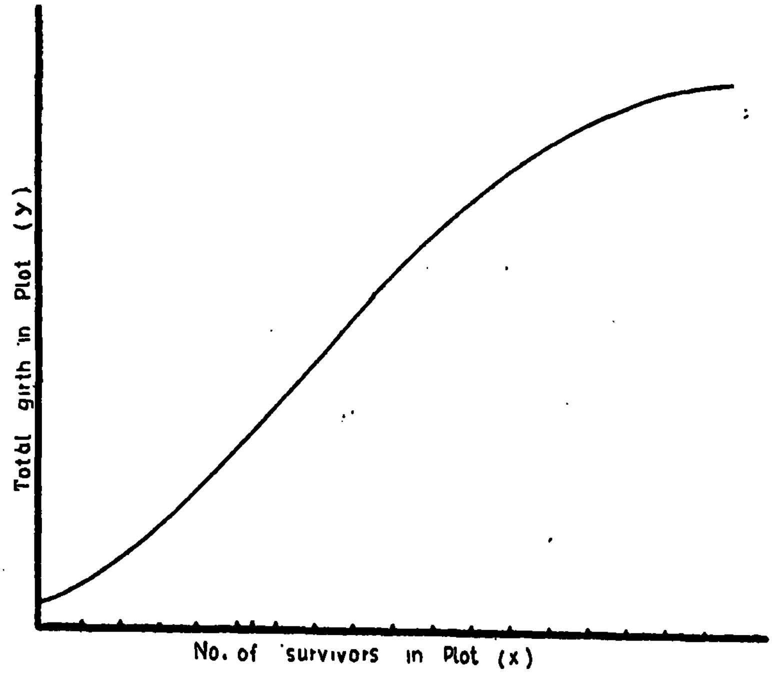

Fig. 3.—The relationship between total girth and number of surviving plants (S-shaped)

Corrections for such situations may be through a triple covariance analysis using a third degree polynomial model such as:

$$y = a + bx + cx^2 + dx^3$$

Errors due to experimental errors

Suppose the presence of any number of gaps or their replacement by relatively younger plants is of no consequence on the remaining plants. Then the relationship between total plant girth and number of surviving plants may be considered linear. Yet due to the experimental error, one cannot expect that the total plant girth per plot is an exact linear function of the number of plants. The relationship is only a statistically linear relationship, in the sense that the total girth may fluctuate about a straight line which gives the real underlying relationship between total girth and plant stand. Therefore any value for total girth that occurs above the regression line will be overestimated and those points below the line will be underestimated, if the mean value approach is adopted.

The best correction will again be through a single covariance analysis using the plant stand as the independent variable.

It will be pertinent to mention herein that errors arising from this factor — namely the experimental error — is common to all the situations mentioned earlier.

Summary and Conclusions

The use of a mean value per surviving plant to correct for differences in plant density arising from mishaps in experimental trees, can give rise to serious errors in the interpretation of data. No doubt in statistical theory, one accepts that the arithmetic mean is an unbiassed, consistent and sufficient statistic and most of us are therefore, led to using this index without the least suspicion, that certain biological interactions common in plant communities can invalidate its use. The mean really commands such respect only when it applies to a random variable and not to some characters in respect of individuals such as trees in a plantation which within a restricted area such as an experimental plot are to some extent inter-dependent. In most cases the presence of gaps or of replaced younger plants will place the neighbouring plants in an advantageous position due to loss of competition and sometimes in a disadvantageous position as explained earlier. Therefore the mean value per plot in respect of the surviving plants ceases to be comparable within the limits of a random experimental error.

The use of the covariance technique through suitable calibrating variates is the only correct approach under the circumstances. One may use the number of survivors as one of the calibrating variates and a suitable transformation of same as a second calibrating variate (Abeywardena 1962) to account for the curvilinearity due to competition, and then a double covariance analysis may be carried out (Pearce 1953). One may also use a very interesting and unique approach suggested by Pearce (1954)*. Or one may try out a suitable amalgamation of both these methods. It may, however, happen that such analyses do not show any significant effect, because the loss of plants may have been very recent and therefore there would have been hardly any time for the effect of the loss of competition to assert itself. The use of calibrating variates such as duration of gaps and or age of replaced young plants, can be helpful in such situations and a multiple covariance analysis may be necessary.

Generally, the experimenter is the best judge of the possible factors operating under a particular set of circumstances and in respect of the plant species he is handling. The best attitude would be to try out various calibrating variates without making any assumptions whatsoever. When one considers the long period and the high cost involved in experimentation with perennials, the extra computing effort involved is more than justified. At any rate, from the point of view of the correct interpretation of data, the suggested statistical approaches, may be extremely helpful.

Even if the factors operating in the particular experimental set up, are restricted to only the linear effect of the stand of plants, a single covariance analysis is recommended in place of the mean value approach which can lead to appreciable error.

*As this reference may not be readily available, an extract of the relevant section is given in an appendix to this paper, A note of warning is necessary in the use of calibrating variates. One must ensure that the calibrating variates are not associated with the treatments, because if it were so, removal of the effect of calibrating variates through covariance analysis, also removes part of the treatment effect. Therefore one must satisfy oneself that the deaths of trees are not caused by the treatments themselves. If the deaths are due to the treatments, one should avoid any corrections but proceed to analyse the plot totals in respect of the survivors.

Acknowledgements

I am thankful to Mr. A.J. Jeevaratnam, Soil Chemist, Rubber Research Institute of Ceylon, the analysis of whose data, has directed my attention to this problem.

References

1. Abeywardena, V. (1961). Some noteworthy features of the "initial flowering" pattern in a coconut plantation. Ceylon Coconut Quarterly, Vol. XII, Nos. 3 and 4.

2. Abeywardena, V. (1962). Statistical control of variability in coconut experiments (with Editor). Emp. J. Expt. Agric.

3. Amos, J. and Hoblyn, T.N. (1928). Manurial trials with Hops. A.R. East Malling Res. Stat. for 1926–27, II Suppl., pp. 165-71.

4. Christidis, B.G. (1939). Further studies on competition in yield trials with Cotton. Emp. J. Expt. Agric. 7, 111–120.

5. Hutchinson, J.B. and Panse, V.G. (1935). A study of Margin Effect in variety trials with Cotton and Wheet. Ind. J. Agric. Sci. 5, 671–92.

6. Lush, J.L. (1954). Statistical methods in animal husbandry. F.A.O., Rome.

7. Pearce, S.C. (1953). Field experimentation with fruit trees and other perennial plants. Tech. Commun. Commonwealth Bur. Hort. Plantation crops.

8. Pearce, S.C. (1954). Resume of Lectures on "Experimentation with perennial plants" conducted at the International Training Centre on Experimental Designs and Survey Techniques of Agricultural Experiments, New Delhi, India.

9. Sharpe, R.H. and Blackman, G.H. (1951). A study of plot size and experimental design with Peccan Yield Data. Proc. Amer. Soc. hort. Sci. 56: 202–6.

10. Taylor, J. (1951). Statistical studies on strawberry crop and vigour measurements. A.R. East Malling Res. Stat. for 1950, pp. 100–7.

Appendix

Incomplete plots:

If plots are incomplete on account of the death of trees, provided that a fairly high proportion of each plot remains, say, 80 per cent or more, it suffices to analyse y = crop per surviving tree adjusted by x, a measure of the loss of competition occasioned by the deaths.

Thus if x represents an experimental tree and o a guard, suppose a plot is like this:

o	o	o		o	o	o	o
o	x^1	x	x^1	x	x	x	o
o		x^1	x	x^1	x	x	o
o	o	o	o		o	o	

Then eleven experimental trees remain and $y = \frac{\text{crop}}{11}$. Four experimental trees have lost one neighbour each (diagonals do not count), so the loss of competition is $\frac{4}{11} = x$.

Again, consider this plot:

o	o	o	o	o	o	o	o
o	x^1		x^3		x^1	x	o
o	x	x^2		x^2	x	x	o
o	o	o		o	o	o	o

Nine trees remain, so $y = \frac{\text{crop}}{9}$.

As to loss of competition, two trees have lost one competitor, two have lost two and one has lost three, so

$$x = \frac{(2\times 1) + (2\times 2) + (1\times 3)}{9} = \frac{9}{9} = 1$$

All that is needed now is to adjust an analysis of variance of y by the values of x using covariance.