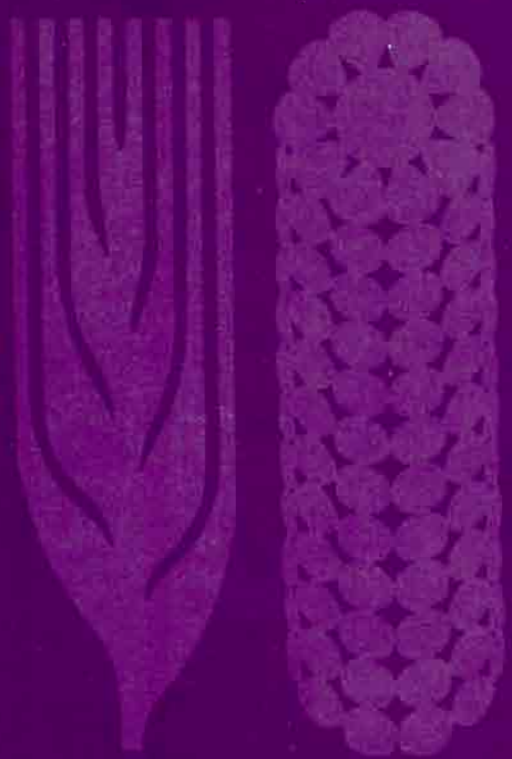
DEVELOPMENT OF DESIGNS TO IMPROVE EFFICIENCY OF OFR

Training Working Document No. 3



CIMMYT

TRAINING WORKING DOCUMENT

DEVELOPMENT OF DESIGNS TO IMPROVE EFFICIENCY OF OFR

Training Working Document No. 3

Prepared by
Roger Mead
Consultant
in collaboration
with CIMMYT staff

CIMMYT
Lisboa 27
Apdo. Postal 6-641,
06600 México, D.F., Mexico

PREFACE

This is one of a new series of publications from CIMMYT entitled *Training Working Documents*. The purpose of these publications is to distribute, in a timely fashion, training-related materials developed by CIMMYT staff and colleagues. Some Training Working Documents will present new ideas that have not yet had the benefit of extensive testing in the field while others will present information in a form that the authors have tested and found useful for teaching. Training Working Documents are intended for distribution to participants in courses sponsored by CIMMYT and to other interested scientists, trainers, and students. Users of these documents are encourage to provide feedback as to their usefulness and suggestions on how they might be improved. These documents may then be revised based on suggestions from readers and users and published in a more formal fashion.

CIMMYT is pleased to begin this new series of publications with a set of six documents developed by Professor Roger Mead of the Applied Statistics Department, University of Reading, United Kingdom, in cooperation with CIMMYT staff. The first five documents address various aspects of the use of statistics for on-farm research design and analysis, and the sixth addresses statistical analysis of intercropping experiments. The documents provide on-farm research practitioners with innovative information not yet available elsewhere. Thanks goes out to the following CIMMYT staff for providing valuable input into the development of this series: Mark Bell, Derek Byerlee, Jose Crossa, Gregory Edmeades, Carlos Gonzalez, Renee Lafitte, Robert Tripp, Jonathan Woolley.

Any comments on the content of the documents or suggestions as to how they might be improved should be sent to the following address:

CIMMYT Maize Training Coordinator
Apdo. Postal 6-641
06600 Mexico D.F., Mexico.

DESIGN OF ON-FARM EXPERIMENTS

1. Experiments at Different Stages of OF Research

Like any other form of research project an OF research project passes through different stages from the exploratory through the detailed examination of components to the verification. These stages are rarely completely separable. For example, at verification there will sometimes be additional treatments of possible promise.

It is, of course, crucial that the objectives at each stage of the experimentation are clearly identified and relevant to that stage. The choice of treatments is discussed in general in section 3. Here it is worth noting that there will be a general tendency for the number of experimental treatments to decline through the programme. At the earliest stages there should be many treatment factors, each with few levels. In subsequent development the number of factors in a single experiment will tend to decline with the number of levels tending to increase. At no stage should there be many factors each with many levels since such an experiment would appear to be asking whether interactions were important and simultaneously assuming that they were while trying to identify best levels of each factor. It is not just the impossibility of managing such a trial but the contradiction in objectives that makes it inappropriate.

Many of the concepts discussed later in this paper are relevant to most stages but there will be differences at least of emphasis. There will be differences in the number and distribution of farms with larger numbers, certainly, for the verification stage. I assume that the questions of choice of sample farms and the definition of recommendation domains are not covered in this paper, being discussed in CIMMYT Training Working Documents Nos. 4 and 5.

2. General Statistical Principles of Precision, Replication and Resource

2.1 Precision of Results

It is extremely important to assess, before the experiment, the precision of the information to be obtained from the experimental results. This involves thinking about three quantities. First the likely background variability, as measured by the Coefficient of Variation, CV, or the plot Standard Deviation, s; second, the difference in yield (or other performance variable) which is important, d, or ^ if the difference is expressed as a percentage of the experimental mean yield; the third component is the number of replications n for each treatment.

Assume that we are interested in the comparison between two treatments each having a total of n observations across the whole experiment. The crucial statistical result is that the standard error of the difference between two mean is:

$$SE\ (X_1 - X_2) = \sqrt{(2s^2/n)} = S\sqrt{(2/n)}$$

To have a realistic chance of identifying whether the true difference between the treatments (μ_1-μ_2) is as large as our critical difference d we must make the SE a good deal smaller then d. How much smaller depends on the significance level we propose to use and the risk we are prepared to accept of missing a true difference as big as d. A useful rule of thumb is to try to make the SE no bigger than d/3. This means that d is 3 standard errors which allows 2 standard errors for achieving a 5% significance level and an extra standard error for bad luck in getting (X_1 - X_2) smaller than d (the risk of missing a true difference of d is one-sixth).

Consider an example where we expect a mean experimental yield of 3000 kg/ha, a CV of 20% which implies s = 600 kg/ha and where we would hope to detect a true difference, if it exists, of 750 kg/ha. The SE should therefore be no bigger than 250 kg/ha and we thus require that n shall be at least big enough to make

$$250 = 600 \sqrt{(2/n)}$$

This implies that

$$n/2 = (600/250)^2 = 5.76$$

and n should be at least 12.

We can approach the problem from the other end, and consider what precision would be achieved and what differences detected for various possible n.

If we can afford 8 observations per treatment our standard error will be

$$\text{SE (for n=8)} = 600 \sqrt{(2/7)} = 300.$$

This would give us a 5/6 chance of detecting a difference of 900 kg/ha using a 5% sig. level.

If we could use 20 observations.

$$\text{SE (for n = 20)} = 600\sqrt{(2/20)} = 186,$$

and we would now have a 5/6 chance of detecting a difference of 560 kg/ha.

There are a few additional comments. We can work with CV and ^ rather than s and d and everything is expressed in percentages but gives the same answer:

$$CV = 20\%, \hat{} = 25\%$$

$$25/3 = 20\sqrt{(2/n)}$$

which gives $n/2 = (20x3/25)^2 = 5.76$.

These calculations assume that the experimenter can assess the appropriate value for d. More crucially the calculations assume that s is known. In practice previous experience with the same crop for similar plots in similar conditions will often provide, through the √EMS, a credible value for s; sometimes less closely related information will have to be used to guess the likely value of s. If the actual experimental value of s is smaller than the assumed value, then our detection chances improve.

Note that although the SE decreases as n is increased, the reduction of SE for an increase of 1 to n gets smaller as n increases (the law of diminishing returns). We can make the statements about detection probabilities and significance probabilities more precise by expressing

$$d \quad \text{as } Z_\alpha + Z_\beta$$

where $Z\beta$ is the standardized normal deviate corresponding to an $*\alpha$ % significance level ($Z_\alpha = 1.96$ for $\alpha = 0.05$) and Z is the standardized normal deviate for a risk of non-detection ($Z_\beta = 1.0$ for $\beta = 1/6$).

Finally if we want to think about precision for interaction effects or other effects than the simple difference for two treatment means, then our formula for the standard error will change. For an interaction of two 2-level factors, which is the difference between two differences,

$$SE = \sqrt{(2s^2/n + 2s^2/n)} = \sqrt{(4s^2/n)}.$$

2.2 Replication, Hidden and Total

The actual number of observations relevant to a comparison may involve hidden replication as well as the apparent, explicit replication. Further, we may have to think which replication is relevant to our question. Suppose we have a set of 16 treatments comprising all combinations of 2^4 factorial structure with factors density, weed control, nitrogen and phosphorus. The full set of combinations are represented in Table 1 for a single site with 2 replications of each combination. Suppose further that we have 5 sites.

If we are interested in detecting an average nitrogen effect at one specific site, we have 16-fold replication for each N level (n=16) being made up of 2 explicit replicates x 8 hidden replicates since each N level occurs for the same eight combinations of density, weed control and phosphorus.

If we are interested in the average nitrogen effect averaging over all five sites than we have 80-fold replication.

If we are interested in the density x nitrogen combinations at one site we have 8-fold replication.

If we are interested in comparing the nitrogen effect at high density, no weed control and no P average across sites, we have 10-fold replication.

Thus the amount of replication, and equivalently, the information, for any comparison must be assessed for each specific comparison of interest.

The use of replications in assessing precision in the previous section has a further potential complication because the variability between plots within a site may be different (smaller) from the variability between sites. Thus, if we are considering the 80-fold replication for the N effect average over all sites our CV will tend to be higher than the 20% within site CV. The increase will depend on the CV of sites and if this were, say, 50% the overall CV would be approximately.

$$\sqrt{(20\%^2 + \frac{1}{16}(50\%^2 - 20\%^2))} = 23\%$$

Note that the increase in the SD or CV depends on the proportion of site replication in the total replication which is 1/16 in the example but would be 1/2 for the fourth example above with a much larger consequent increase in the CV. The amount by which the between-site CV is greater than the within-site CV may vary widely, but a factor of 2 or 3 might be a reasonable guess.

2.3 Efficient Use of Resources

In section 2.1 we considered the choice of n, the treatment replication, which must be considered in the context of different forms or replication (2.2). We now consider the efficiency of use of total resources in an experiment.

Suppose that in an experiment there are a total of N plots. Consider the total df (N-1). We use these in three ways:

(i) Estimating the error or random variance, s^2
(ii) Controlling, or identifying and allowing for causes of variation. This includes blocking, covariance adjustment, possible losses of plots.
(iii) Answering questions through treatment comparisons. Here, questions include those relating to main effects and interactions, but also possibly the modification of treatment effects by environments.

If we consider (i) first then it is usually accepted that a good experiment should have at least 12 df for error (some might say 10 or 15 instead of 12). It is equally important to recognise that there is very little benefit in having more than 20 df for error. An experiment with more than 20 df for error is inefficient. Surplus d.f should be transferred to (ii) to reduce s^2 and hence improve precision or to (iii) to provide answers to more questions.

For example, a simple Randomised Complete Block Design with 4 blocks and 12 treatments in each block is an inefficient design because it allocates 33 df to error. To redesign the experiment to make better use of the resources, we could try to use more df in (ii) or (iii).

One way of using more df in (ii) would be to use smaller blocks, dividing each complete block of 12 treatment plots into incomplete blocks of 4 or 6 treatment plots. The construction of incomplete block designs [(see Mead (1988) chapters 7 and 15] but will be worthwhile only if reducing block size reduces s^2 as should often be possible. More df could also be used in (ii) by identifying covariates which might explain some of the plot-to-plot variation.

More df could alternatively be used in (iii) by including an additional treatment factor at two levels and assessing the main effect of the factor and interactions with the original treatment factors. The upper level could be applied to six treatments in block 1 and the lower to the other six treatments with the pattern reversed in block 2, and the whole pattern or a similar one repeated in blocks 3 and 4. Further discussion of such confounded designs is given in Mead (1984, 1988).

I believe it is crucial in OFR that we do not simply transfer the often thoughtless and inefficient recipe designs used widely in research station experimentation. We must design experiments efficiently to use resources fully.

On station experiments are often regarded as good. I believe, on the contrary that they are often unimaginative, inefficient and boring and they achieve information only because of the overall high level of control and by being big.

Good designs could, in fact, reap very much greater rewards in OFR than they have been allowed to on research stations.

3. Plot Sizes

3.1 Plot size, choices and implications

It is quite widely believed that the variability which results when using small plots for OFR experiments is such as to make small plot OFR experimentation inappropriate. It is not clear that there is anything peculiar about the variability of plots on farms except that it is rather larger than that on research stations. The normal statistical expectation would be that the relation between the plot standard deviation, s, and plot area would be of the form

$$s = K/\sqrt{\text{Area}}.$$

This is essentially the same pattern as for replication

$$s = \sigma/\sqrt{n}.$$

Data from uniformity trials on farms collected and analysed by Hector Barreto seem to confirm the relationship between plot standard deviation and plot area. Hence, we should expect the usual statistical benefits of using smaller plots, allowing better control of the random variation, to apply on farms.

There are, however, other considerations. One is that the researcher (with the farmer) may not be able to use an appropriate form of blocking because of inadequate time to examine the particular farm situation and likely pattern of yield variation. Hence, the experimental design will not be efficient in controlling random variation and small plot benefits will be dissipated.

A second consideration is that small plots may seem quite unrealistic to the farmer and he may not apply the care of management which he would to larger plots.

Another consideration is that the total area of an experiment on farm is less than one on-station, and it is also true that the absolute level of variability, per area, on farms is greater than on station. Hence, it is inevitable that precision of results on farms will be reduced compared with that expected from station experiments. The disappointment with the level of precision on farms, ignoring information about expected precision which can be calculated as in section 2.1, may be partly responsible for the belief that small plots are the cause of poor precision in OFR.

The choice of plot size for future OFR experiments must continue to be a matter of judgement. If the level of control of variation through the researcher's knowledge of the particular farm conditions, expressed in careful choice of blocking, is good, then smaller plots may be appropriate. If such control is not feasible or if farmer preference and difficulty of management make small plots unacceptable, then we may have to use larger plots.

An almost inevitable consequence of using larger plots is that the number of plots per farm will be reduced and it is then possible that the precision of information *on a single farm* will be inadequate to produce useful conclusions *for that farm*. The implication would then be that we have to expect that most information would derive from analysis of the total (multiple farm) experiment. This further implies that the design of the whole experiment be considered as a whole rather than repeating an identical experiment at a number of farms. It is, of course, possible more generally that designing the whole experiment, with non-identical site-experiments, will be beneficial.

Certainly the use of larger plots does not necessitate a reduction in the total number of treatments (corresponding to the total number of questions asked). We may well need to use subsets of the total set of treatments at each farm and the design of experiments with different treatment subsets is discussed later. We must be prepared to design OFR Experiments to fit the particular resources and questions according to the principles of Sections 2.1 - 2.3.

4. Managing Variation

4.1. Information About Variation at the Farm Level

In some situations it may not be possible for the experimenter to make any assessment of the variability within the area made available for experimentation by the farmer. This necessarily makes designing efficient experiments much more difficult. If it can be assumed that the person responsible for local arrangements for the experiment both understands the concept of blocking and is capable of identifying the likely patterns of variation between plots then it is reasonable to use blocking in the design. In such a situation, where the experimenter cannot control, directly, the decisions about plot and block design I think

the correct philosophy is to try to limit the potential for the choice of unsuitable blocks. This can probably be achieved by setting an upper limit of 8 plots per block and requiring that the blocks be made compact.

However I shall assume that the experimenter does have the opportunity to assess variation within the proposed experimental area and that (s)he, in collaboration with the farmer, can make judgements identifying those plots likely to produce similar results. To avoid the suggestion that a "Block" must be rectangular in nature, I propose to call sets of plots judged to be similar "Groups". It is crucial that this identification of groups be separate from decisions about the choice of treatments. These two aspects of design inevitably interact later but we must first try to ensure that we have the best possible grouping of plots (as well as the most relevant set of treatments.

Let us consider an example from one of the on-farm experiments in the Poza Rica area. There were 24 plots in three rows of eight plots, as shown:

Row 3	17	18	19	20	21	22	23	24
Row 2	9	10	11	12	13	14	15	16
Row 1	1	2	3	4	5	6	7	8

Row 1 is at the bottom of the hill, row 2 above row 1, and row three above both.

The correct grouping system, identified by the experimenter, was three groups

(1,2,3,4,5,6,7,8) (9,10,11,12,13,14,15,16)
(17,18,19,20,21,22,23,24).

The experiment was actually designed in four blocks of six plots

(1,2,3,4,5,6)	(9,10,11,12,13,14)
(17,18,19,20,21,22)	(7,8,15,16,23,24)

You can probably guess how many treatments there were!

The correct design would have three blocks of eight plots each with each block containing all six treatments plus extra plots for two of the treatments, the extra two being different in each block. The irony of this particular situation is that the actual allocation of treatments to blocks was inevitably equivalent to the correct design (can you see why?) and the analysis that should be applied to the results should treat the data as if the design had been intended to be the correct design (i.e. 3 blocks of eight plots, not 4 blocks of six plots). However there is one difference between the allocation of treatments to plots in the ideal design and in that actually used and this is that the randomisation of the design in three blocks would include all possible allocations of treatments to the eight plots (see design example 1, section B).

Naturally experimenters cannot always expect to achieve the ideal design when trying to use an inappropriate block structure. Another example in the Poza Rica farms had 24 plots as shown:

Row 5	19	20	21	22	23	24
Row 4	13	14	15	16	17	18
Row 3	7	8	9	10	11	12
Row 2	4	5	6			
Row 1	1	2	3			

The site was clearly very variable with the rows climbing steeply from row 1 to row 5 and with poor growth at the right hand side of the area which curled some way round the hill. Sensible groupings would appear to be

(A) row 5,
(B) rows 1 and 2

with either	(C1)row 3	or (C2) plots (7,8,9,13,14,15)
and	(D1)row 4	and (D2) plots (10,11,12,16,17,18)

The actual design had two blocks of twelve plots (for the twelve treatments) and there was plainly a great deal of variation within the blocks (to be fair some of that variation would have remained even within the better blocks.

One more example from the Poza Rica experiments to show how too swift assumptions about natural patterns may lead to faulty conclusions. The 14 plots (for seven treatments) were as shown:

14	13	12	11	10	9	8
1	2	3	4	5	6	7

There was a small road running along the side of plots 1 to 7. Each row of seven plots covered a long area and it seemed unlikely that plots (1 to 7) and (8 to 14) would be sensible blocks. However when walking between the two blocks there was a very strong suggestion that those plots on the left (plots 1 to 7) for each treatment showed better crop growth than those on the right. This impression can obviously be verified at harvest but if true demonstrates how first impressions can be misleading. Further examination of the site showed that the second block (8 to 14) was clearly, if only slightly, higher than the first, providing further justification for the blocking pattern used.

In each on-farm experiment it is important, at the planning stage, to identify how the plots should be grouped so as to both minimise variation within groups and maximise that between groups. This may, and perhaps often should, produce groups of different sizes and later in the design process these ideal groupings may be modified. Nevertheless it is vital that the ideal grouping be first identified.

4.2. Recording and Use of Ancillary Information

Whenever experimenters, or even statisticians, observe field plots during an experiment there are many differences and patterns of growth which are apparent. Two which we observed during observation of the Poza Rica experiments concerned patchiness of Johnsongrass and a possible trend along each block away from the edge of the field. Field notebooks should, and often do, include anecdotal information collected in a systematic manner. If that anecdotal information were recorded in a crude quantitative form it could be used, later, through covariance analysis to improve the precision of treatment comparisons. The minimum form of record would be presence(1) or absence(0) of a characteristic. More usefully a three-.or two- point scale to record nil(0) mild(1) substantial(2) or total(3) level of the defect characteristic should allow adjustment for the effect of the characteristic.

The possible distance adjustment occurred for a tillage and cover crop experiment, for which the plots were as shown.

Orange	15	16	17	18	19	20	21
	14	13	12	11	10	9	8
Orchard	1	2	3	4	5	6	7

The visual impression was of increased growth at the orange orchard end of the experiment. If the trend is gentle then a covariate of the distance from the left-hand end of the experiment may be adequate. If the reduction in yield near the orange orchard is more sudden and severe then using the reciprocal of distance may be appropriate (see Example 10.2 in Mead (1988))

4.3 Discarding or Adjusting Data

It is expected, as part of the general philosophy of OFR, that considerable variability of results will be experienced, and there is an argument that, consequent on this expectation, all data should always be used precisely so that this expected variability should be displayed. A counter argument is that sometimes extreme observations or sets of data are patently not representative of the same population as the rest of the data and this should be recognised and the analysis and interpretation amended accordingly.

There are different levels at which the discarding of data may be considered. Within a site there are sometimes obvious causes for reduced yields on some plots. A very common example for maize is plots having a markedly low number of plants. Although total plot yield compensates partially for lower density through reduced competition there is usually a clear tend relating yield and plot density. This is an appropriate situation for covariance analysis and adjustment with total discarding of plot data only in extreme cases. In covariance the relationship between the principal variate (yield) and the concomitant variate (density) is assessed for each treatment (allowing for block effects if this is necessary) and the average trend is calculated. The treatment means are then adjusted to a common density level enabling valid comparison of treatments unaffected by particular deviations of density. The progress of the technique is demonstrated for four treatments each with two plots in Figure 1.

Rep 1		Rep 2	
Yield (kg/ha)	**Density/plot**	**Yield**	**Density**
2.25	33	1.45	22
1.90	30	2.15	40
1.95	16	2.80	35
2.55	32	2.85	39

The four slopes (in g/ha per plant/plot) are:

0.8/11
0.25/10
0.85/19
0.1 / 7

and the average slope is

$$\frac{0.8 + 0.25 + 0.85 + 0.1}{11 + 10 + 19 + 7} = \frac{2}{47} = 0.425$$

The adjustments are:

Yield	Mean Density	Adjustment to 35 plants/plot	Adjusted Yield
1.85	27.5	+ 7.5 x 0.0425	2.17
2.025	35	Nil	2.02
2.375	25.5	+ 9.5 x 0.0425	2.78
2.6	36.5	- 1.5 x 0.0425	2.52

It may be decided to adjust yields to some value other than 35, for example, to average achieved plot density. The principles and treatment differences are unchanged and the level of the covariate (density) to which the yields are adjusted should be chosen subjectively to obtain the most relevantly representative yields (to adjust to a target of 40 plants/plot would result in over-optimistic yields).

The linearity assumption required for covariance analysis may be improved if 1/density is used as the covariate. Whichever gives the greater Sums of Squares for the covariance term in the analysis may be used.

If covariance adjustment is used at several sites, the adjustment rates = covariance coefficient may be different at different sites. This may be due to random variation or to real differences in yield-density curves. Unless there are compelling reasons for believing that the covariance coefficients should be the same there is no reason against using different adjustments at different sites.

If there is no obvious covariate, it is still possible to make adjustments for any observed and quantifiable concomitant information. Thus, for example, if the end plot of each block give a unusually low yields a covariate taking the value zero for end plots and one for other plots can be used adjusting yields to a covariate value of one.

The decision whether or not to use covariance adjustment of treatment mean yields is, ultimately, a subjective one. It should certainly involve looking at the plots of the data (as in Fig. 1). The effectiveness of covariance adjustment will vary, but there will usually be some benefit in accuracy or precision if the data plot suggests a relationship. Covariance adjustment remains valid if the covariate is apparently dependent on treatments. The adjusted values then represent what would have been expected if a uniform value of the covariate had occurred and this can sometimes provide valuable information (in parallel with the treatment means of the covariate and of the unadjusted mean yields).

The basis for discarding values completely is more difficult and should normally be on the basis of the researcher's knowledge of unusual circumstances rendering a plot unrepresentative. In cases of real doubt it is legitimate to calculate analyses with and without the dubious data.

When the discard possibility refers to a whole site the situation is rather different. It is not usually appropriate to consider adjusting site results to a common level because one of the purpose of using multiple sites is to examine behavior of treatments over varying sites. It is much more appropriate then to consider the set of treatment effects and relate them to possible causative variables measures at each site. Essentially rather than discarding "unrepresentative" sites we should seek to separate them from the rest and analyze them in parallel with the main data.

Finally, various rules have been advocated for discarding sites based on CV, overall mean, check treatment performance, or % Error SS of Total SS. Counter examples demonstrating the inappropriateness of any of

those rules can be easily constructed. The safe approach is to retain all data but to seek to understand the effects of causative factors and to identify different groups of sites showing different patterns of results.

5. Choice of Experimental Treatments

There are four main types of experimental treatment structure. First, the essentially unstructured set of alternative treatments, such as a set of varieties or herbicides, where each treatment is of equal potential importance. Second, the complete factorial structure including, equally, each combination of levels of the factors included in the experiment. The third type could be defined simply as "the rest" between these two extremes. Typical examples are control treatments in the unstructured set; stepwise combinations, where each treatment is a particular modification of the previous treatment; subsets of a factorial structure omitting inappropriate combinations. The fourth type, which may also be within types two or three, is for levels of a quantitative factor where the particular levels are not chosen primarily for their direct interest but rather as representatives of the range of interesting levels of the quantitative factor.

The only general rule is that the selected set of treatments shall provide the best possible information about the questions which the experiment is purposed to answer. This statement,of course, assumes that the questions precede the choice of treatments rather than the reverse (which is not scientifically justifiable).

There is relatively little to say about type 1 structures except that the number of treatments should be determined by the number of interesting alternatives, always within the limitations of resources. The range of incomplete block design structures now available for testing large numbers of varieties, using any appropriate block sizes, is so comprehensive that there is no excuse for tailoring the number of varieties to suit a particular design structure.

5.1 Complete Factorials

Complete factorial structures, possibly omitting one or two unsuitable combinations, always provide a very powerful method of acquiring information because of their twin advantages: first they allow us to investigate whether there are important interactions, and second, whether or not there are interactions, the information about each separate factor effect will be more precise with factorial structure.

For initial experiments within a research programme factorial structured treatments will be suitable because they provide information about the existence of interactions between factors, thus allowing those interactions which are found to be unimportant to be ignored in subsequent experiments. The second advantage of factorial structures in giving more efficient information about main effects and two.factor interactions through the use of every combination in each effect estimate will, of course, also be beneficial in initial experiments.

However this second advantage becomes much more important in subsequent experiments where it is more efficient to continue to ask several questions in each experiment rather than relapsing to the classical scientific approach of asking only a single question in each experiment.

5.2 Quantitative Factor Levels

For the choice of levels of a quantitative factor we should be concerned to maximise the information about the pattern of response as a whole or about a particular characteristic of the response, such as the position of the maximum. General statistical theory shows clearly that both forms of information are maximised by

(1) choosing as wide a range of values of the quantitative factor as is consistent with the assumption that the general pattern of response over that range can be summarised by a simple form of response curve.

(2) using as few different levels as are required to estimate the response curve with one extra for assessing the adequacy of the response curve (or one for luck!).

These requirements are widely applicable and should be ignored only to use particular levels of importance.

- For the actual choice of levels it will often be at least approximately correct to use equally spaced levels (possibly on the log scale). If the pattern of the response curve is expected to be strongly skewed then the levels should be closer together where the response is changing rapidly and further apart in areas of lesser change.

5.3 Incomplete Factorial Structure

When the set of treatments is to be a subset of a factorial structure (that is several factors are varied in the set of treatments, but less than all the possible combinations are to included) then the consideration of the precision of comparisons from different subsets is very important. Precision increases according to the total number of combinations providing information about each comparison. This is the power of hidden replication. For a very simple example consider two alternative subsets each of four combinations from three two-level factors.

	Factor		
Subset(1)	**A**	**B**	**C**
Treatment 1	0	0	0
Treatment 2	1	0	0
Treatment 3	1	1	0
Treatment 4	1	1	1

	Factor		
Subset(2)	**A**	**B**	**C**
Treatment 1	0	0	0
Treatment 2	1	1	0
Treatment 3	1	0	1
Treatment 4	0	1	1

The precision of the estimate of the difference between levels 0 and 1 of factor A (or B or C) is more than twice as good for subset (2). That is the variance of the estimate of the difference using (2) is less than 50% of that using (1). The advantage derives from the use in (2) of all four combinations for estimating the difference as compared with using only two combinations in (1) plus the non-independence of the three estimates of differences in (1).

In general I believe the choice of treatments for an incomplete factorial structure has to reflect a balance between the objective of comparing, and being seen to compare, particular treatment combinations and that of estimating effects precisely. In a situation where previous experimentation has established that factors A, B and C act almost completely independently and where each main effect is believed to be substantial then the benefits for presentation of subset (1), may outweigh the consideration of precision. For example a sequence of treatments following the expected adoption sequence of farmers may be more understandable for farmers.

The statistical theory on simple subsets of factorial structures which are efficient for the estimation of main effects and two-factor interactions is well established. Nice fractions containing four, eight or sixteen combinations from structures with three, four, five or six two-level factors are easily found. Some examples are shown:

3 factors: 4 combinations
(000,011,101,110) or the complement (100,010,001,111)
4 factors: 8 combinations
(0000,0011,0101,0110,1001,1010,1100,1111) or complement
4 factors: 4 combinations
(0000,0110,1001,1111) or (0010,0101,1001,1110

5 factors: 16 combinations

(00000,00011,00101,00110,01001,01010,01100,01111,
10001,10010,10100,10111,11000,11011,11101,11110)

5 factors: 8 combinations

(00000,00110,01011,01101,10011,10101,11000,11110).

Suppose it is decided that a particular number of combinations are to be used and, further, that the presentation purposes both the combination of the lower level for all factors and the combination of the upper levels of all factors are to be included (note that this will not normally be statistically beneficial. The principles for choosing the other combinations are

(1) the two levels of each factor should be nearly equally represented,
(2) the four combinations of levels for each pair of factors should be equally represented,
(3) to remember the ideal properties of the "nice" fractions in maximising information about main effects.

Many practical situations using subsets of factorial structures require two-level factors and the logic of selecting appropriate subsets is clearer for two levels per factor. Nevertheless it is possible to select subsets from three- or four- level factors using the same principles. Thus, suppose we require twelve combinations from a 2x3x4 structure. A suitable set would be

(000,101,102,003,110,011,012,113,120,021,022,123)

which includes all the 2x3 combinations twice, all the 3x4 combinations once and the 2x4 combinations each once or twice.

5.4 An Example

To consider further the arguments pertinent to the choice of experimental treatments for type 3 structures we shall use an example of a verification trial from Ipiales (Woolley et al. 1988). The experiment was for a beans/maize intercropping mixture and the actual treatments used were

	Variety		Density			Seed
	Beans	**Maize**	**Beans**	**Maize**	**Fertiliser**	**Treatment**
1)	1	A	8	16	100	
2)	1	A	8	16	100	
3)	2	A	12	16	100	
4)	2	A	16	16	100	
5)	3	A	16	16	100	
6)	2	A	16	16	300	
7)	2	B	16	16	300	
8)	2	A	16	16	300	Yes

The treatments were designed in a step-wise fashion to assess the effects of a sequence of changes, depending on (a) the size of the effects detected in previous trials and (b) the expected adoption sequence by farmers. Treatment 1 was intended to be the individual farmer's practice in contrast to treatment 2 which was the to be the average practice of the group; in fact they emerged as virtually identical. Treatment 3 introduced an improved bean variety. Treatment 4 changed the proportion of beans/maize. Treatment 5 introduced a possible alternative, earlier,bean variety which might take better advantage of the increased proportion of beans. Treatment 6 added more fertilizer to treatment 4. For treatment7 an alternative maize variety (at the higher fertilizer level) was tried. Finally(treatment 8) a seed treatment was added to treatment 6. The logic of the stepwise evolution of treatments is simple and easily understood. In statistical terms it is also, unfortunately, inefficient in the use of resources. Each question which the treatments were

selected to answer is answered with minimum precision at each site because the answer involves the simple comparison of two experimental treatments. A full factorial experiment with 3 bean varieties, 2 maize varieties, 3(?) bean densities,2 fertiliser levels and + or - the seed treatment would require 72 experimental treatments and is plainly unthinkable.

The sequence of treatments 3 -4 -5 -6 -7 includes trying alternative varieties of beans and of maize increasing bean density and increasing fertiliser level. Because these changes occur in a particular order there is no opportunity to test different orders which would be appropriate if the relative sizes of the main effects differ from site to site. Moreover, if we wish to demonstrate the benefit of

(a) increased fertilizer,
(b) bean variety 2,
(c) maize variety B, and
(d) increased bean density

a subset of factorial structure provides much better estimates of the four effects than a sequence. Some possible changes to the treatment structure could involve:

(1) including all 8 combinations of bean(2or3) maize (A or b) and fertiliser(1 or 3) either at each site or in sets of 4 combinations per site;
(2) considering the seed treatment as an extra factor and treating half the combinations in each replicate;
(3) eliminating treatment 1 or 2;
(4) other variations on the lines of (1) incorporating the density change(treatment 4 to 5).

The detail of any experiment must always be determined through discussion and joint decision of the experimenter and statistician. However, a possible design in two blocks of eight plots per block would be

	Variety		Density			Seed
Block	**Beans**	**Maize**	**Beans**	**Maize**	**Fertiliser**	**Treatment**
1	1	A	8	16	100	NO
1	2	A	12	16	100	YES
1	3	B	12	16	100	NO
1	3	A	16	16	100	YES
1	2	B	16	16	100	NO
1	2	A	16	16	300	YES
1	3	A	12	16	300	NO
1	3	B	16	16	300	YES
2	1	A	8	16	100	YES
2	2	A	12	16	100	NO
2	3	A	12	16	100	YES
2	2	A	16	16	100	NO
2	2	B	16	16	300	YES
2	2	A	12	16	300	NO
2	3	B	12	16	300	YES
2	3	B	16	16	100	NO

All but three of the factorial combinations of bean variety, maize variety, bean density and fertiliser level are included, with two standard treatments repeated in each block, and the seed treatment imposed as an extra across the sets of treatments in each block.

6. Replication

There are always two levels of replication to consider for on-farm trials. Replication within each farm and replication between farms (and years). The purposes of these two forms are rather different. Replication within a farm provides a (usually) rather limited level of information about the precision of the results from that farm and also gives some protection against loss of individual plot information. Replication between farms provides information about the overall precision of the average results over farms and also allows the estimation of the variability of results between locations (and years).

The replication of the set of experimental treatments within sites was discussed in the report from my previous consultancy My conclusions then remain valid. Where there is a minimum of five sites, chosen fairly carefully to represent the variation between sites, then the use of two replicates of the set of treatments per site is sensible, except in those experiments where the primary interest is in the variation of effects over sites when a large number of sites is needed and within-site variation has little benefit. If fewer than five sites are used then it is likely that the replication may need to be more than two to achieve the necessary within-site precision.

One point that must never be forgotten is that factorial structures always provide hidden replication and when quite large factorial structures(at least 16 treatment combinations) are being considered, as they must be if maximum use is to be made of resources, then the insurance benefits of two explicit replicates are much less important since even with several missing values the factorial structure permits the reconstruction of values for all combinations.

7. Different Treatment Subsets at Different Farms

When discussing the choice of experimental treatments in section 3 we considered how a subset of the possibly interesting treatment combinations should be selected in such a way as to give good information about as many of the more important treatment effects as possible. Suppose we have a particular situation where we only have room for six combinations from four two-level factors, and these must include both (0000) and (1111). Suitable subsets would be

(0000,0011,0101,1000,1110,1111)

or (0000,0010,0100,1011,1101,1111)

or (0000,0100,0111,1001,1010,1111)

or (0000,0011,0110,1010,1100,1111).

If the experiment is to be at a number of farms there is a choice between selecting one subset and using it at all farms or using different subsets at different farms. Since the different subsets each provide only partial information and the partiality varies there is clearly advantage in changing the subsets between farms so that the combined information will be greater.

The following material from the 1989 Consultancy Report provides some additional ideas on the design and analysis of different factorial subjects on different farms.

7.1 Some Possible Development of Designs

With the development of computers it is possible to develop designs for experiments beyond the ideas of the 1930's which account for almost all the experimental designs used in agricultural research to-day (thirty years after the advent of computers).

The crucial ideas of experimental design are the use of blocking based on recognition of patterns of likely similarity among the available plots and the use of factorial structures, particularly subsets, to provide efficient information about main effects and 2-factor interactions. The advantages of factorial subsets can be seen by comparing a half-replicate of a 2^4 with various sets of non-factorial treatments. Consider the following three sets of treatment combinations (FP = Farmer's Practice).

Design 1	**Design 2**	**Design 3**
FP	FP	FP
FP + A	FP + A	FP + A + B
FP + B	FP + A + B	FP + A + C
FP + C	FP + A + B + C	FP + A + D
FP + D	FP + A + B + C + D	FP + B + C
		FP + B + D
		FP + C + D
		FP + A + B + C + D

If we consider only the estimation of (+A), (+B), (+C), and (+D), then designs 1 and 2 each provide less than 25% of the information from design 3. Allowing for the difference in numbers of observations designs 1 and 2 are still less than 40% as efficient as design 3.

The reasons why design 3 provides so much more information are:

(i) That it includes equal numbers of observations with and without A
(ii) Those with A and those without A each include B, C and D twice.

The first point ensures that as much information as possible about A is available; the second ensures that that information is completely unpolluted by the effects of B, C, and D.

Suppose we consider a fourth design which is allowed only six observations. We would like to have three with A and three without A and similarly for B, C and D. We would also like to minimize the interference between our four effects. The following design does quite well.

Design 4
FP
FP + A + B
FP + A + C
FP + A + B + D
FP + C + D
FP + B + C + D

The three observations for A include also two B, one C and one D, while those without A include one B, two C and two D. Given that each of B, C and D occur twice, this arrangement cannot be improved.

In thinking about constructing efficient designs with factorial subsets, we must concentrate on getting the balance of "with" and "without" right for each factor and then on minimizing interference. A simple half-replicate, such as design 3 allows a perfect solution, but other subsets can be nearly as good.

Of course all this discussion has ignored the possibility of interaction. Fractions of the complete factorial set constructed in the way outlined also provide the best possible information about interaction [Note that design 3 has two observations each for (i) A but not B, (ii) B but not A, (iii) both A and B, and (iv) both A and B, and that interference from C and D is zero]. Designs 1 and 2 in contrast provide no information on interaction.

If we are considering not the choice of a subset of treatments but different subsets for different blocks within a farm or for different farms, then the same principles apply. All other things being equal, we would prefer not to repeat subsets, but to use subsets not involving the same treatments. Thus, if design 3 were used in one block (or one farm), the ideal subset for a second block (or farm) would be:

Design 5

FP + A
FP + B
FP + C
FP + D
FP + A + B + C
FP + A + B + D
FP + A + C + D
FP + B + C + D

The combination of one block of Design 3 and one block of design 5 produces a classical confounded design with the four-factor interaction confounded. This would be a very good design if the circumstances were to be exactly appropriate. However, just like any other recipe design, it should be used only when the conditions of proper blocking, total resources and relevance of questions are suitable. If blocks of size 5 or 6 or 10 or 12 are clearly more suitable, then we should construct designs for those block sizes, and the arguments for numbers of treatments per farm are identical.

Of course there are questions about how designs such as 3,4 or 5 will be perceived by the farmer (perhaps also by the researcher's colleagues). This may lead to some modifications in designs and some explanation of designs in terms of capacity to examine changes, both individually and in combination. There is considerable further scope for developing designs and explanation of these principles.

The analysis of designs involving subsets of factorial structures allocated either to different farms or to different blocks within a farm can be completed on any computer program that can handle multiple regression analysis. This requires the definition of variables to represent the effects of interest together with dummy variables representing block and site differences.

To illustrate the analysis, we consider an experiment at three sites at each of which duplicate plots of a (different) subset of treatment combinations are used to comprise the experiment for that site. The three site experiments have six, six and nine treatments respectively; the first uses design 4, the last design 5 plus an FP treatment, and the second a design like design 4 chosen to complement the other two by including almost all the factorial combinations in the total experiment.

The designs and the yield data (artificial) are given in Table 2. The data format for any multiple regression program is shown in Table 3. The results (from GENSTAT) are shown in Table 4.

This is the basic analysis which provides estimates of A, B, C, and D main effects. Interaction effects can be estimated by constructing additional columns in Table 3 from the columns representing the relevant factors. Note that the first five columns after the yield column are the dummy factors for blocks and sites.

Analyses of variance to show SS for individual effects can also be constructed though the t values provide equivalent information.

There are somé correlations between the different effect estimates and these could be checked by requesting the correlation matrix for the estimates. For the model fitted here, the correlations are not large and can be ignored.

8. (Non-complete) Block Designs

We consider here the situation where one or more replicates of a set of treatments are to be divided into blocks and where the block size is less than the number of treatments. First, suppose that the experimental treatments are simply an unstructured set or, if there is some structure the important comparisons are between particular combinations rather than main effects and interactions. Then the division of each replicate into two or more blocks should be such that the divisions in different replicates are as different as possible and those treatments whose comparison is more important should tend to occur together in a block. The sense of "as different as possible" is that the treatments occurring together in a block in one replicate should be distributed evenly between the various blocks in each other replicate.

For structured treatment sets we first identify the treatment contrasts which are important. In a factorial structure these will almost always be the main effects, and probably also two-factor interactions. In other treatment structures the treatment contrasts will correspond to the questions which prompted the choice of the particular treatments. For these important contrasts we must arrange that each block provides maximal information. Thus, for a main effect, a1-a2, each block should include equal numbers of a1 and a2 observations. For an interaction effect, (a1-a2)(b1-b2), the four combinations, a1b1, a1b2, a2b1, a2b2 should all occur equally frequently in each block. For a contrast between a control group of treatments and an innovative set of treatments, each block should contain the same proportion of control:innovative.

In some cases the obvious block size does not allow a complete replicate of the set of experimental treatments to be contained in a set of blocks. In these cases we try to arrange that each pair of treatments occurs together in a block as nearly equally frequently as possibly. The requirement of equal occurrence for main effects still applies.

Examples of the construction of designs are included in part B.

9. Precision in Incomplete Block Designs

In incomplete block designs we trade the hope of a reduced value of sigma, the random variance (estimated by the error mean square) against some loss of information because we cannot compare each treatment with every other treatment in each block. One exception to this balancing act is confounded designs where the effects that can be estimated in each block suffer no loss of information to offset the gain from a smaller value of sigma.

Although we cannot estimate in advance the gain achieved througb a reduction in sigma, we can assess the loss of information from having to compare treatments occurring in different blocks indirectly. If treatment A occurs in block 1 and treatment B in block 2 and if treatments C,D and E occur in both then A and B may be compared by comparing each with the average of (C,D,E). The use of the intermediary treatments reduces the precision of the A-B difference by 33%. In incomplete block designs each treatment occurs several times and the web of comparisons through intennediaries becomes very complex.

To assess the loss of information from the use of a proposed design we can pre-test the design using a statistical analysis package. Any package capable of handling the analysis of a general block-treatment design would provide the information, but the simplest method available at CIMMYT is to use the

statistical package REML. The method is illustrated in the attached output for a design comparing twelve treatments in six blocks of six plots per block which is attached to the end of this document (3A).

The information required by REML is the block and treatment identification for each plot and a set of data values. For the illustration the plot allocation to blocks is (in plot order)

1	1	1	1	1	1	2	2	2	2	2	2	3	3	3	3	3	3
4	4	4	4	4	4	5	5	5	5	5	5	6	6	6	6	6	6

and the treatment allocation (same order) is

1	2	7	8	9	11	3	4	5	6	10	12	1	3	6	8	10	11
2	4	5	7	9	12	1	4	6	7	9	10	2	3	5	8	11	12

For data we use any set of simple numbers (all different)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36

The output summarises the standard errors for comparing pairs of treatments by giving the average, maximum and minimum standard errors. Since we have used nonsense data these will be nonsense standard errors. However, the relative values will be correct because relative precision depends only on the design and not on the data. Further REML prints the value of σ^2, the random variance. If we divide each standard error by the square root of sigma squared we obtain the standard errors as multiples of sigma. We can therefore assess exactly the loss of information which we expect to more than outweigh by the reduction in the value of sigma

In the attached example the standard errors for the nonsense data are

AVERAGE 0.385 MAXIMUM 0.404 MINIMUM 0.361

The value of σ^2 is 0.196 so that σ is 0.443 and the standard errors are σ multiplied by

AVERAGE 0.869 MAXIMUM 0.912 MINIMUM 0.815

Note first that the range of standard errors is about 5% either side of the average so that a single standard could be used in summarising the results of the experiment. Second the minimum possible standard error for comparing two treatments with three observations each is

$$\sigma\sqrt{(2/3)} = 0.816\sigma$$

which is the same (apart from rounding error) as the minimum achieved in our example design.

In general incomplete block designs are much more efficient than might initially be expected. A rough guide to precision in incomplete block designs is derived by considering the minimum possible variance between two treatment means

$$\text{MINVAR} = \sigma^2 \times 2/r$$

where r is the replication per treatment;
and the maximum possible variance which is

$$\text{MAXVAR} = \sigma^2 \times 2/k$$

where k is the average number of times that treatment pairs occur together in a block.

An excellent approximation to the average variance for treatment pair differences is then

$$VAR = MINVAR + (MAXVAR - MINVAR)/t$$

where t is the number of treatments.

For the example r = 3. The number of pairwise comparisons in each block is 15, so that over the set of blocks there are 6x15 = 90 pairwise comparisons within blocks and there are 66 treatment pairs. Thus k = 90/66. Hence

$$MINVAR = \sigma^2 \times 2/3 = 0.667\sigma^2$$

$$MAXVAR = \sigma^2 \times (2\times 66/90) = 1.467\sigma^2$$

$$VAR = \sigma^2\, 0.667 + \sigma^2\, (1.467 - 0.667)/12$$

$$= 0.733\sigma^2$$

giving an approximate standard error of $0.856\ \sigma$.

10. Loss of Plots/Sites

The loss of individual plot data causes, if anything, rather less problem for incomplete block designs than for complete block designs. For complete block designs the pattern of block-treatment structure is no longer complete when plots are missing and the data should be analysed as an incomplete block design. The alternative of estimating missing values is only approximately correct, is a throwback to the pre-computer days, and should not be necessary. For incomplete block designs the loss of plot data produces a different incomplete block design but no change in principle. We simply analyse the data we do have.

In either situation, if several, or all, plots of a particular treatment are lost the information about that treatment is badly affected, but the two design types suffer equally. The loss will be reduced when factorial treatment structure is used (complete or incomplete) because of the hidden replication benefits.

When different treatment subsets are used at different ocations the total loss of some sites should not cause problems provided the different treatment subsets have been chosen so that each location provides information on all or most of the factor main effects and two-factor interactions.

11. Analysis and Computers

The analysis of experimental data is, rightly, increasingly handled by the use of computer packages. These vary from those which can analyse a very restricted set of tightly specified designs to general statistical packages which can handle almost any design structure.

Where computer facilities are available the analysis of incomplete block designs can be managed using a general package, the most powerful being REML (the form of analysis information being illustrated in the example attached to this document), GENSTAT and, rather less informatively, by SAS. In the absence of any of these packages any block-treatment design structure can be analysed by a multiple regression package, as illustrated in my report on my 1989 consultancy, by defining a regression variable for each block and each treatment except block 1 and treatment 1. The regression coefficients then estimate the difference of each block, or treatment, from the first block, or treatment.

To the best of my knowledge (limited) the only smaller package designed for PC's which offers the possibility for analysing incomplete block designs is INSTAT (and even there I am not sure if that option is yet available in the presently commercially available version). However it can only be a matter of a short time before the better PC packages have facilities for analysing incomplete block designs.

Finally, it must be emphasised that it is possible to analyse data from any incomplete block design using only a small pocket calculator using the method of sweeping. This method is described in detail in part C. The only arithmetical operations involved are

(i) the calculation of means,
(ii) subtraction, and
(iii) the summing of squares.

For a large or complex design these operations are repeated many times. If computer facilities are available then of course they should be used. However sweeping is always a possible method of analysis and should be understood by all users of analysis of variance, not least because it displays the logic of the analysis clearly and because it is the principle utilised in the better statistical analysis packages.

Table 1. Plot treatments at a site

Density	Weed Control	Phosphorus	Zero N		80 kg/ha N.	
			Rep 1	Rep 2	Rep 1	Rep 2
High	No	Zero	x	x	x	x
High	No	40 kg/ha	x	x	x	x
High	Yes	Zero	x	x	x	x
High	Yes	40 kg/ha	x	x	x	x
Low	No	Zero	x	x	x	x
Low	No	40 kg/ha	x	x	x	x
Low	Yes	Zero	x	x	x	x
Low	Yes	40 kg/ha	x	x	x	x

Table 2. Designs and results for three sites.

Site 1	Block 1	Block 2
FP	1900	1300
FP + A + B	2500	2700
FP + A + C	3100	3300
FP + A + B + D	2400	3300
FP + C + D	2900	2000
FP + B + C + D	2400	2600
Site 2		
FP	2200	1600
FP + A + D	3500	2800
FP + A + B + C	2800	3600
FP + A + C + D	4100	2600
FP + B + C	2200	2500
FP + B + D	3400	2800
Site 3		
FP + A	3700	2800
FP + B	2300	1500
FP + C	1800	2800
FP + D	3700	2500
FP + A + B + C	3500	3600
FP + A + B + D	3500	4600
FP + A + C + D	4000	3300
FP + B + C + D	2600	3600
FP	2500	2100

Table 3. Data information for multiple regression for designs of Table 2

	Site 1	Site 2		Site 3		Factor			
Yield	Block	Block	Block	Block	Block	A	B	C	D
1900	0	0	0	0	0	0	0	0	0
1300	1	0	0	0	0	0	0	0	0
2500	0	0	0	0	0	1	1	0	0
2700	1	0	0	0	0	1	1	0	0
3100	0	0	0	0	0	1	0	1	0
3300	1	0	0	0	0	1	0	1	0
2400	0	0	0	0	0	1	1	0	1
3300	1	0	0	0	0	1	1	0	1
2900	0	0	0	0	0	0	0	1	1
2000	1	0	0	0	0	0	0	1	1
2400	0	0	0	0	0	0	1	1	1
2600	1	0	0	0	0	0	1	1	1
2200	0	1	0	0	0	0	0	0	0
1600	0	0	1	0	0	0	0	0	0
3500	0	1	0	0	0	1	0	0	1
2800	0	0	1	0	0	1	0	0	1
2800	0	1	0	0	0	1	1	1	0
3600	0	0	1	0	0	1	1	1	0
4100	0	1	0	0	0	1	0	1	1
2600	0	0	1	0	0	1	0	1	1
2200	0	1	0	0	0	0	1	1	0
2500	0	0	1	0	0	0	1	1	0
3400	0	1	0	0	0	0	1	0	1
2800	0	0	1	0	0	0	1	0	1
3700	0	0	0	1	0	1	0	0	0
2800	0	0	0	0	1	1	0	0	0
2300	0	0	0	1	0	0	1	0	0
1500	0	0	0	0	1	0	1	0	0
1800	0	0	0	1	0	0	0	1	0
2800	0	0	0	0	1	0	0	1	0
3700	0	0	0	1	0	0	0	0	1
2500	0	0	0	0	1	0	0	0	1
3500	0	0	0	1	0	1	1	1	0
3600	0	0	0	0	1	1	1	1	0
3500	0	0	0	1	0	1	1	0	1
4600	0	0	0	0	1	1	1	0	1
4000	0	0	0	1	0	1	0	1	1
3300	0	0	0	0	1	1	0	1	1
2600	0	0	0	1	0	0	1	1	1
3600	0	0	0	0	1	0	1	1	1
2500	0	0	0	1	0	0	0	0	0
2100	0	0	0	0	1	0	0	0	0

Table 4. GENSTAT output for data from Table 2.

Estimates of regression coefficient

	Estimate	SE	t
Constant (site 1 block 1)	1672	257	6.51
Site 1 block 2	0	296	0.00
Site 2 block 1	500	296	1.69
Site 2 block 2	117	296	0.39
Site 3 block 1	629	270	2.33
Site 3 block 2	540	270	2.00
A	851	159	5.36
B	121	159	0.76
C	211	159	1.33
D	541	159	3.41

Analysis of Variance Summary

	df	SS	MS
Regression	9	14266827	1585203
Residual	32	8382935	261967
Total	41	22649762	

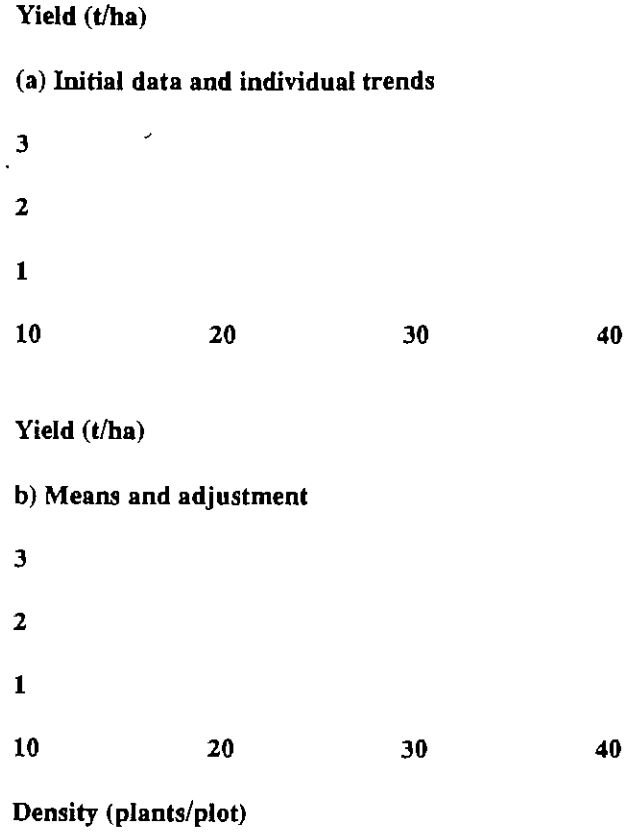


Fig. 1. Covariance Adjustment

Table 5

```
REML - Analysis by Residual Maximum Likelihood  (  96K version  )
(C) Scottish Agricultural Statistics Service
    University of Edinburgh

'TITLE'
Analysis of Design 1 with unstructured treatments
'UNIT'  36
'FACTOR'
Block     6  1 2 3 4 5 6
 FACTOR Block     CREATED
Treat    12  A B C D E F G H I J K L
 FACTOR Treat     CREATED
'VARIATE'    Yield
'FIXED'      Block  +  Treat
'DEPENDENT' Yield
FIXED MODEL READ
'READFREE' 1 Block
 1  1  1  1  1  1
 2  2  2  2  2  2
 3  3  3  3  3  3
 4  4  4  4  4  4
 5  5  5  5  5  5
 6  6  6  6  6  6
 DATA SET READ
'READFREE' 1 Treat
 A  B  G  H  I  K
 C  D  E  F  J  L
 A  C  F  H  J  K
 B  D  E  G  I  L
 A  D  F  G  I  J
 B  C  E  H  K  L
 DATA SET READ
'READFREE' 1 Yield
 1  2  3  4  5  6
 7  8  9 10 11 12
13 14 15 16 17 18
19 20 21 22 23 24
25 26 27 28 29 30
31 32 33 34 35 36
 DATA SET READ
'PRINT' 5
'SE'    2
'DEC'   3
'AT END'
'MEAN EFFECT' Treat
'CANONICAL"
'ENDPRINT'
'GO'
```

artificial data

Echos the input data and directives to the output stream as they are read

```
*** No RANDOM model specified - RESIDUAL term only will be used
INPUT READ
    36 EXPERIMENTAL UNITS
FIXED DF =     17


REML - RESIDUAL MAXIMUM LIKELIHOOD COMPONENTS OF VARIANCE
              ITERATION NO      2

Analysis of Design 1 with unstructured treatments
 ITERATIONS HAVE CONVERGED

MEAN EFFECTS (B.L.U.E.'S) of Treat
 A          B          C          D          E          F          G
   15.917     16.361     16.750     16.972     18.083     18.306     18.583

 H          I          J          K          L          MARGIN
   19.028     19.917     20.306     20.694     21.083     18.500

STANDARD ERROR OF DIFFERENCES BETWEEN PAIRS
AVERAGE       0.385     MAXIMUM     0.404     MINIMUM     0.361

CANONICAL DECOMPOSITION  T.INV_INF_MX.T  TO BE DIAGONAL

DIAGONALISED INVERSE_INF MATRIX  -  (D)
                    SIGMA
                    SQUARED
SIGMA      SQUARED       0.004

  Coefficient matrix - (T)

    SIGMA
    SQUARED
      1.000

TC = T. (Components_from_previous_iteration)  and  2*TC*TC/D  -  i.e.
Approximate Stratum Variances     and effective df

SIGMA      SQUARED          0.196              19.000
```

Best Linear Unbiased Estimates

Output from directive 'MEAN EFFECT'

Output from directive 'CANONICA

ERROR mean square

ERROR degrees of freedom

NOTE: In this example, the MINIMUM standard error of differences between pairs (0.361) is the standard error of differences between pairs of treatments concurring once, as for instance treatments A and B. Whereas the MAXIMUM (0.404) is the standard error of differences between pairs of treatments not concurring at all, as for instance A and E.

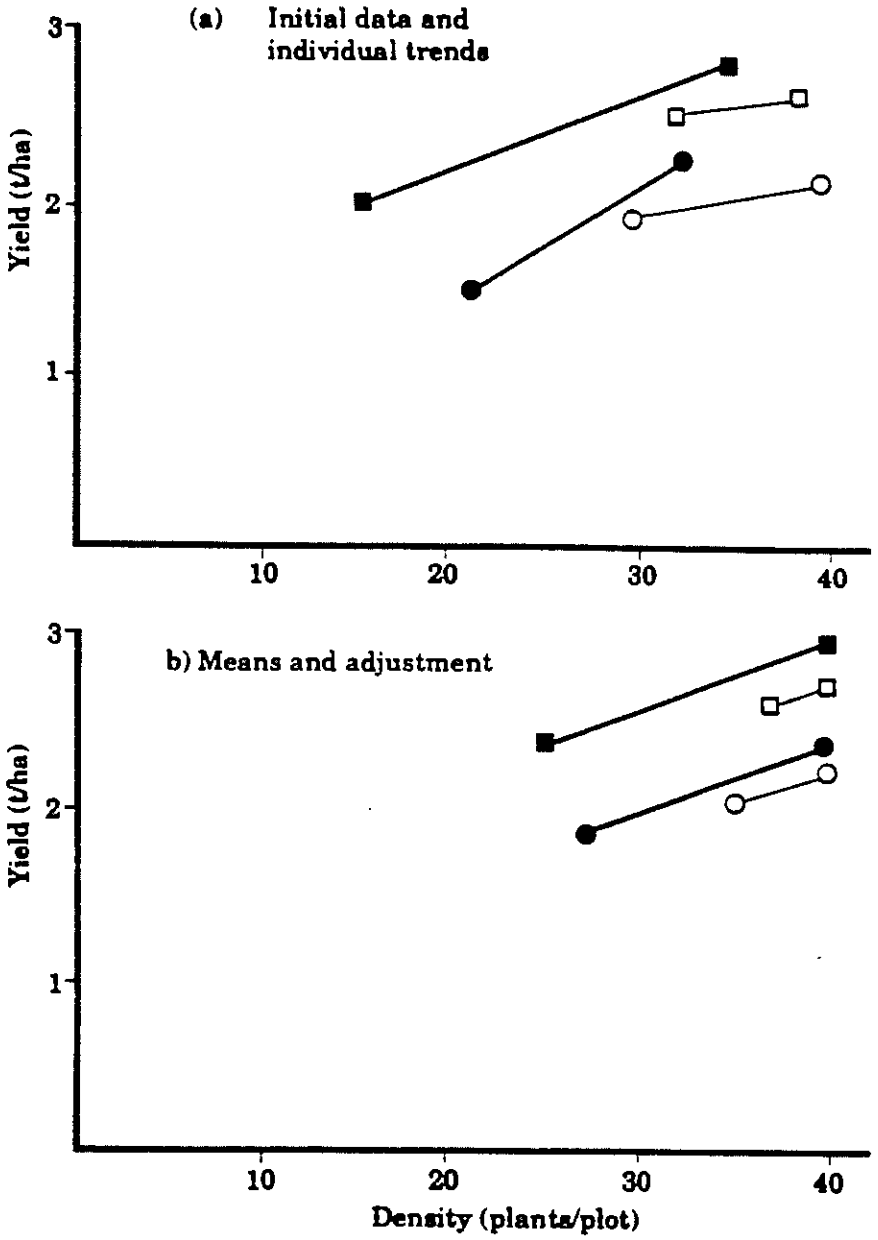


Fig. 1 Covariance Adjustment.

EXPERIMENTAL DESIGN PROBLEM EXAMPLES

These examples are intended to illustrate the general principles for fitting subsets of treatments into sets of blocks, the sizes of the blocks being determined so as to provide homogeneous plots within a block. In most cases it will be assumed that

(i) the set of treatments is pre-defined, possibly with factorial structure (complete or incomplete),
(ii) the experiment is to include 2, 3 or 4 replicates of the treatment set,
(iii) each replicate is to be divided into two or more blocks of the same or similar size.

The initial set of problems have been generated from the problems that (should) have been solved for various on-farm experiments at Poza Rica and Chalco in early 1990. Other problems from the past, or for the future, or purely hypothetical, will be added as they are suggested to me. (Such suggestions will be welcome!).

1. Principles

1.1 Treatments Without Structure

Ideally, each pair of treatments should occur together (in a block) equally frequently, or at worst, with frequencies differing by at most one. Particularly with relatively small numbers of treatments (e.g.twelve), this is not always possible to achieve. So we need some "rules" for the division into blocks in each replicate such that the resulting design will come as near to the equal pair-wise occurrence as possible.

Where each of several replicates is split into blocks the splits in the different replicates should be as different as possible. This is to be interpreted in the sense that, for the first two replicates, each block group of treatments in the first replicate should be split as equally as possible between the blocks of the second replicate. For further replicates the division of treatments into blocks should be as different as possible (in the same sense) to each of the divisions in previous blocks

1.2 Complete Factorial Structure

In each replicate the division into blocks should be such that for each factor in each block the levels of that factor should occur as equally as possible. This requirement is intended to maximize the information about each factor main effect. Further, for each pair of factors for which the two-factor interaction is likely to be important, all combinations of levels from those two factors should occur as evenly as possible in each block. Where there are still arbitrary choices to be made the equal occurrence in each block of all combinations of levels of three factors should be aimed at.

1.3 Other Structure

For incomplete factorial structures the general principle about equal occurrence in each block of the levels of each factor and of the combinations of levels of pairs of factors still applies. When the set of treatments includes some non-factorial structure, the important treatment contrasts should be identified. For each contrast each block should contain a similar balance between the groups of treatments compared in the contrast.

2. Examples

2.1 Twelve Treatments in Six Blocks of Six

This is the actual problem solved for the Chalco experiments being initiated at the beginning of April 1990.

Treatments:
Twelve, being three sowing dates x four varieties

	V1	V2	V3	V4
D1	1	2	3	4
D2	5	6	7	8
D3	9	10	11	12

The factorial structure is not important for the design of this experiment. All comparisons between treatment combinations are important. Comparisons between treatments (1,2,5,6,7,8,11 and 12) are expected to be more important as these combinations are expected to be more successful.

Replicates and Blocks:
Each replicate is to occupy a 4 x 3 grid of plots. The plots are approximately square. It is thought that the total area of each replicate may be rather too large to be properly homogeneous and that using two blocks of 2 x 3 within each replicate might provide greater homogeneity within blocks (and correspondingly differences between blocks within each replicate). If the differences between blocks within replicates turn out to be negligible the analysis can revert to that for a RCBD since each replicate is considered as a whole as well as being split into two blocks.

Design:
We wish to divide the twelve treatments into two groups(blocks) of six in each replicate in such a way that the divisions are as different as possible in the three replicates. The choice is too wide and we may find it helpful to use the two-factor structure to start the design. For the first replicate let the division be such that each block includes two combinations for each date and at least one combination for each variety. We try

Block 1 (1,2,7,8,9,11) Block 2 (3,4,5,6,10,12)

Now the next pair of blocks must each include three combinations from block 1 and three from block 2. It is still helpful to use the date and variety structure. So we try

Block 3 (1,3,6,8,11,12) Block 4 (2,4,5,7,9,12)

Quite a number of treatment pairs have not yet occurred together in a block (1 and 4, 2 and 3, 5 and 8, 6 and 7 and so on). We try to remember to include these as well as splitting treatments in the third replicate so that each block includes three from each of blocks 1, 2, 3 and 4. There are still plenty of good solutions and we try

Block 5 (1,4,6,7,9,10) Block 6 (2,3,5,8,11,12)

An alternative design with many similar patterns, but motivated by a strong desire to compare treatments 1 with 12, and 2 with 11, produced the following

Block 1 (1,2,7,8,11,12)	Block 2 (3,4,5,6,9,10)
Block 3 (1,4,6,7,10,12)	Block 4 (2 3 5 8 9 11)
Block 5 (1,3,5,8,10,12)	Block 6 (2,4,6,7,9,11)

Pre-testing Precision:
It is possible to give confidence in a proposed design by calculating, before the use of the design, the precision that will be achieved. The relative precision of different treatment comparisons within a design, and of alternative designs, is a property of the designs. Of course the absolute precision achieved will depend on the data and obviously that is not available before the experiment. However we can calculate the relative precision in advance with various computer programs (details given in main paper on "Design of on-farm experiments").

For the first design the range of standard errors for estimated treatment differences is

0.82 σ to 0.91 σ with a mean of 0.87 σ

where σ is the standard deviation of the random variation, estimated by the square root of the Error Mean Square.

The smallest achievable standard error, with three observations per treatment would be

$$\sqrt{(2\sigma^2/3)} = 0.82\ \sigma$$

so that the precision of the design is really very good. Certainly we should intend to use only a single standard error when presenting the treatment results from the analysis of the six block design.

For the second design with particular emphasis on comparing treatments 1 with 12 and 2 with 11 (this was the design actually used at Chalco) the range of standard errors for estimated treatment differences is

0.82 σ to 0.91 σ with a mean of 0.87 σ

exactly the same, to two figures as for the first design.

We consider four other designs to see how quickly the very high and consistent precision in both designs thus far deteriorates as we take less thought over the design. We shall think of the treatments in terms of their Varieties by Dates structure.

	V1	V2	V3	V4
D1	1	2	3	4
D2	5	6	7	8
D3	9	10	11	12

Suppose that we decided in the first replicate to put varieties 1 and 2 in block 1 and varieties 3 and 4 in block 2. In the second replicate varieties 1 and 3 in block 3, varieties 2 and 4 in block4, and the other variety pairings in the third replicate, producing the following (rather like a split-plot design!)

Block 1 (1,2,5,6,9,10)	Block 2 (3,4,7,8,11,12)
Block 3 (1,3,5,7,9,11)	Block 4 (2,4,6,8,10,12)
Block 5 (1,4,5,8,9,12)	Block 6 (2,3,6,7,10,11)

The range of standard errors for estimated treatment differences is

0.82 σ to 0.88 σ with a mean of 0.87 σ,

fractionally better than the two earlier designs but for all practical purposes unchanged.

Suppose we think in a "split-plot" pattern keeping similar date treatments together.

Block 1 (1,2,3,4,5,6)	Block 2 (7,8,9,10,11,12)
Block 3 (1,2,3,4,11,12)	Block 4 (5,6,7,8,9,10)
Block 5 (1,2,9,10,11,12)	Block 6 (3,4,5,6,7,8)

The range of standard errors for estimated treatment differences is

0.82 σ to 0.93 σ with a mean of 0.88 σ.

This time the precision is marginally worse than our first two designs but again the change is insignificant.

Suppose we just try pretty patterns within the structured set of treatments.

Block 1 (1,3,6,8,9,11)	Block 2 (2,4,5,7,10,12)
Block 3 (1,2,7,8,9,10)	Block 4 (3,4,5,6,11,12)
Block 5 (1,3,6,8,9,11)	Block 6 (2,4,5,7,10,12)

The range of standard errors for estimated treatment differences is

0.82 σ to 0.97 σ with a mean of 0.90 σ.

A little bit worse both in mean S.E. and in the increased range but really the penalty for lack of thought, and even for repearing the division of replicate 1 in replicate 3 is very small.

Finally we actually try to make the divisions into two blocks unnecessarily similar in the different replicates.

Block 1 (1,2,6,7,11,12)	Block 2 (3,4,5,8,9,10)
Block 3 (1,3,5,8,9,11)	Block 4 (2,4,6,7,10,12)
Block 5 (1,4,5,8,9,10)	Block 6 (2,3,6,7,11,12)

The range of standard errors for estimated treatment differences is

0.82 σ to 1.10 σ with a mean of 0.90 σ.

Well, we have produced at least one rather poor precision comparison but the mean is still not much worse than our best efforts and even using the mean S.E. for the maximum S.E. would hardly be a disaster. At least for this design problem the moral is that if one makes any real attempt to produce a design according to the defined principles it is actually rather difficult not to arrive at a good design.

2.2 Sixteen Treatments in a 4x4 Lattice

This is a problem for which there is a classical statistical design solution (lattice) but it is included here to illustrate the methods. The experiment, at Poza Rica, is to compare 16 varieties for drought tolerance. Four replicates, each split into four blocks are to be used.

This time we have no structure to guide us so we make an arbitrary split in the first replicate.

Replicate 1

Block 1	**Block 2**	**Block 3**	**Block 4**
(1,2,3,4)	(5,6,7,8)	(9,10,11,12)	(13,14,15,16)

For the second replicate each block must include one variety from each of the first four blocks.

Replicate 2

Block 5	**Block 6**	**Block 7**	**Block 8**
(1,5,9,13)	(2,6,10,14)	(3,7,11,15)	(4,8,12,16)

For the third replicate each block must include one variety from each of the first four blocks and one variety from each of the second four blocks. This requires slightly more thought than the second replicate but can be solved for the first block(9) and systematically thereafter.

Replicate 3

Block 9	**Block 10**	**Block 11**	**Block 12**
(1,6,11,16)	(2,5,12,15)	(3,8,9,14)	(4,7,10,13)

That was probably the hardest stage, and for the fourth replicate the choices are reduced and the problem gets a little easier. If we had made a less fortunate choice in the third replicate then we could have found ourselves with no choice in the fourth replicate. We would then have had to try a different third replicate.

Replicate 4

Block 13	**Block 14**	**Block 15**	**Block 16**
(1,7,12,14)	(2,8,11,13)	(3,5.10,16)	(4,6,9,15)

This completes the required design. Note that if a fifth replicate were needed there is one more division into four blocks which brings together all those pairs not previously linked. Note also that if we had only needed three replicates we could have stopped after the third because of the sequential nature of the construction.

Precision:
The range of standard errors for the estimated treatment differences is

0.791 σ to 0.817 σ with a mean of 0.796 σ.

The minimum possible S.E. would be

$$\sigma \sqrt{(2/4)} = 0.707\ \sigma.$$

But, of course, with no pair of treatments repeated together we cannot hope to be very close to that. This is a classical design of known high efficiency so that we should not be surprised that the range is very small. We can recognize also that for the design in blocks of four to be superior to the RCB in blocks of 16 the error variance for the blocks of four needs to be reduced by a factor of only

$$(0.707/0.796)^2 = 0.79$$

which should be more than likely with a sensible choice of blocks.

2.3 Only 15 varieties in Blocks of 3 and 4, or all 5

Suppose the number of varieties had been 15. There are two interesting alternatives. One would be to use the design for example 2 simply omitting one of the treatments and using the resulting mixture of blocks of three and four plots. The other would be to use three blocks of five plots per replicate.

For the first design we omit treatment 13 (arbitrarily) from the design, renumbering the subsequent treatments, and the resulting design is

Replicate 1

Block 1	**Block 2**	**Block 3**	**Block 4**
(1,2,3,4)	(5,6,7,8)	(9,10,11,12)	(14.15.16)

Replicate 2

Block 5	**Block 6**	**Block 7**	**Block 8**
(1,5,9)	(2,6,10,13)	(3,7,11,14)	(4,8,12,15)

Replicate 3

Block 9	**Block 10**	**Block 11**	**Block 12**
(1,6,11,15)	(2,5,12,14)	(3,8,9,13)	(4,7,10)

Replicate 4

Block 13	**Block 14**	**Block 15**	**Block 16**
(1,7,12,13)	(2,8,11)	(3,5,10,15)	(4,6,9,14)

For the design in blocks of five plots, we start by using an arbitrary split into the three blocks.

Replicate 1

Block 1	**Block 2**	**Block 3**
(1,2,3,4,5)	(6,7,8,9,10)	(11,12,13,14,15)

Now each block in the second replicate must have one or two varieties from each of the first three blocks.

Replicate 2

Block 4	**Block 5**	**Block 6**
(1,6,7,11,12)	(2,3,8,13,14)	(4,5,9,10,15)

Note that we inevitably had to repeat some joint occurrences (6 with 7, 2 with 3,etc.). It is probably useful at this stage to keep a note of which varieties have occurred with variety 1, with 2, and so on. Whether or not this is done we move on to the third replicate.

Replicate 3

Block 7	Block 8	Block 9
(1,4,8,13,15)	(2,5,6,10,11)	(3,7,9,12,14)

In the fourth replicate we try both to include those pairs of treatments which have not previously occurred together and to avoid any third repetitions of pairs

Replicate 4

Block 10	Block 11	Block 12
(1,4,9,11,14)	(2,6,8,12,15)	(3,5,7,10,13)

The ranges of standard errors are:

In blocks of 3 and 4

0.790 σ to 0.828 σ with a mean of 0.803 σ.

In blocks of 5

0.730 σ to 0.809 σ with a mean of 0.774 σ.

Again the precision of both designs is good compared with the minimum possible S.E. of 0.707 σ (and remembering that the σ in smaller blocks should be a good deal smaller). The unequal blocks of the first design and the loss of balance compared with the exact lattice have had only a marginal effect. The slightly less friendly blocks of five have produced a larger range (10% compared with 5%) but the average S.E. comes down (relative to σ) quite a bit with the blocks of 5.

2 4 Six Treatments in Three Blocks of Eight Plots.

An experiment from Poza Rica, mentioned in part A, in the section on variation at the farm level. Six varieties are to be compared and the natural blocking pattern for the 24 plots is three groups (rows) of eight plots per group. Each treatment will be replicated four times and the design problem is how to allocate sets of treatments to the three groups of eight plots.

The allocation must allow as many comparisons between different treatments in a block as possible. Therefore each block must include each treatment at least once.

Block 1	treatments	1	2	3	4	5	6	?	?
Block 2	treatments	1	2	3	4	5	6	?	?
Block 3	Treatments	1	2	3	4	5	6	?	?

We have one more observation for each treatment and the six remaining plots occur two in each of the three blocks. We therefore have no choice but to add two different treatments to each block. The choice of which pair of treatments to duplicate together is arbitrary; the treatments that are duplicated together will be slightly more precisely compared than other treatment pairs. The resulting block-treatment allocation is then

Block 1	treatments	1	2	3	4	5	6	1	2
Block 2	treatments	1	2	3	4	5	6	3	5
Block 3	treatments	1	2	3	4	5	6	4	6

When the treatments are randomized in each block all the eight "treatments" listed above are considered equally. The resulting randomization could look like

Plot	1	2	3	4	5	6	7	8
Block 1	5	1	4	2	1	3	2	6
Block 2	5	3	1	5	3	6	4	2
Block 3	6	4	4	1	2	6	5	3

Randomization always produces some odd-looking patterns and provided the blocking system correctly identifies the underlying pattern of plot-to-plot variation any randomisation is acceptable. If some randomisation results make us uncomfortable then the answer is to redefine the blocking system, not to try another randomisation. For example , in the above case we could decide to work with six blocks of four plots (a half-row per block) or to impose a column classification as well as a row classification. Neither of these would be appropriate here, I believe, since the rows do genuinely appear to be the most appropriate definition of blocks.

The range of standard errors for estimated treatment differences is

0.707 σ to 0.718 σ with a mean of 0.717 σ.

Compared with the minimum possible S.E. of 0.707 σ the use of the correct form of blocking has produced virtually no penalty of variable precision.

2.5 Fourteen Treatments in Blocks of 4, 5 and 6

The experiment includes 14 treatments, being seven varieties combined with two levels of nitrogen. The important treatment comparisons will be the difference between the two nitrogen levels for each variety and the differences between varieties for each nitrogen level. The nitrogen main effect is well-known and does not need to be reconfirmed.

The 28 plots for the two replicates of the 14 treatments are in two rows of ten plots and two rows of four plots. Differences between rows are likely to be large since the rows are at different contour levels.

Arrangement of plots:

28	27	26	25						
21	22	23	24						
20	19	18	17	16	15	14	13	12	11
1	2	3	4	5	6	7	8	9	10

Each row of four plots should probably be treated as a block and the first replicate should be completed with the block of plots 15 to 20. The second replicate has one block of four plots (11 to 14) and the other ten plots should be split into two blocks of five plots. The block pattern is therefore

1	1	1	1						
2	2	2	2						
3	3	3	3	3	3	4	4	4	4
5	5	5	5	5	6	6	6	6	6

We now have to divide the fourteen treatments into blocks of 4,4 and 6 in the first replicate, and into blocks of 4, 5 and 5 in the second replicate. The allocations should be such that treatments which occur together in a block in the first replicate do not again occur together in a block in the second replicate. We shall see that this requirement cannot be completely satisfied.

The allocation for the first replicate must be

(1,2,3,4), (5,6,7,8) and (9,10,11,12,13,14)

where we can decide later which actual treatments correspond to the labels 1 to 14. In the second replicate the treatments in a block in the first replicate should be evenly spread between the three blocks of the second replicate. This leads quite directly to

(1,5,9,10), (2,3,6,11,12) and (4,7,8,13,14).

Five pairs of treatments (2,3), (7,8), (9,10), (11,12) and (13,14) occur together twice and we should try to ensure that these are treatment combinations which we would particularly like to be precisely compared. Note, however, that the random variance with the blocks of 4, 5 and 6 should be much smaller than the random variance within complete blocks of 14 plots (1 to 14, and 14 to 28) as originally planned so that treatment comparisons should be more precise in the proposed design.

The range of standard errors for estimated treatment differences is

1.00 σ to 1.26 σ with a mean of 1.15 σ.

Compared with the precision results for our previous designs the variation here is rather disappointing. The minimum possible S.E. with blocks of 14 is 1.00 $\sigma(14)$ so we would be very confident that our more sensible blocks will reduce σ sufficiently that all S.E.'s will be smaller with the new design. The decision on whether to use the average S.E. is marginal but the maximum would be only 10% higher than the average so I would decide to use the average on the basis that if out S.E.'s are only 10% out from an ordinary analysis of variance we're doing pretty well.

2.6 Factorial Structure in an Unreplicated Trial

This trial at Poza Rica Station is described as an unreplicated observation trial and I have not seen the location of the trial. However since there are 36 treatment combinations I would question why it was not designed as an experiment. With 36 treatment combinations it is by no means clear that direct replication is necessary since hidden replication may be sufficient.

The 36 treatment combinations are

6 herbicides x 2 cover crops x 3 planting dates.

Assume that we are interested in main effects and in the combined effects of each pair of factors. The 36 plots should probably be grouped in six blocks of six plots per block, though detailed examination of the site might suggest alternative blocking patterns. In deciding the allocation of treatment combinations to blocks we would try to arrange

(i) all six herbicides (h) in each block,
(ii) both cover crops (c) to occur three times in each block,
(iii) each planting date (d) to occur twice in each block,
(iv) all six combinations of cover crop x planting date to occur in each block.

Other requirements for equal occurrence of combinations of pairs of factors in each block are impossible in blocks of six plots (Using three blocks of twelve plots would allow all combinations of herbicide x cover crop in each block). The design for six-plot blocks is constructed by allocating the six cover crop x planting dates to each block and then distributing the herbicide treatments so that no herbicide level is repeated in a block.

Block 1	Block 2	Block 3	Block 4	Block 5	Block 6
c1p1h1	c1p1h2	c1p1h3	c1p1h4	c1p1h5	c1p1h6
c1p2h2	c1p2h4	c1p2h6	c1p2h3	c1p2h1	c1p2h5
c1p3h3	c1p3h5	c1p3h1	c1p3h2	c1p3h6	c1p3h4
c2p1h4	c2p1h3	c2p1h5	c2p1h6	c2p1h2	c2p1h1
c2p2h5	c2p2h6	c2p2h4	c2p2h1	c2p2h3	c2p2h2
c2p3h6	c2p3h1	c2p3h2	c2p3h5	c2p3h4	c2p3h3

(The choice for the allocation of herbicides to block,c,p combinations is very wide and is equivalent to a Latin Square solution.)

The simple analysis of variance for the experimental data has the structure shown:

Source	df
Blocks	5
Herbicides	5
Cover crops	1
Planting dates	2
C x P	2
Error	20

All information on main effects and the Cx P interaction is fully efficient. Some information on the H x C and H x P can be recovered if a statistical package such as Genstat is used.

If the design with twelve plots in each of three blocks is used then the design will be:

Block 1	Block 2	Block 3
c1p1h1	c1p1h2	c1p1h3
c1p1h4	c1p1h5	c1p1h6
c1p2h2	c1p2h6	c1p2h4
c1p2h3	c1p2h1	c1p2h5
c1p3h5	c1p3h4	c1p3h1
c1p3h6	c1p3h3	c1p3h2
c2p1h3	c2p1h6	c2p1h4
c2p1h5	c2p1h2	c2p1h1
c2p2h6	c2p2h3	c2p2h5
c2p2h1	c2p2h4	c2p2h2
c2p3h4	c2p3h1	c2p3h6
c2p3h2	c2p3h5	c2p3h3

The analysis of variance structure is shown

Source	df
Blocks	3
Herbicides	5
Cover crops	1
Planting dates	2
H x C	5
C x P	2
Error	17

All information on the main effects and on the interactions of C with H and with P are fully efficient. Again some information on H x P could be recovered with Genstat.

SWEEPING METHODS FOR ANALYSIS OF VARIANCE

The conventional approach to analysis of variance is to calculate sums of squares for recognizable components of the total variation and to estimate the random variance from the remaining, "error", sum of squares. In a simple design, such as the Randomized Complete Block Design we can recognize that because of the orthogonality of blocks and treatments (each treatment occurs once in each block) the sums of squares for blocks and for treatments can be calculated quite independently. In more complicated, but still orthogonal,designs, such as two replicates of a factorial structure with four two-level factors arranged in four blocks of eight plots per block, we have to identify from the properties of the design which interaction sum of squares cannot be calculated (because of the confounding system).

In designs where each block contains a (different) subset of the set of treatments, blocks and treatments are not orthogonal. The sums of squares for blocks and for treatments are not now calculable independently and we have to think about the order in which we fit the terms "Blocks" and "Treatments" in the same way as for fitting terms in multiple regression. That is, we calculate the sum of squares for Blocks (ignoring treatments) and then the sum of squares for Treatments (after allowing for block differences).

For some particular designs, such as lattices or Balanced Incomplete Block Designs, standard methods for calculating the analysis of variance are given in text books (from Cochran and Cox onwards) and are available in some statistical computing packages. For other designs, with less regular patterns of treatment subsets in blocks, the analysis of variance can be calculated using powerful packages such as GENSTAT or (somewhat tediously) through a multiple regression package (an example of the multiple regression approach is given in Document 3A, section 7).

Whether our experiment is simple or more complex there is an element of the "sausage machine" about the calculations for the analysis of variance. This is particularly true of the calculation of the error sum of squares. There is an alternative approach to the analysis of variance and the estimation of treatment and block effects which, I believe, provides more insight into the concepts of the analysis and particularly the error sum of squares. It is not new and is in fact the basis of some of the better (and more flexible) statistical packages, but it does not appear to be widely known. Using this method we can analyse any design structure with no more than a pocket calculator (though for really complex structures the calculations may be rather tedious). The method is that of "Sweeping" the data.

In sweeping we identify the sets of effects (Blocks, Treatments, Main Effects, Main Plot Effects, Rows, Columns) which we wish to allow for in our analysis. Each yield will be labelled by one effect from each set: that is, each yield is in one block, has one treatment, etc. Essentially we define a model expressing the yield for each plot as a sum of several components.

For each set of effects in turn we estimate the effects and then subtract from each yield the value of the appropriate effect. After adjusting the yields to allow for all the relevant sets of effects we are left with the residuals which represent the random variation, not explicable by the sets of effects which we have considered, and the error sum of squares is simply the sum of the squared residuals. At any intermediate stage of the analysis the sum of squares of the currently adjusted yields provides a measure of the variation not yet accounted for.

Example 1: Randomized Complete Block Design

We start with a RCBD example for eight varieties in three blocks (G.Edmeades Ghana data 83T1 site15). We shall consider three components in our model:-

(1) the general mean
(2) the block effect
(3) the treatment effect

The separate consideration of the overall mean is not strictly necessary but is adopted to emphasize the general principles.

	Block		
	1	2	3
Varieties			
1	270	275	360
2	390	360	425
3	290	300	235
4	250	305	240
5	220	130	330
6	315	270	315
7	365	290	285
8	285	275	365
		298	
	-28	-23	+62
	+92	+62	+127
	-8	+2	-63
	-48	+7	-58
	-78	-168	+32
	+17	-28	+17
	+67	-8	-13
	-13	-23	+67
Mean	0	-22	+21

We first calculate the overall average.

We now sweep out this average from each plot value.

The resulting residuals are the set of deviations from the overall mean.

The sum of squares of these "residuals about the overall mean" is the total sum of squares = 89951.

Next calculate the mean for each block (note that apart from rounding error they sum to zero).

We now sweep out these values from the plot values in the corresponding blocks (note that we are sometimes subtracting negative numbers, e.g. -23 -(-22) = -1; 62 -(-22) = +84)

			Mean
-28	-1	+41	+4
+92	+84	+106	+94
-8	+24	-84	-23
-48	+29	-79	-33
-78	-146	+11	-71
+17	-6	-4	+2
+67	+14	-34	+16
-13	-1	+46	+11

The sum of squares of these residuals = 82293,
and the change from the total SS is
the block sum of squares = 7658.

Finally we calculate the means for each treatment as shown above and sweep these from the residuals to give the final residuals:

-32	-5	+37
-2	-10	+12
+15	+47	-61
-15	+62	-46
-7	-75	+82
+15	-8	-6
+51	-2	-50
-24	-12	+35

Notice that because we have swept out the effects of blocks and treatments the block and treatment totals of these residuals are all zero (apart from rounding error).

The sum of these final residuals is the error SS = 34779.
and the change from the previous sum of squares is the Treatment sum of squares = 47514

Explaining this example in detail has spread it out so now we bring all the calculations together in a compact form. At the same time we combine the first two stages by calculating the block means directly.

Treat	Block 1	2	3	Means						
1	270	275	360	-28	-1	-41	+4	-32	-5	+37
2	390	360	425	+92	+84	+106	+94	-2	-10	+12
3	290	300	235	-8	+24	-84	-23	+15	+47	-61
4	250	305	240	-48	+29	-79	-33	-15	+62	-46
5	220	130	330	-78	-146	+11	-71	-7	-75	+82
6	315	270	315	+17	-6	-4	+2	+15	-8	-6
7	365	290	285	+67	+14	-34	+16	+51	-2	-50
8	285	275	365	-13	-1	+41	+11	-24	-12	+35

Means

298 276 319 (Mean = 298)

Deviations

0 -22 +21

The calculation of the block deviations from the overall mean provides an alternative way of calculating the block sum of squares. We calculate the squares of the block effects for each plot and sum them, giving

$$8(0^2 + 22^2 + 21^2)$$
$$= 8(0+ 484+441) = 7600.$$

Because we have worked without decimals this is only approximately equal to the value calculated previously. The corresponding calculation for the treatment sum of squares is

$$3(4x4+94x94+23x23+33x33+71x71+2x2+16x16+11x11)$$
$$= 3(16+8836+529+1089+5041+4+256+121)$$
$$= 47676.$$

again approximately as before.

Finally we calculate the treatment means and the standard error of a difference between two means. The treatment means are calculated by adding the treatment effects (+4, +94, -23, -33, -71, +2, +16 and +11) to the overall mean to obtain

Treatment	1	2	3	4	5	6	7	8
Means	302	392	275	265	227	300	314	309.

The standard of a difference is calculated in the usual way,

$$(2(34779/14)/3) = 41.$$

Example 2: Four two-level factors in 4 blocks of 8.

This is a confounded design with much more structure than the simple RCBD. The data are from an experiment on factors of production in Ghana (G:Edmeades 79T1 Site 1).

The data are as shown:

Treatment				Block			
				1	2	3	4
A	B	C	D				
1	1	1	1	570		540	
1	1	1	2		550		700
1	1	2	1		585		520
1	1	2	2	900		640	
1	2	1	1		530		420
1	2	1	2	530		615	
1	2	2	1	500		425	
1	2	2	2		940		655
2	1	1	1		405		250
2	1	1	2	300		560	
2	1	2	1	305		400	
2	1	2	2		620		560
2	2	1	1	485		440	
2	2	1	2		460		410
2	2	2	1		570		390
2	2	2	2	670		675	

We would normally think of the analysis in two stages. First the calculation of Block, Treatment and Error sums of squares, and second the calculation of Main Effects and Interactions and their sums of squares. With sweeping we do the same. First calculate block means and subtract them from the yields.

532	582	537	488	MEAN = 535
				Means
+38		+3		+20
	-32		+212	+90
	+3		+32	+18
+368		+103		+236
	-52		-68	-60
-2		+78		+38
-32		-112		-72
	+358		+167	+262
	-177		-238	-208
-232		+23		-104
-227		-137		-182
	+38		+72	+55
-47		-97		-72
	-112		-78	-105
	-12		-98	-55
+138		+138		+138

The treatment means have been calculated above and we subtract them from the current residuals to obtain the purely random residuals.

+18		-17	
	-122		+122
	-15		+14
+132		-133	
	+8		-8
-40		+40	
+40		-40	
	+96		-96
	+31		-30
-128		+127	
-45		+45	
	-17		+17
+25		-25	
	-17		+17
	+43		-43
0		0	

The initial analysis of variance calculated from the block deviations from the overall mean. the treatment effects and the final residuals is

	SS	df	MS
Blocks	35350	3	11783
Treatments	543798	14	38843
Error	135209	14	9653

Note that there are only 14 df for the treatment SS because one treatment effect, the four factor interaction, is identical with the difference between blocks (1-2+3-4).

For the second stage of the analysis we consider the treatment means in systematic order.

Treatment Effect				Treatment Combination(ABCD)			
+20	+90	+18	+236	1111	1112	1121	1122
-60	+38	-72	+262	1211	1212	1221	1222
-208	-104	-182	+55	2111	2112	2121	2122
-72	-105	-55	+138	2211	2212	2221	2222

The means for the two levels of factor A are:

Level 1 (+20 +90 +18 +236 -60 +38 -72 +262)/8 = +66
Level 2 (-208 -104 -182 +55 -72 -105 -55 +138)/8 = -67.

The SS for the Main Effect of A is $(66^2 + 67^2)x16 = 141512$
and we subtract the means from the treatment effects to get:

-46	+24	-48	+170
-126	-28	-138	+196
-141	-37	-115	+122
-5	-38	+12	+205

The means for the two levels of factor B are -9 and +10, the SS for the Main Effect of B is 2888 and the reduced effects are

-37	+33	-39	+179
-136	-38	-148	+186
-132	-28	-106	+131
-15	-48	+2	+195

The means for the two levels of factor C are -50 and +50, the SS for the Main Effect of C is 80000 and the reduced effects are

+13	+83	-89	+129
-86	+12	-198	+136
-82	+22	-156	+81
+35	+2	-48	+145

The means for the two levels of factor D are -76 and +76, the SS for the Main Effect of D is 184862 and the reduced effects are

+89	+7	-13	+53
-10	-64	-122	+60
-6	-54	-80	+5
+111	-74	+28	+69

Notice that these reduced effects are now generally much less than the treatment effects with which we started. We can observe how rapidly they diminish at each stage by summing the squares of the reduced effects. After subtracting the effects of the four main effects the remaining sum of squares is:

$(89^2 + 7^2 + 13^2 + \ + 28^2 + 69^2) \times 2 = 132694$ on 10 df.

compared with the total treatment SS of 543798 on 14 df and the error SS of 135209 on 14 df.

At this stage we might decide that the reduced SS is now sufficiently close to what would be expected, based on the error mean square, (the F ratio is

$$(132694/10) / (135209/14) = 1.37)$$

that we should not examine the interaction effects. However we shall continue a little further, if only for illustrative purposes.

To, calculate, and adjust for, a two-factor interaction effect, we must first calculate the means for the four combinations of levels of the two factors. Consider the AxB interaction, for which the four combinations are the four rows of the table of treatment effects. The four means are +34, -34, -34 and +34 (we should not be surprised that the numbers are all the same since there is only one df for this interaction effect). The reduced effects are:

+55	-27	-47	+19
+24	-30	-88	+94
+28	-20	-46	+39
+77	-108	-6	+35

Observation of the pattern in this table suggests that the four columns are each either all + or all -. This corresponds to the CxD interaction (check with the original table of treatment combinations) and this would seem to be the next effect to consider. The means for the four columns are +46, -46, -46 and +46. and the reduced effects are:

+9	+19	-1	-27
-22	+16	-42	+48
-18	+26	0	-7
+31	-62	+40	-11

and are clearly now very small. The sum of squares of these reduced effects is now 27110 on 8 df and there is clearly no point in searching for further effects. If we had wished to examine other effects we would use the table of treatment combinations to identify the four sets of four values from which we should calculate means.

The tables of means which we require to summarize the results are two-way tables for (1) factors A and B and (2) factors C and D. The means are calculated from the mean effects previously calculated and the overall mean. Thus for the four combinations of A and B the combination means are calculated:

Factor	**A**	**B**			
Levels	1	1	535 +66 -9 +34	=	626
	1	2	535 +66 +10 -34	=	577
	2	1	535 -67 -9 -34	=	424
	2	2	535 -67 +10 +34	=	502
	1	Mean	535 +66	=	601
	2	Mean	535 -67	=	469
	Mean	1	535 -9	=	526
	Mean	2	535 +10	=	545

The standard form of presentation for the two-way table is:

Factor B	**Factor A**		
	1	**2**	**Mean**
1	626	577	526
2	424	502	545
Mean	601	469	

Standard error for comparing marginal means:

$$\sqrt{(2(135209/14)/16)} = 35$$

Standard error for comparing means in the table:

$$\sqrt{(2(135209/14)/8)} = 49$$

The two-way table of means for factors C and D is constructed in the same manner:

Factor D	Factor C		
	1	2	Mean
1	455	462	459
2	515	708	613
Mean	485	585	

The standard errors are exactly as for the table for factors A and B

Example 3 Incomplete Block Design

When each block contains a different set of treatments the result of sweeping, first by blocks and then by treatments, will not leave residuals which sum to zero both for each block and for each treatment. Consider a very simple example with four treatments in four blocks, arranged so that each block includes only three of the treatments.

	Block			
	1	2	3	4
Treatment				
A	410	260	360	
B	510	370		320
C	640		590	430
D		510	640	430

To understand the process better we shall consider first the situation where all 16 combinations are available.

	Block			
	1	2	3	4
Treatment				
A	410	260	360	190
B	510	370	480	320
C	640	470	590	430
D	650	510	640	430

Sweeping by blocks and treatments we get

Means	552	402	517	342	Mean = 454
	(+98)	(-52)	(+63)	(-112)	
	-142	-142	-157	-152	-148
	-42	-32	-37	-22	-33
	+88	+68	+73	+88	+79
	+98	+108	+123	+88	+104
	+6	+6	-9	-4	
	-9	+1	-4	+11	
	+9	-11	-7	+9	
	-6	+4	+18	-16	

The mean of these residuals in each block and each treatment is effectively zero apart from rounding error. The error sum of squares is the sum of squares of the residuals:

$$(+6)^2 + (+6)^2 + (-9)^2 + \ldots + (+18)^2 + (-16)^2 = 1352$$

Now consider the incomplete block situation where we have only three treatments in each block.

	Block			
Treatment	**1**	**2**	**3**	**4**
A	410	260	360	
B	510	370		320
C	640		590	430
D		510	640	430

Sweeping by blocks and treatments we get:

Means	520	380	530	393	Mean = 456
	(+64)	(-76)	(+74)	(-63)	
	-110	-120	-170		-133
	-10	-10		-73	-31
	+120		+60	+37	+72
		+130	+110	+37	+92
	+23	+13	-37		
	+21	+21		-42	
	+48		-12	-35	
		+38	+18	-55	

It can be seen immediately that the means in each (block) column are not zero. This requires that we sweep again to eliminate differences between blocks. Why has this happened? If we look back at the block means first in the complete case and then in the incomplete case we see that the means for block 1 are 552(complete) and 520(incomplete). The mean for the incomplete case is too low because treatment D, which is the best treatment, was missing in block 1. Hence, we have not fully allowed for the high level of yields in block 1 and the further sweeping will do this.

So we sweep again by blocks.

Mean	+31	+24	-10	-44
	-8	-11	-27	
	-10	-3		+2
	+17		-2	+9
		+14	+28	-11

Now, for exactly the same reason, we find that the means for the treatments are not zero. So we sweep again by treatment and continue to sweep alternately by block and treatment until we achieve residuals which give zero means for blocks and treatments.

Treatment Means					
-15	+7	+4	-12		
-4	-6	+1		+6	
+8	+9		-10	+1	
+10		+4	+18	-21	
Means	+3	+3	-1	-5	
	+4	+1	-11		-2
	-9	-2		+11	0
	+6		-9	+6	+1
		+1	+19	-16	+1
	+6	+3	-9		
	-9	-2		+11	
	+5		-10	+5	
		0	+18	-17	
Means	+1	0	0	0	
	+5	+3	-9	0	
	-10	-2		+11	0
	+4		-10	+5	0
		0	+18	-17	0

So we have arrived at last! The error SS is calculated as usual by summing the squares of the final residuals

$$(+5)^2 + (+3)^2 + (-9)^2 + \ldots + (+18)^2 + (-17)^2 = 1094.$$

The analysis of variance needs some care because the SS for blocks and treatments depend on the order in which they are fitted. If, as in the procedure above, we first fit blocks (ignoring treatments) the block SS is calculated from the block totals or means. The treatment SS is calculated from the change in the sum of squares of the residuals from those obtained after sweeping for blocks (the first time) to those after the sweeping is completed. Thus

the block SS is $((64)^2 + (-76)^2 + (74)^2 + (-63)^2) \times 4 = 78468$,

the SS of residuals after sweeping blocks only = 110667,

the SS of residuals after the final sweeping = 1094,

and the treatment SS(adjusting for blocks) = 109573.

Hence we have the analysis of variance

	SS	df	MS
Blocks (ignoring treatments)	78468	3	26156
Treatments (adjusting for blocks)	109573	3	36524
Error	1094	5	219

The block and treatment effect estimates are built up from the results of the repeated sweeps. Thus for blocks we have

Block	1	2	3	4
1st sweep	+64	-76	+76	-63
2nd sweep	+31	+24	-10	-44
3rd sweep	+3	+3	-1	-5
4th sweep	+1	0	0	0
Total	+99	-49	+65	-112

Treatment	1st	2nd	3rd	Total
A	-133	-15	-2	-150
B	-31	-4	0	-35
C	+72	+8	+1	+81
D	+92	+10	+1	+103

Notice how closely these estimates of block and treatment effects correspond to those from the complete block data set, which contains the same data points as the incomplete block data plus four extra observations.

	Complete	Incomplete
Block 1	+98	+99
Block 2	-52	-49
Block 3	+63	+65
Block 4	-112	-112
Treatment A	-148	-150
Treatment B	-33	-35
Treatment C	+79	+81
Treatment D	+104	+103

We would expect such agreement because we are trying to estimate the same quantities, the only change being that in the incomplete design we have less information on which to base our estimation. The reduced information is clearly sufficient to obtain estimates close to those based on fuller information. We can also compare the final residuals for the complete and incomplete cases. Again the agreement is good and not surprising.

Complete				Incomplete			
+6	+6	-9	-4	+5	+3	-9	
-9	+1	-4	+11	-10	-2		+11
+9	-11	-7	+9	+4		-10	+5
-6	+4	+18	-16		0	+18	-17

Finally we would wish to compare the treatment mean yields. The calculation of treatment means is already almost completed. We merely have to add the overall mean to the estimates of the treatment effects. The calculation of standard errors is more difficult and is one aspect of the sweeping technique where an exact solution is not possible with manual calculation. However, we can use an approximation which generally gives excellent results.

We calculate two variances for each treatment pair, one based on the total number of observations for each treatment (MIN), the other based on the number of blocks in which both treatments appear (MAX). If the number of treatments in the experiment is t, then the variance for a difference between two treatment means is

$$MIN + (MAX - MIN)/t.$$

For our incomplete block design

$$MIN = 2(219)/3 = 146$$

$$MAX = 2(219)/2 = 219$$

$$Var = 146 + (219 - 146)/4 = 164.25.$$

Hence we have treatment means and standard error

Treatment	A	B	C	D
Mean	306	421	537	559

Standard error of difference = 12.8

Example 4: Reblocking an on-farm experiment

Sometimes it may become clear on observing an experiment that the blocking should have been arranged differently. It is then possible to redefine the blocking to match the pattern which is believed to correspond to the real field variation. This should not be done lightly and particularly it is a dangerous procedure if many alternative post-experiment blocking systems are tried, and the most successful one used, or if the re-blocking is based on the numerical yield data rather than on practical assessment of the plot patterns.

The danger derives from the prospect that by trying too hard to define the correct re-blocking system the estimate of the random plot variance, deduced from the error mean square, will be biased downwards. The error mean square is, under normal randomization procedures, an unbiased estimate of the plot variability, after allowing for block differences and treatment effects. It will underestimate the normally expected plot variability if the form of analysis is pressured too much to make it small rather than appropriate. It is possible to try too hard!

When we use a different blocking system from that intended in the original design specification it is very likely that the treatments will not be complete in each block. In the example considered here the original design was for five herbicide treatments in three randomized complete blocks. On inspecting the plot Jonathan Woolley and I felt that there were clear patches running diagonally across the blocks. The experimental plan was

Block 3	Plot 11 T 1	Plot 12 T 4	Plot 13 T 3	Plot 14 T 2	Plot 15 T 5
Block 2	Plot 10 T 2	Plot 9 T 5	Plot 8 T 4	Plot 7 T 3	Plot 6 T 1
Block 1	Plot 1 T 5	Plot 2 T 1	Plot 3 T 2	Plot 4 T 4	Plot 5 T 3

The patches perceived by us were approximately

(1) plots 4,5,6,7,15;
(2) plots 2,3,8,13,14;
(3) plots 1,9,10,11,12.

The plots were scored by each of us and the combined score used as a measure of performance. The data for the revised blocking were

	New Block		
	1	2	3
Treatment			
1	4	13	9
2		10 , 17	9
3	3 , 3	9	
4	5	5	8
5	6		7 , 5

We sweep as usual alternatively by blocks and by treatments. The figures hereafter are multiplied by 10 to avoid decimals and reduce spatial confusion.

Block Mean	42	108	76				
				Treat Mean		**Residuals**	
Residuals	-2	+22	+14	+11	-13	+11	+3
		-8 +62	+14	+23		-31 +39	-9
	-12 -12	-13		-14	+2 +2	-4	
	+8	-58	+4	-15	+23	-43	+19
	+18		-6 -26	-5	+23		-1 -21
				Block Mean	+7	-6	-2
				Treat Mean		**Residuals**	
Residuals	-20	+17	+5	+1	-21	+16	+4
		-25 +45	-7	+4		-29 +41	-11
	-5 -5	+2		-3	-2 -2	+5	
	+16	-37	+21	0	+16	-37	+21
	+16		+1 -19	-1	+17		+2 -18
				Block Mean	+2	-1	0
				Treat Mean		**Residuals**	
Residuals	-23	+17	+4	-1	-22	+18	+5
		-28 +42	-11	+1		-29 +41	-12
	-4 -4	+6		-1	-3 -3	+7	
	+14	-36	+21	0	+14	-36	+21
	+15		+2 -18	0	+15		+2 -18
				Block Mean	0	0	0

The sum of squares of the residuals is 6010 (60.1 in terms of the original data). For comparison the Error SS for the analysis using the original blocks was 58.3. The new blocks do not appear to be an improved description of the pattern of variation between the plots.

References

Mead, R. (1984). Confounded experiments are simple, efficient and misunderstood. *Experimental Agriculture* 20: 185-201.

Mead, R. (1988). *The design of experiments: Statistical Principles for Practical Application.* Cambridge University Press.

Woolley, J.N., Beltran, J.A., Vallejo, R.A., Prager, M. (1988). Identifying Appropriate Technologies for Farmers: The case of the bean and maize system in Ipiales, Colombia, 1982-1986. CIAT Working Document 31. CIAT, Cali, Colombia.

Centro Internacional de Mejoramiento de Maíz y Trigo
International Maize and Wheat Improvement Center
Lisboa 27, Apartado Postal 6-641, 06600 México, D.F. México