

IMPRECISE EXPERIMENTAL DESIGNS AND FEEBLE CONCLUSIONS

R.O. THATTIL

Department of Crop Science, Faculty of Agriculture,
University of Peradeniya, Peradeniya, Sri Lanka

ABSTRACT

This paper discusses how the precision of an experiment can be improved using good design techniques. Ideas on blocking, factorial experiments, use of covariates and importance of systematic spacing designs are given. The analysis and interpretation of results are discussed with examples and emphasis given to the use of contrasts which match the objectives of a given experiment. The confidence interval approach in testing of hypotheses also discussed.

KEY WORDS: Experimental Designs, Analysis, Interpretation

INTRODUCTION

The Tropical Agriculturist is the oldest scientific journal pertaining to Agricultural Experimentation in Sri Lanka. It is therefore, imperative that the papers published in the journal should be of top quality. This article is therefore written with the aim of giving agricultural experimenters, ideas on designing precise experiments and using 'proper' data analysis so that conclusions are valid.

To reach general conclusions the experiment should be repeated over locations and seasons. At a given site all necessary design techniques should be utilized to control variability. If a systematic variation such as soil variability is not controlled by blocking, that variability too enters into residual variability leading to non significance of the treatment effect. The experimenter should be also aware that findings from trials done at research stations cannot be directly imposed at the farm level. Yet, variability at farm level can also be controlled using covariates. At the research station, the experimenter has more control over the experiment and therefore, no lame excuses can be given for using poor designs. The following sections examine design aspects, techniques of data analysis and interpretation.

EXPERIMENTAL DESIGNS AND CONCLUSIONS

PRECISE EXPERIMENTAL DESIGNS

Blocking

Blocking is a technique that is used to control systematic variability. If soil variability occurs systematically, blocking is done perpendicular to the direction of variability so that the resulting blocks are more or less homogenous. The variability due to blocks is therefore considered explicitly in the model. In such a situation if blocking was not done the residual variability will be inflated as the block variability too is added on to it.

In cases where more than one direction of variability exist, latin square designs can be employed. If the number of treatments is low, replicated latin squares should be used to obtain the required number of degrees of freedom for the residual. If a large number of treatments are used incomplete latin square designs are useful. If all the treatment cannot be accommodated in a block then there is a need to use incomplete block designs. The use of the principle of latin squares in designing experiments is given in Thattil (1992).

Precise estimation of the random error component is essential, for tests of significance to be valid. The estimate is obtained as a residual (total variability – explained variability). Therefore, if variability due to a systematic effect is not considered explicitly in the model it will be added on to residual variability.

Consider the following partition of variability

Total variability = Treatment variability + variability due to soil fertility
+ variability due to moisture + variability due to other factors

Since residual variability is calculated as the difference between total variability and the rest, the residual variability will be an overestimate if any of the sources of variability is not controlled by the design.

Depending on the variability of the experimental site, a block can accommodate about 8 – 12 treatments. However, if the site is very homogenous

(e.g. Green house) blocking may not even be necessary. When the number of treatments become large, incomplete block designs should be used since the block may not be able to accommodate all the treatments. Among incomplete block designs the most useful are the balanced incomplete block (BIB) designs. In BIB designs every treatment pair is compared with the same precision. It may sometimes not be possible to obtain the required number of blocks necessary for BIB designs. In such cases partially balanced incomplete block (PBIB) designs can be used. Many textbooks are available that discusses design and analysis of most type of designs. A few useful books are given in the reference section.

It should be also noted that conclusions reached from greenhouse and pot experiments may not be valid under field conditions. However, such experiments can be used as a preliminary tool.

Factorial Experiments

Single factor experiments may in some cases give useful results. However, to obtain information on interactions between several factors experiments using a factorial structure is an absolute necessity. Results from single factor trials do not reflect what will happen under field conditions.

In many factorial experiments it is a fashion to employ split-plot type designs. Unless there is practical necessity to have different plot sizes for the different factors, such designs should be discouraged as it has many disadvantages. Some text books state that there is higher precision on estimating interaction effects which is not true. Where complete block designs can be employed the precision for the interaction effects are the same while the main effects will also have high precision.

In large factorial experiments all treatment combinations cannot be accommodated in a complete block. For such cases confounding of higher order interactions can be used to accommodate the treatment combinations in smaller blocks. Obviously a full replicate will then consist of more than one block. If the

EXPERIMENTAL DESIGNS AND CONCLUSIONS

experiments investigates many factors even a single replicate will not be possible. In such cases fractional factorial designs can be employed to obtain useful information on the non – confounded effects. Texts dealing with confounded designs and fractional factorial designs are given in the reference list.

Many experimenters shy away from factorial designs as they are not comfortable with the analysis and interpretation of results. The next section gives guidelines on interpretation of results. The main idea in designing experiments is to ensure that all treatments or treatment combinations (in factorial experiments) are compared under similar conditions. It is to ensure similar conditions that techniques such as blocking are used. In large factorial experiments information on higher order interactions are sacrificed (by confounding) to obtain useful information on main effects and lower order interactions. When interaction is present, main effects should be examined at fixed levels of the other factor(s).

Experiments investigating optional spacing of crops cannot be performed with conventional designs. Systematic designs should by necessity be employed for spacing experiments. For intercropping experiments where optional spacing for two crops are investigated the designs become more complicated. Useful references are provided by Nelder (1962) and Thattil (1988) for spacing experiments. Mead and Riley (1990), reviews statistical ideas relevant to intercropping research, while Thattil (1985) provides useful ideas on the design and analysis of intercropping experiments. The results from systematic spacing designs can be used to design conventional designs consisting of only a few levels of spacing.

ANALYSIS AND INTERPRETATION OF DATA

Statistical Analysis

Many experimenters begin and end the interpretation of results by the single statement “The treatment effect is significant”

Consider the following hypothesis test made in the 2 – sample case (pooled t test).

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{Verses} \quad H_1 : \mu_1 - \mu_2 \neq 0$$

If the Null hypothesis (H_0) is rejected the conclusion is that the mean difference is significantly different from zero; no other information is given. It is more informative to give an estimate of the mean difference and a confidence interval (C.I.) which tells the range of the possible difference at a given level of confidence (say 95% confidence interval). The C.I. not only gives a range for the expected difference, but also contains the significance test. If the interval does not contain the hypothesized (H_0) value, it would have been rejected. Many experimenters stop with the statement "The difference is significant, highly significant etc."! Highly significant from zero!

With more than 2 treatments, the Analysis of Variance (ANOVA) procedure is used and the conclusion made is that the treatment effect is significant. Many experimenters follow this up with the least significant difference (LSD) procedure and state which pairs are significantly different. It is only an extension of the 2 sample case where one states that the difference is significantly different from zero. However, when the treatment number is large, there is every chance that one or more pairs will be falsely declared as significant. The probability of such an event occurring is $1 - (1 - \alpha)^m$ where m is the number of pairwise comparisons made and α is the significance level. If $\alpha = .05$ and $m = 45$ (for 10 treatments) this probability works out to 0.9, an almost certain chance to make at least one false declaration of significance! To keep the comparison wise error rate constant, no matter how many comparisons are made, the Duncan's multiple range test is used giving an open license to make all possible simple comparisons most of which are not even properly defined as objectives of the experiment! If t treatments are investigated the degrees of freedom (d.f.) for the treatment source is given by $t - 1$. As such one should make only $t - 1$ comparisons.

Given the objectives of an experiment, these can be easily translated as contrasts involving 2 or more treatments. Contrasts can be tested for significance and confidence intervals given for each of the contrasts.

No more than $t - 1$ contrasts (comparisons) should be made. If the contrasts

EXPERIMENTAL DESIGNS AND CONCLUSIONS

from an orthogonal set it will be an added bonus. The number of contrasts in an orthogonal set will be $t - 1$.

Example :

An experimenter evaluates 3 local varieties (V_1, V_2, V_3) and 2 imported varieties (V_4, V_5) of a certain crop. His objectives are to compare,

1. Local Vs imported varieties
2. Comparison within local and imported varieties

The set of contrasts (comparisons) will then be as given below.

Contrasts	Varieties				
	V_1	V_2	V_3	V_4	V_5
1	2	2	2	-3	-3
2	2	-1	-1	0	0
3	0	1	-1	0	0
4	0	0	0	1	-1

Contrast 1 examines the first objective while contrasts 2, 3 and 4 are within group comparisons. The number of comparisons is equal to $t - 1 = 4$. One need not force the comparisons to be an orthogonal set, if it does not match the objectives. Also, all the comparisons in an orthogonal set need not also be done if it does not match the objectives. Finally C.I.'s can be given for each contrast. To keep the experimental error rate constant, simultaneous confidence intervals using Bonferoni's method can be used.

Use of Covariates

When the variability is not systematic, blocking will not help to control variability. In such cases use of covariates is essential in controlling variability. Consider a rubber plantation; the yield of latex depends on the girth of the tree. The girth classes will not be found in homogenous patches. Blocking will not therefore be possible. However, the girth of the tree (supplementary variable measured in addition to the measurement on the response variable) can be used as a covariate to adjust the yield. In tree crop experiments, past yields of the crop

will be an ideal covariate, if such records exist.

In on farm trials, larger plots are used and blocking will not be possible where the number of treatments are larger than 4 or 5. In such cases too, use of covariates will help to control variability. When comparisons are made treatment means adjusted to the covariate has to be employed.

Analysis of Quantitative factors

When the treatments are quantitative more powerful techniques such as regression analysis can be utilized. One can then examine the nature and strength of the relationship between the independent variable(s) and the dependant variable. If necessary non-linear regression analysis can be carried out depending on the nature of response. Draper and Smith (1981) is an excellent text book for regression analysis.

Experimenters usually take an unusually large number of levels of the factor. This is not optimal. The number of levels used should not exceed the number of parameters of the fitted curve by more than one. The extra level is to examine the lack of fit of the model. At each level there should be multiple observations, if not lack of fit cannot be tested. The spacing between levels will also depend on the type of model used. Some experimenters use quantitative factors but the analysis is carried out as for qualitative factors!

Statistical software such as SAS (statistical analysis system) and SPSS (Statistical package for social sciences) contain orthogonal as well as non – orthogonal statistical analysis. However, it is the experimenter who will be responsible for interpretation of the output. The above software can solve univariate as well as multivariate problems.

In some instances measurements are made on multiple response variables. In such cases multivariate analysis should be done followed by univariate analysis if necessary. In cases where the scale of measurement is ordinal, non – parametric techniques should be employed while if the data is of the categorical type (count data), discrete multivariate techniques (see Bishop et al., 1975) should be used. There are many text books available that describe these methods.

EXPERIMENTAL DESIGNS AND CONCLUSIONS

OTHER CONSIDERATIONS

The researcher should keep in mind that experiments carried out in research stations should address the field conditions of the farmer. There is no point in doing pot experiments or experiments on a single site for a single season and trying to extrapolate to the general situation in the country.

The coefficient of variability (CV) of the experiment reflects the efficiency of the design used. In field crop experiments the CV should not exceed 15 – 20%, while for tree crops even a CV of 30 – 35% is acceptable. For experiments done in green houses where conditions are fairly homogenous even 10% is a high CV.

$$\text{CV of experiment} = \frac{\sqrt{\text{Residual MS}}}{\text{mean of observations}} \times 100$$

In incomplete block designs where the block size is smaller than the corresponding complete block design a lower CV can be expected. The design should also provide adequate d.f. for the error term. It is of utmost importance to remember that fairness in comparison of treatments is essential. Thus, the treatments should be applied to uniform material. If complete blocks cannot be used, treatment means should be adjusted with respect to the blocks. If uniform blocks are not available adjustment should be made using covariates. Even for unbalanced designs it is prudent to use similar block sizes, number of replicates as far as possible constant and number of times any pair of treatments occur together in the block as same as possible.

REFERENCES

Draper, N. and H. Smith. 1981. Applied Regression Analysis. 2nd Ed. John Wiley and Sons. Inc. New York.

Bishop, Y.M.M. et al 1975. Discrete Multivariate Analysis. The MIT Press, Cambridge.

Finney, D.J. 1960 An introduction to the theory of experimental design. University of Chicago press. Chicago and London.

- Hinkelmann, K. and O. Kempthorne. 2000. Design and Analysis of experiments, Advanced Experimental Design. John Wiley and Sons, New York.
- Mead, R. and J. Riley 1981. A review of Statistical ideas relevant to intercropping research. Journal of the Royal Statistical Association. 144: 462 – 509.
- Nelder J.A. 1962. New kinds of systematic designs for spacing experiments. Biometrics. 18:283-307.
- Thattil, R.O. 1985. Design and analysis of intercropping experiments. Unpublished Ph.D thesis. Virginia Polytechnic Institute and State University, USA.
- Thattil, R.O. and W.A.J.M. de Costa. 1988. Spacing experiment on Maize x Mungbean interbreeding system using a “3 – way” systematic design.
- Thattil R.O. 1992. Experimental designs for intercropping under trees. Multipurpose tree species in Sri Lanka : Research and Development. 37 – 42.