THE DESIGN AND MIXED-MODEL ANALYSIS OF EXPERIMENTS

IV. General principles in designing experiments

(ref. Mead, secs 6.1, 9.2, 12.1)

IV.A Basic principles in designing experiments

The three basic principles, introduced by R.A. Fisher in 1935, are: replication, randomization and blocking. What I want to do now is outline the role each of these plays in experiments.

Why use statistical principles in the design of experiments? — The short answer is uncontrolled variation.

I have two neighbouring vines in the field? They have been pruned the same, same fertilizer application rate, same trellis, watering regime, rootstock. Indeed, as far as is humanly possible they have been treated exactly the same.

The fruit from each vine is harvested and weighed separately. It is extremely unlikely that the yields will be the same. A small difference is likely to be observed. Let's suppose that the yields of the two vines are 2 kg. and 2.25 kg., respectively. This difference is likely to have been caused by a large number of small *uncontrollable* differences, viz. slight differences in

1. soil — structure, depth, nutrients, moisture
2. pest attack — birds, virus, etc.
3. plant material — differences between original cuttings
4. management differences — stuck dripper, slightly unequal fertilizer etc.

Definition IV.1: **Uncontrolled variation** is variation between units treated as similarly as possible that arises from all the minor differences which we are unable to control. ■

Example IV.1 Pruning experiment

The reason uncontrolled variation is a problem can be seen by considering the following scenario. Suppose someone wants to run an experiment to compare two vine pruning treatments. She takes two neighbouring vines (not the pair above) one of which she prunes using standard practice (A) and the other she prunes more heavily (B). This is the only conscious difference in her treatment of the vines. Everything else is kept as similar as possible.

B	A
x	x

She records the yields of the vines which turn out to be 2.02 kg. for A and 2.48 kg. for B.

Does this different in yield indicate that the pruning treatments lead to yield differences? YES/NO.

Answer: Don't know as the yield difference may be due entirely to uncontrolled differences between the vines or may be due to a combination of uncontrolled differences and pruning treatment differences?

That is, the difference between the prunings is *mixed up* with vine (uncontrolled) differences.

Definition IV.2: Two effects are said to be **confounded** when it is not possible to separately estimate them. ■

In this case, the pruning effect is confounded with the vine effect.

The problem of the confounding of treatment effects with uncontrolled variation is widespread in the biological, physical and social sciences.

How does one overcome it?

Answer: Use statistical principles in the design of the experiment. N.B. We do not eliminate uncontrolled variation, rather we adopt strategies that enable us to live with it.

The statistical principles to be used are the three I mentioned previously:

a) Replication
b) Randomization
c) Blocking

a) Replication

- provides a measure of uncontrolled variation
- the application of each treatment several times

Example IV.1 Pruning experiment (continued)

So we might lay out the experiment, using a row of 20 vines, as follows:

A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Each vine is harvested separately and the yields are as follows:

Pruning Level	**Vine yields**										Range
A	**1.88**	2.48	2.02	2.28	2.22	1.96	2.36	2.48	2.42	2.08	0.60
B	1.96	2.24	2.16	2.36	2.30	1.90	2.44	**2.56**	2.50	2.16	0.66

$$\text{Total Range} = 2.56 - 1.88 = 0.68$$
$$= \text{Range within Pruning levels}$$

Now what is the experimenter interested in here? What is the question to be answered? Does pruning treatment lead to **yield difference**.

What statistics are we going to use to measure this? *Answer:* mean or median.

Let's use mean: $\bar{y}_A = 2.218$, $\bar{y}_B = 2.258$

What is our conclusion? Clearly, mean for heavy pruning (B) is slightly higher than that for standard pruning (A).

But what could contribute to such differences? *Answer:* Pruning and uncontrolled differences? Well, has pruning contributed in this instance? *Answer:* Can't tell from just looking at the means. Need to get a measure of uncontrolled variation.

The following dotplot of the yields will provide us with a simple method of assessing uncontrolled variation.

The assessment of uncontrolled variation is based on measuring spread.

```
                                 .
                    .  .         .   .  .  .   .  .  .              B

                                                 x
                   x   x  x  x      x  x   x  x  x                  A
---------------------------------------------------------------------
     |         |         |         |         |         |         |
    1.6       1.8       2.0       2.2       2.4       2.6       2.8
```

Now the cause of differences between the ten vines receiving A is uncontrolled variation (same for B). So given that the spread of the 10 observations in each treatment is due to uncontrolled differences, "Is there any evidence of a pruning contribution to yield spread *or* is the spread in all 20 likely to be uncontrolled variation only?". *Answer:* the latter i.e. difference between $\bar{y}_A$ and $\bar{y}_B$ can be attributed entirely to uncontrolled variation.

In fact, for A, $s = .221$
for B, $s = .219$
for both, $s = .215$

N.B.! The spread in the replicate observations provides a measure of uncontrolled variation.

Example IV.2 Pruning experiment in second vineyard

Suppose, the experiment is run in a second vineyard and the results are as follows:

Pruning Level	Yields									
A	**1.66**	**2.26**	1.80	2.06	2.00	1.74	2.14	2.26	2.20	1.86
B	2.06	**2.74**	2.26	2.46	2.40	**2.00**	2.54	2.66	2.60	2.26

Now, for this experiment, we again calculate the means to look at the question of interest.

$$\bar{y}_A = 1.998, \; \bar{y}_B = 2.398$$

Again, the question arises as to whether the difference is due to uncontrolled variation alone, or whether pruning treatments are contributing to yield differences also.

Lets look at the dotplot in this instance.

```
                                      .
                         .  .         .      .  .   .  .  .   .       B

                                      x
        x   x  x  x      x  x   x  x  x                               A
-------------------------------------------------------------------
     |         |         |         |         |         |         |
    1.6       1.8       2.0       2.2       2.4       2.6       2.8
```

Again the differences between 10 observations within A (or B) is due to uncontrolled variation. Thus, uncontrolled variation results in a range of 0.74 and 0.60 (or s of 0.221 and 0.250).

Can the total spread be attributed to uncontrolled variation or will the total spread be somewhat larger than the spread that we know is due to uncontrolled variation? Answer: The latter. So conclude that it appears that pruning treatments have lead to increased yield.

Summary

- Uncontrolled variation is inevitable in biological, physical and social sciences.
- Cannot use a single observation of each treatment in experiments.
- Replicate to measure (not eliminate) uncontrolled variation using a measure of spread.
- Compare total spread with spread from uncontrolled variation alone to decide if treatment has an effect.

Problem

While a measure of uncontrolled plot variability can be obtained the observed treatment difference might be caused by systematic effect.

Example IV.1 Pruning experiment (continued)

not so good part of field — better part of field

A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

⇒⇒⇒⇒⇒⇒⇒
fertility trend

Thus there will be large differences between the group of plots irrespective of whether or not prunings are different.

Any other systematic arrangements, such as those shown below, have this problem.

not so good part of field — better part of field

A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

⇒⇒⇒⇒⇒⇒⇒
fertility trend

A	A	A	A	A	A	A	A	A	A
x	x	x	x	x	x	x	x	x	x

B	B	B	B	B	B	B	B	B	B
x	x	x	x	x	x	x	x	x	x

b) Randomization

- overcomes systematic effects

Definition IV.3: **Randomization** is assignment of treatments to units so that every unit has the same probability of receiving each treatment. ■

Randomization results in an arrangement with no particular pattern. This is usually done by assigning each treatment with a number and then producing a random sequence of these numbers using a table of random numbers. The simplest statistical design involving randomization is the Completely Randomized Design (CRD).

Completely Randomized Design

Definition IV.4: A **completely randomized design** is one in which each treatment occurs a specified, possibly unequal, number of times. ■

Example IV.1 Pruning experiment (continued)

Let's randomize the treatments for our two-treatment, 20-vine example. In this example we could use the tossing of a coin to choose the treatment for each plot. So for the first plot, toss the coin and if it comes up head the plot receives A, otherwise B. This is repeated until A and B occur on 10 plots each. The result might be as follows:

A	B	A	A	B	B	B	A	A	B	A	A	A	A	B	B	A	B	B	B
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

The resulting arrangement is one with no particular pattern.

Note:

The situation here is similar to the arrangements given above in that we still have a measure of our uncontrolled variation and we can compare the total range with the range within each variety. The difference is that we are more confident that larger difference between A plots and B plots will be due to pruning rather than systematic effect.

c) Blocking

- to improve experiment.

Not absolutely necessary but very important in improving experiments.

Definition IV.5: **Blocking** is the *grouping* of all plots into groups called BLOCKS, the plots within a group being as similar as possible. ■

More about making them as similar as possible later.

Randomized Complete Block Design (RCBD)

Definition IV.6: A **randomized complete block design** is one in which the number of plots per block is equal to the number of treatments and every treatment occurs once and only once in each block, the order of treatments within a block being randomized. ■

Example IV.1 Pruning experiment (continued)

To produce this experiment, one must randomly select one plot in each block for each treatment.

Block

1		2		3		4		5		6		7		8		9		10	
B	A	B	A	B	A	B	A	A	B	B	A	A	B	A	B	B	A	A	B

The grouping of two neighbouring plots into a block is often used as it is likely that plots close together are more similar than plots further apart.

While blocks might be contiguous blocks of land, they might also be litters of animals, groups of people of the same weight or contiguous periods of time.

IV.B Experimental strategy

(Box, Hunter and Hunter ch. 9;
Mead and Curnow sec.14.8, 14.9)

What we have discussed so far assumes that we know what the treatments are to be. However, part of the process of designing an experiment involves deciding on the treatments to be included. I now discuss some aspects of making this decision.

The first point to note is that an experimental program very rarely consists of a single experiment. The process of scientific enquiry involves coming up with hunches and testing the hunches. Invariably, the hunches are not completely correct and so some new theory is developed, based on the old hunch and the results of the experiments conducted. A further experiment then has to be conducted to test the new theory. Thus scientific investigation is an iterative process leading to the evolution of new knowledge. So we are not having to worry about putting all our eggs into one basket and running the one definitive experiment which would necessitate getting the treatments spot on the first time.

A second point to consider in thinking about how the treatments are to be decided is that the types of experimental strategy can be categorized, according to the aim of the program, into three categories:

a) Selection: obtain the best r out of t treatments where t is large and there is no particular interest in any one treatment (e.g. a variety breeding program)

b) Comparison: a small set of treatments that differ qualitatively are to be compared to establish differences

c) Optimization: find the optimum level out of a range of treatments

a) Selection

The treatments (varieties) here are specified by the researcher (the plant breeder). As I have suggested there will be a large number of treatments and usually only two or three replicates of each. This generates specific problems of design and analysis that are currently undergoing active development. The subject will not be further discussed in this course, except to say that the selection is usually carried on over several years experiments with poor treatments being excluded at each stage.

b) Comparison

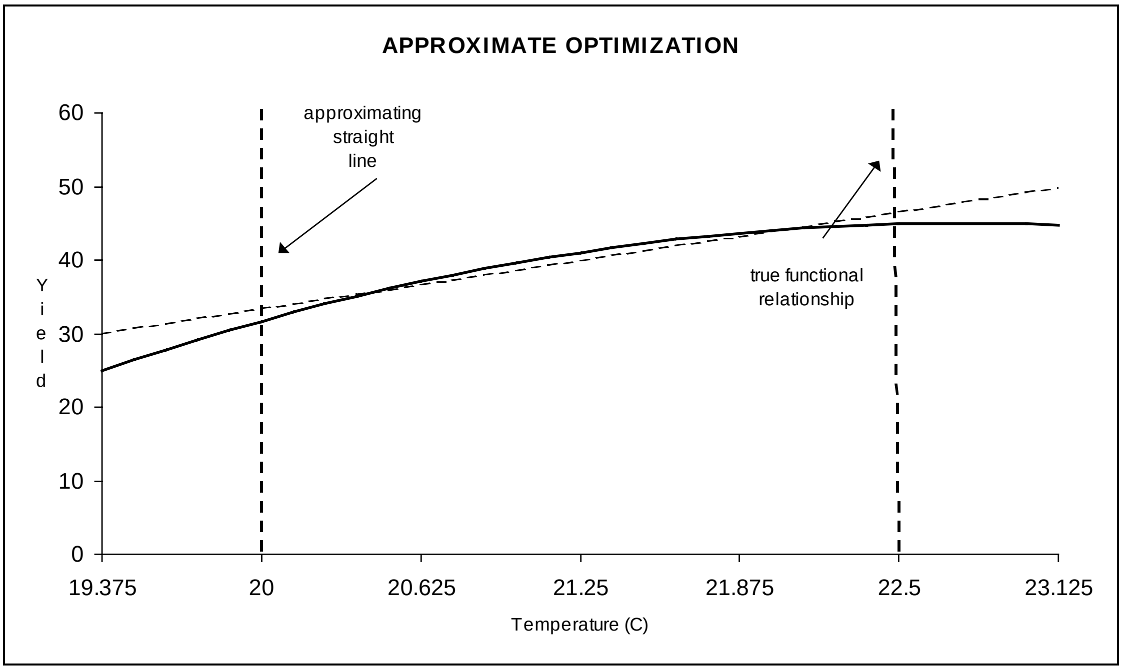
In this sort of experiment the researcher will have specified a particular set of treatments of interest. For example, it might be that a set of recommended varieties is to be tested at a particular location to determine the most suitable varieties for that location. It will often be that a control treatment will be included. This establishes the baseline for comparison. For example, the variety currently grown in the district is included. You know how it performs generally and you will have its performance under the conditions of the experiment to which you can compare the new varieties.

It is also a good idea to bear in mind that it might be beneficial to incorporate other factors to maximize the extent of the conclusions that will be able to be drawn from the experiment. Thus, you might include fertilizer factors, rather than use just the standard fertilizer application. It may be that some of the new varieties perform better under different fertilizer regimes to the standard variety.

Further, it is likely that several years experiments will be conducted to observe the performance of the varieties under different conditions (dry year, wet year etc.). The results of the trials in one year may cause some modification of the trial in subsequent years. Some obviously unsuitable varieties dropped, new varieties added, fertilizer application varied and so on.

c) Optimization

Here one is concerned with finding the optimum levels of a continuously variable, quantitative factor. This involves establishing the response function over the region of interest. To do this one often uses empirical functions to approximate the true relationship because we often do not know the true relationship and the empirical function provides a practically useful approximation. The situation is illustrated diagrammatically below.



At times straight lines will not provide a sufficiently good approximation and we will use a quadratic to provide a curved approximation. It is also possible to fit other curves, but we will not cover these in this course. The important point is that there is no intention that the equation is the true function expressing the relationship; they provide a convenient approximation.

The basic problem of experimental design is to decide what levels of the factor are best going to reveal the relationship between the response variable and the factors of interest over the region of interest.

To some extent there is a circular problem here in that if we knew the response function we would know which levels were best. For example, if one knew that the response function is linear then the best thing to do would be to have just two levels at either end of the region of interest and to replicate these heavily. If the response function was known to be quadratic, just three well-replicated levels spread evenly over the region of interest would be optimum as only three points are needed to specify a quadratic curve.

Fortunately, the fact that we do not know the response function is not crippling, particularly when experiments are run sequentially so that the information gained in one experiment can be used to influence the design of the next. The initial experiments will therefore be partly concerned with establishing the form of the function; that is, those factors that do not influence the response and the way those that do.

Thus, in the initial experiment to investigate a response function, the levels are likely to cover a wide range and be sufficient to allow the testing for deviations from the response function. Of course, the more levels one has the less replication one is able to employ for a given sized experiment. This can get to the stage where the replication is so paltry that one is unable to detect any difference. On the other hand, for widely spaced levels, the response differences are likely to be large so that some precision can be sacrificed. Generally, then one should include sufficient levels to establish the response but not so many so as to sacrifice precision. In subsequent experiments, as the range of interest narrows, so should the number of levels and there should be a concomitant increase in replication.

IV.C Summary

In this chapter we have:

- discussed the necessity of using replication and randomization in an experiment and the desirability of using blocking;
- described the two most basic experimental designs: completely randomized design and randomized complete block design;
- outlined considerations in selecting the set of treatments to be included in an experiment; in particular this depends on whether selection, comparison or optimization is the goal of the experiment.