STATISTICAL MODELLING

II. Designing experiments — some general aspects

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

II.A General considerations

Experiments, as opposed to surveys, involve researchers manipulating situations, by applying treatments, in an effort to draw conclusions about what their manipulations have caused. Surveys merely observe some aspect of the world as it is. Experiments represent a very important technique in the acquisition of scientific knowledge.

a) Basic purposes of experimentation

- To provide valid comparison of the effects of treatments
- To provide valid information about the relationship between variables of interest

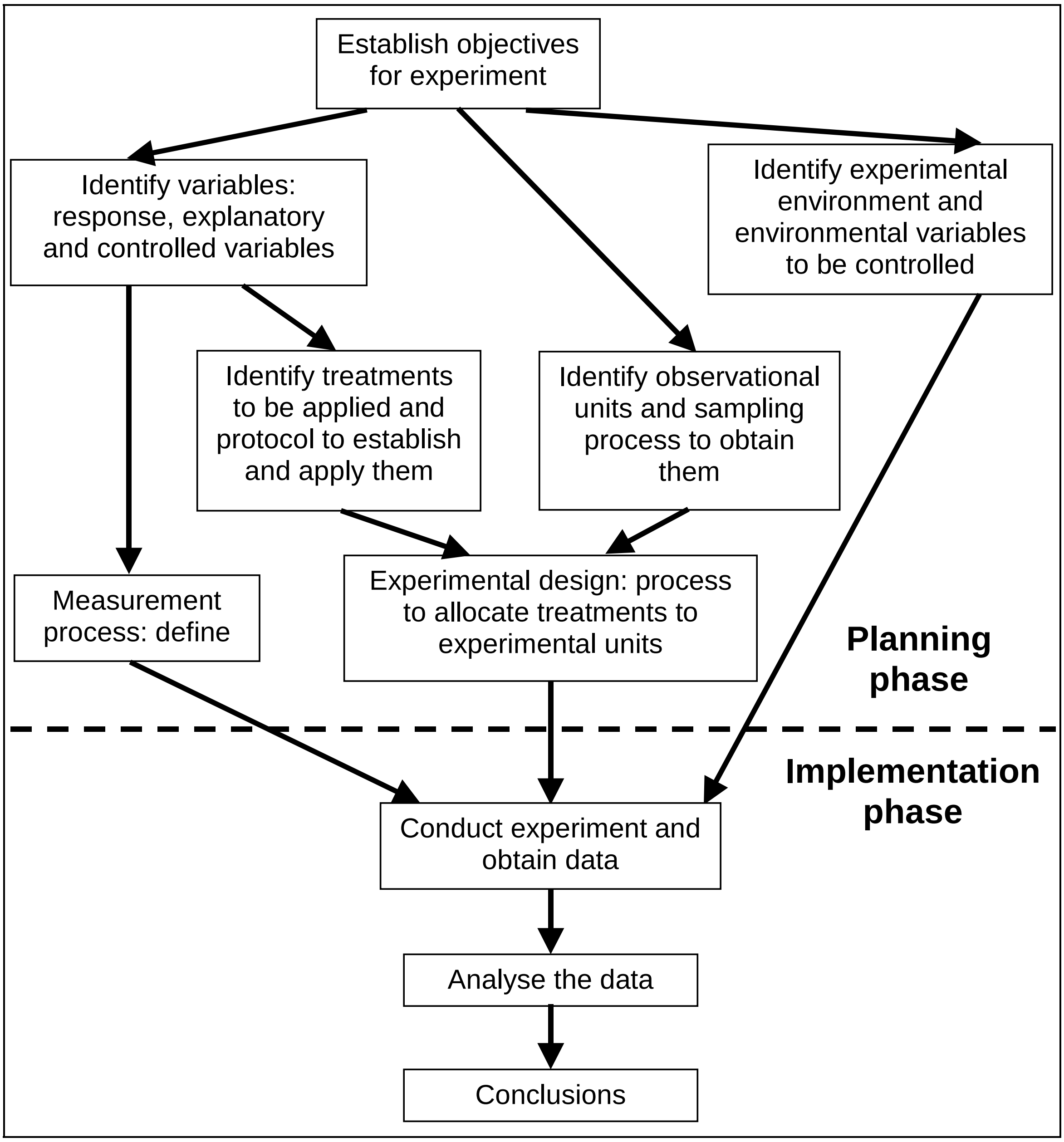
b) Basic requirements

- The experimental conditions should represent the situational conditions of the problem of interest
- The comparison of treatments should be free from other possible explanations due to the presence of other variables (confounding)
- The treatment comparison should be made with as little influence of random variation as possible

- The level of uncertainty in the conclusions should be assessable
- The experiment should be as simple as we can make it

c) The experimental process

Below are the key steps in conducting an effective experiment and their relationship is illustrated in the diagram below. Those above the dashed line make up the **planning phase** of the experiment and those below the **implementation phase**.



Example II.1 Factorial experimental design for concrete durability research

- A clear understanding of the problem being studied and the objectives for the experiment

- Identification of the variables of interest including the response (dependent) variable, the explanatory (independent) variable (or factor) to be studied and the other variables to be controlled.
- Identification of the observational units to be used and the sampling that is to be done to obtain them taking into account the need to control variation in them.
- Identification of the treatments to be applied and the protocol for establishing and applying them.
- Allocating the treatments to the experimental units through an experimental design, using blocking to overcome any major variation in the experimental units.
- Identification of the measurement process to be used and controlling the key sources of variation that might affect the measurement process.
- Identification of the experimental environment and how this is to be controlled.
- Carrying out the experiment and collecting the data, implementing any control protocols recognized as necessary for this step and paying attention to any unusual aspects of this step.
- Analysing the data
- Drawing conclusions and acting on them in relation to the original problem.

(*Proceedings of the Institution of Civil Engineers Structures and Buildings*, 1994, **104**, Nov., 449–462.)

Some key aspects of the experimental program were

- **Objective**: to investigate the effects of three factors (water-cement ratio, aggregate-cement ratio and aggregate size) on the durability of the concrete as measured by permeation properties (sorptivity, air permeability) and the freeze — thaw deterioration and the depth of carbonation. The permeation is considered an important part of durability due to the common failure mechanism of the penetration and spread of aggressive substances into the concrete.
- **Response variables**: permeation properties (sorptivity, air permeability) and the freeze — thaw deterioration and the depth of carbonation.
- **Explanatory variables (factors)**: There are three factors in this experiment and each factor was tested at three different levels or values. These factors and their levels were: Water-cement ratio (0.4, 0.55, 0.7), Aggregate-cement ratio (3.16, 4.65, 6.14) and Size of aggregate (mm) (6, 10, 20).
- **Treatments**: All possible combinations of the levels of the factors leading to 27 mixtures made up in total.
- **Observational/Experimental units**: run to produce a 300 mm x 600 mm x 50 mm slab — all slabs were made up from one batch of materials, materials were oven dried to avoid variability in moisture content, same person used to perform run, same degree of compaction achieved through vibration for 15 seconds, may have depended on mix combinations hence affecting results (in this experiment, and many others, the observational and experimental units are the same)
- **Experimental design**: There was only 27 runs made, no runs were repeated and the mixtures were assigned to the runs in a random order.
- **Measurement process**: Permeation testing, after 7 months using Autoclam permeability system with test samples dried to remove variation due to different

moisture contents; accelerated carbonation tests and freeze thaw tests were performed under controlled environmental conditions.

- **Analysis**: by graphical analysis and analysis of variance.
- **Conclusions**: A number of conclusions are drawn about the effect of the variables on durability. It was found that the effect of the factors studied on permeation gave very similar conclusions to the tests of durability suggesting that permeation is a good indicator of durability.

d) Definition of some key terms

We will refer to the example given above of concrete durability to explain some of the key terms:

Factor: the explanatory variable(s) manipulated or set by the experimenter. In the above example there were three factors: the water-cement ratio, the aggregate-cement ratio and the size of the aggregate.

Levels of a Factor: the values that an individual factor takes. In the example the factor water-cement ratio was set at three different levels: 0.4, 0.55, 0.7 — the other two factors had three levels too.

Treatment: a combination of one of the levels from each of the factors and this combination is applied to particular experimental units. In the above experiment the combination of a particular value for the water-cement ratio e.g. 0.4 with a particular value of the aggregate-cement ratio e.g. 3.16 and a particular value for the size of the aggregate, say 6 mm, would constitute one treatment. A run would be performed to make up a slab from this treatment. In the experiment each of the 3 levels of the three factors was tried in combination with each level of the other factors leading to a total of 27 treatments.

Observational unit: the native physical entity that yields a single value of the response variable. In our case it is a run since it produces a single value of each of the response variables — for example, sorptivity.

Experimental unit: the unit, which may be a collection of observational units, to which a single treatment is randomly allocated. The experimenter usually applies one or more conditions or treatments to a set of experimental units. In our case the experimental unit is a run to produce a single slab of concrete since one of the treatments is assigned to each run.

As mentioned above, the observational unit and experimental unit for an experiment are often the same.

II.B Three key principles in designing experiments

The three key statistical 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 a pilot plant for making paper for paper grocery bags. I make two specimens, keeping everything, as far as is humanly possible, exactly the same. The tensile strength of the paper for two runs:

12 psi and 15 psi.

This difference is likely to have been caused by a large number of small *uncontrollable* differences, viz. slight differences in

1. environment — ambient temperature
2. raw materials — slight differences in composition and amounts
3. settings on and behaviour of pilot plant

Definition II.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 II.2 Paper bag experiment

The reason uncontrolled variation is a problem can be seen by considering the following scenario. Suppose tensile strength of paper produced from the process is currently 15 psi and in control. However, management would like to improve the strength of the paper. The engineers think that hardwood concentration of pulp will affect tensile strength. Currently 10% hardwood is used in the process and economics dictates that hardwood concentrations is in the range 5–20%. It is decided to investigate whether there is a difference between the strength of paper when 10% compared to 15% hardwood is used. A specimen at of paper is made using the pilot plant at 10% hardwood and then a second specimen made at 15% hardwood. The only conscious difference between the productions of the two specimens is the hardwood concentration. Everything else is kept as similar as possible. The strength of the two specimens is determined and suppose that they are:

12 psi and 18 psi.

Does this different in yield indicate that the hardwood concentrations lead to strength differences? YES/NO.

Answer: Don't know as the strength difference may be due entirely to uncontrolled differences between the runs or may be due to a combination of uncontrolled differences and hardwood concentration differences?

That is, the difference between the concentrations is *mixed up* with run (uncontrolled) differences.

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

In this case, the Concentration effect is confounded with run 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, i.e. to several experimental units.

Example II.2 Paper bag experiment (continued)

So employing the first principle, each concentration is replicated six times as shown in the following table. That is, the 6 specimens for 10% hardwood are produced first, followed by the 6 for 15% concentration.

Run	1	2	3	4	5	6	7	8	9	10	11	12
Concentration	10	10	10	10	10	10	15	15	15	15	15	15

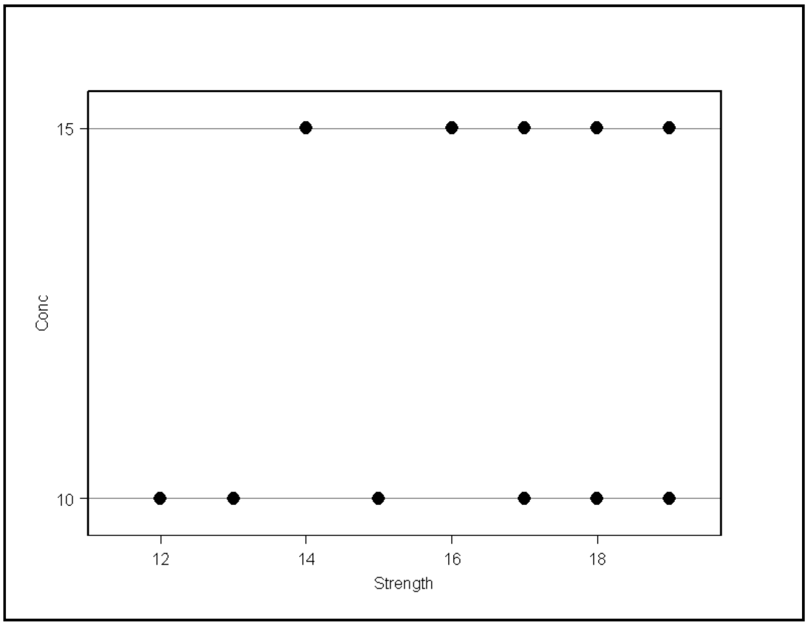
The measured strengths are given in the following table:

Conc			Strength (psi)				Range
10%	12	17	13	18	19	15	7
15%	14	18	19	17	16	18	5

Does concentration increase or decrease the strength? What statistics are we going to use to measure this? *Answer:* mean or median.

$$\bar{y}_{10}=15.7\ ,\ \bar{y}_{15}=17.0$$

Clearly, mean for 15% hardwood is larger than that for 10% hardwood. But, is the difference is due to uncontrolled variation alone, or are concentrations contributing to strength differences also? Cannot tell from just looking at the means whether concentrations are having an influence. Need to get a measure of uncontrolled variation. The following dotplot of the strength will provide us with a simple method of assessing uncontrolled variation.



Now the cause of differences between the six runs for 10% concentration is uncontrolled variation (same for 15%). So given that the spread of the six observations in each treatment is due to uncontrolled differences, "Is there any evidence of a concentration contribution to strength spread *or* is the spread in all 12 likely to be uncontrolled variation only?". *Answer:* the latter i.e. difference between $\bar{y}_{10}$ and $\bar{y}_{15}$ can be attributed entirely to uncontrolled variation.

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

Example II.3 Second paper bag experiment

Suppose a second experiment is conducted to investigate the difference between 10% and 20% hardwood concentration. In this experiment six specimens are made with 10% hardwood concentration and then six further specimens made with 20% hardwood concentration. The results are shown in the following table.

Run	1	2	3	4	5	6	7	8	9	10	11	12
Concentration	10	10	10	10	10	10	20	20	20	20	20	20

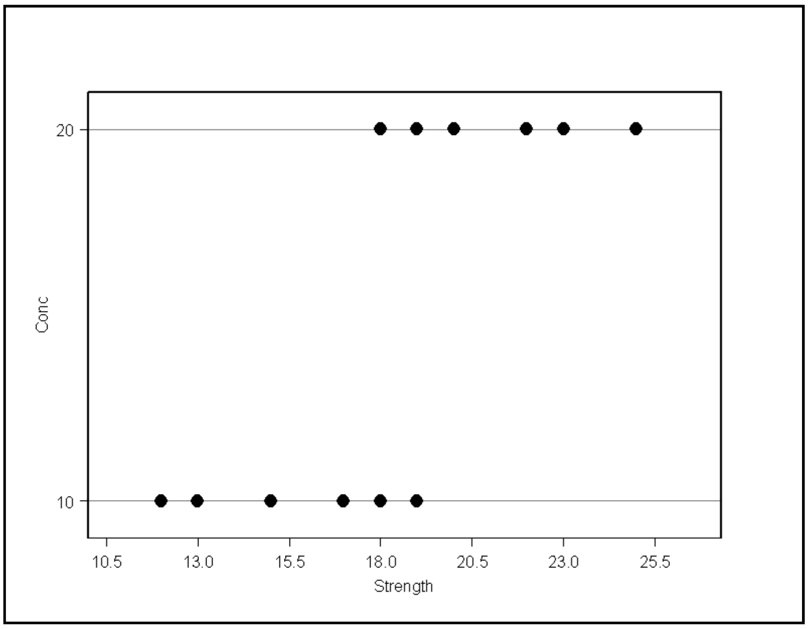
Conc			Strength (psi)				Range
10%	12	17	13	18	19	15	7
20%	19	25	22	23	18	20	7

The question of interest is once more: does concentration increase or decrease the strength? Now, for this experiment, we again calculate the means to look at the question of interest.

$$\bar{y}_{10}=15.7\ ,\ \bar{y}_{20}=21.2$$

Again, is the difference is due to uncontrolled variation alone, or are concentrations contributing to strength differences also?

Lets look at the dotplot in this instance.



Again the differences between six observations within 10% (or 15%) is due to uncontrolled variation. Thus, uncontrolled variation results in a range of 7 psi in both cases.

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 the 20% hardwood concentration has lead to increased strength.

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 variability can be obtained the observed treatment difference might be caused by systematic effect.

Example II.2 Paper bag experiment (continued)

For example

	low temperature											high temperature
Run	1	2	3	4	5	6	7	8	9	10	11	12
Concentration	10	10	10	10	10	10	20	20	20	20	20	20

⇒⇒⇒⇒⇒⇒⇒
temperature trend

Thus there will be a large difference between the two groups of runs irrespective of whether concentrations are different.

Other systematic arrangements

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

	low temperature											high temperature
Run	1	2	3	4	5	6	7	8	9	10	11	12
Concentration	10	20	10	20	10	20	10	20	10	20	10	20

⇒⇒⇒⇒⇒⇒⇒
temperature trend

	low temperature											high temperature
Run	1	2	3	4	5	6	7	8	9	10	11	12
Concentration	10	10	10	20	20	20	20	20	10	10	10	10

⇒⇒⇒⇒⇒⇒⇒
temperature trend

What if trend is to increase and then decrease?

b) Randomization

- overcomes systematic effects

Definition II.3: **Randomization** is assignment of treatments to experimental 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 II.4: A **completely randomized design** is one in which each treatment occurs a specified, possibly unequal, number of times. ■

Example II.2 Paper bag experiment (continued)

Let's randomize the treatments for our two-treatment, 12-run example. In this example we could use the tossing of a coin to choose the treatment for each run. So for the first run, toss the coin and if it comes up heads the run receives 10%, otherwise 15%. This is repeated until 10% and 15% occur on six runs each. The result might be as follows:

Run	1	2	3	4	5	6	7	8	9	10	11	12
Concentration	10	20	10	10	20	20	20	10	10	20	10	20

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 a larger difference between a run at 10% and those at 20% will be due to hardwood concentration rather than systematic effect.

c) Blocking

- to improve the experiment by reducing the amount of variability affecting the treatments.

Not absolutely necessary but very important in improving experiments as it allows some control of uncontrolled variation.

Definition II.5: **Blocking** is the *grouping* of experimental units into groups called BLOCKS, the units within a group being as similar as possible. ■

Numerous situations in which it can be applied:

- Machines and test equipment
- Batches of raw material
- People
- Time periods

all commonly form blocks. The aim is to isolate differences between these so that they do not affect treatment differences.

Randomized Complete Block Design (RCBD)

Definition II.6: A **randomized complete block design** is one in which the number of experimental units 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 II.2 Paper bag experiment (continued)

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

Block

1		2		3		4		5		6	
20	10	20	10	20	10	20	10	10	20	20	10

The grouping of two neighbouring runs into a block is often used as it is likely that runs close together are more similar than runs further apart. Here we are thinking that it is likely that environmental conditions in the laboratory will be very similar for the two runs, the test equipment will have changed little, the raw materials will be come from the same place in the batch, the operator will be in a similar state for both runs and so on. Consequently two consecutive runs are likely to be more similar than two widely separated runs. That is blocks are made up of experimental units (runs) as similar as the experimenter can make them so that treatments can be compared with the minimum effect of uncontrolled variation thus minimizing the threat to internal validity.

II.C Choosing the factors and their levels

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

As outlined in section II.A, *General Considerations*, we have to choose the explanatory variables and determine the treatments. This amounts to determining a) the factors to be investigated, and b) the levels of each of the factors. In this section we discuss some aspects of deciding the treatments to include in an experiment.

While some experiments involve just a single treatment factor, others involve two or more factors. For example, the paper bag experiment discussed in the previous experiment involves the single factor Concentration, whereas the concrete durability experiment involved the three factors Water-cement ratio, Aggregate-cement ratio and Size of aggregate.

Experiments in which we are interested in the effects of two or more factors are often performed as **factorial experiments**. On the other hand **single-treatment-factor** experiments involve just a single treatment factor. For the first half of this course we will concentrate on the latter type of experiments.

a) Factorial experiments

In factorial experiments the treatments are all combinations of the levels of all the factors. Generally, the number of treatments is equal to the product of the numbers of levels of the factors in the experiment. Once you have determined the treatments, then they can be assigned to experimental units using either of the two experimental designs that we have discussed, i.e. the completely randomized or the randomized complete block designs.

Example II.4 Investigating the moulding properties of sand

Consider an experiment designed to investigate the effect of moisture content and temperature on the moulding properties of sand. Suppose we choose two levels of moisture content (e.g. dry and moist) and two levels of temperature (low and high). The response (or dependent) variable is some measure of the moulding ability of the sand. A factorial experiment here would generate 4 groups of measurements, a different group for each of the four combinations of the factors: dry/low, dry/high, moist/low, moist/high. This experiment would be referred to as a $2 \times 2 = 2^2$ factorial experiment. The four treatments might then be assigned to the experimental units using a randomized complete block design. How many units per block would there be in this experiment? ■

The advantages of factorial experiments are that they are more efficient than the traditional 'one-at-a-time' approach and they also work better when the factors interact with each other — for more details see chapter VII, *Factorial experiments*. We are not restricted to just 2 variables, nor to two levels for each variable — an arbitrary number of factors with possibly different numbers of levels for different factors is allowed. Obviously as both the number of levels and the number of factors rises the experiment can become very large. For example, our concrete durability experiment with three factors each at three levels involved $3 \times 3 \times 3 = 3^3 = 27$ treatments.

b) Choosing the levels of a factor

Control treatment

A control treatment is one that involves no treatment to the units or the treatment with the "standard" or "commonly-accepted" practice. It is usual that a control treatment be included to establish the baseline for comparison. That is, we want to be sure that the treatments are better than doing nothing or an improvement over current practice. For example, 10% hardwood concentration in the manufacture of paper bags is the current practice and so this treatment is included so that we can be sure whether or not any other concentration is a real improvement over what is being done now. In this case, the non-addition of hardwood is not a feasible treatment, as it is a requirement of the manufacturing process that some hardwood is input. Similarly, if a new measurement method is being investigated, it needs to be compared to the currently-used method.

In a single-factor experiment, the control treatment is just one level, the control level, of the treatment factor. In a factorial experiment one has to establish the control level for each factor. The control treatment is then the treatment that has the control level for every factor.

Determining the other levels of a factor

As far as determining the levels for a particular factor is concerned, it is useful to categorize experiments into three categories on the basis of their general objective:

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)

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

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

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 of experiments with poor treatments being excluded at each stage.

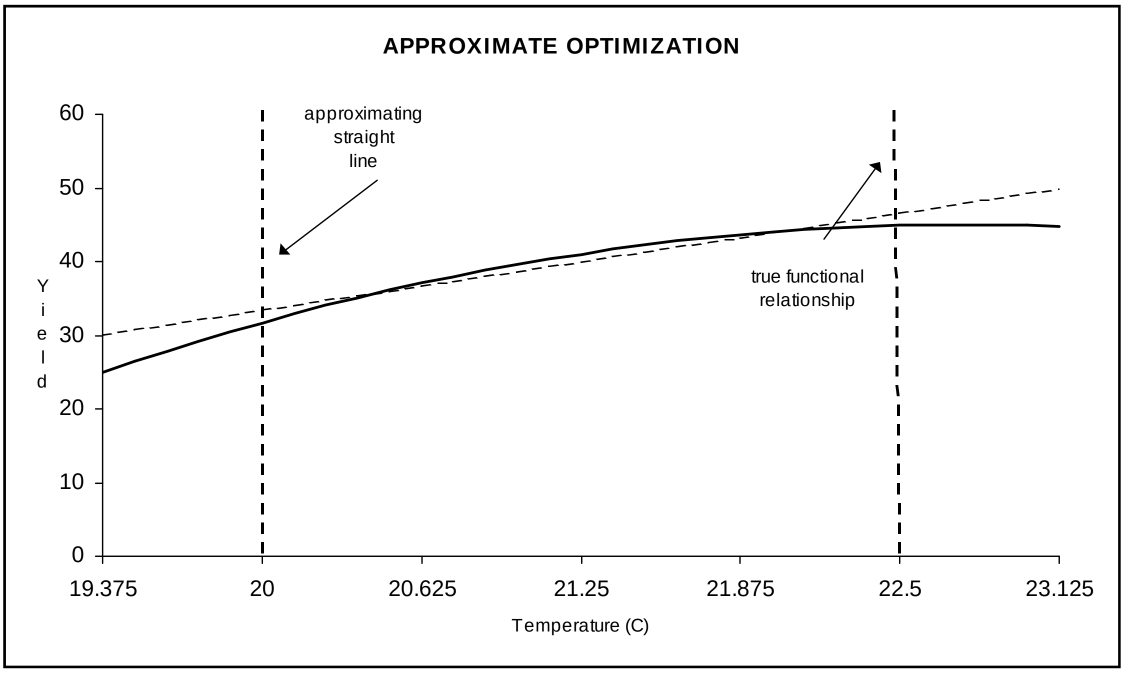
Comparison

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 materials that might be used in manufacturing a component to determine the most suitable material for that component. Do not forget to include the control treatment. This establishes the baseline for comparison. For example, the material currently used 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 materials.

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 operating temperature and other environmental factors, rather than use just the standard operating conditions. It may be that some of the new materials perform better under a wider range of operating conditions than the standard material.

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 main principles are:

1. Establish the range of the quantitative factor over which the response is to be established.
2. The levels of each factor should be chosen so that they cover the maximum range

3. The number of levels should be at least one more than the maximum number of parameters in models that are under consideration.
4. The levels should be equally replicated.

These principles take into account that:

1. Large differences in a factor are likely to be associated with large differences in the response variable and larger differences are relatively easier to detect than smaller differences. On the other hand a simpler model is likely to apply over a narrow range of a factor.
2. Testing for model failure is desirable and requires at least one more level than the number of parameters in the model. In practice, it will be rare that a model for a factor will have more than three parameters and so about 4 levels will all that should be used on most occasions.
3. You need sufficient replication to provide enough power for detecting differences. For a fixed number of runs in an experiment, the more levels of a factor that you have, the less replicates of each level there will be.

Thus, in an 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. The response differences are likely to be large so that some precision can be sacrificed. This can get to the stage where the replication is so paltry that one is unable to detect any difference. 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.

c) Sequences of experiments

A 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.

II.D Experimental validity

The assessment of the quality of an experimental design requires knowledge of the factors that influence or cause variation in the measured outcomes. With the paper planes for example, we were measuring flight time. In order to evaluate the experimental design we used we need to understand what factors affect flight time and then what we can do about these factors in the design of the experiment. Two concepts help us with this.

Internal validity: conclusions can be appropriately drawn from within this experiment about the relationship between the independent and dependent variable
External validity: conclusions from the experiment can be appropriately generalised to a wider situation of interest.

a) Internal validity

There are two main threats to the internal validity of an experiment. These are the effect of uncontrolled factors on our conclusions through random variation and the effect of variables that change systematically at the same time as we make systematic changes to the variables in the experiment.

Uncontrolled factors: In the example with the paper bags we might identify that the ambient temperature and variations in the amounts of chemicals used are factors that influence the strength. If these conditions are uncontrolled throughout the experiment and they vary, then this will cause variation in the strengths. This variation may obscure differences caused by the hardwood concentration or it may cause us to think that there is a difference when there is none. Variation due to the effect of uncontrolled factors is then a potential threat to the internal validity of this (and every other) experiment.

Systematic changes: If we are making comparisons between two treatments, e.g. two hardwood concentrations, then we want to avoid other possible explanations for a difference. If we introduce other possible explanations for our results then we have a problem of confounding. In the paper bag experiment we first suggested that we make all those with the lower concentration first and then all of those with the higher concentration next. If we observe a difference between the two concentrations we need to make sure that there is no other variable that also changed as we changed from one concentration to the other. For example, suppose that we had made all those with the lower concentration first and then had a break and came back to do those with the higher concentration next. Whilst we were away someone turned on the air-conditioning in the lab causing the lab to warm up. We now have another possible explanation for the difference in the strength between paper bags made using the different concentrations. The effects of the two variables concentration and temperature are said to be confounded, to be mixed up in a way that we cannot separate.

b) External validity

The issues concerning external validity relate to which variables we have controlled and what affect this will have in the generalisation of our results. For example to control the effect of variation due to different paper-plant plants we used one plant to produce the bags. We cannot be sure that the plant will not have an effect on their strength. We are therefore limited to the extent to which we can generalise from this experiment to say something about the best concentration for different plants.

c) Key methods for overcoming internal and external validity problems

Methods for ensuring internal validity

- Replication: for measuring the variability due to the uncontrolled factors so that we can properly assess the relationship between the explanatory and response variables (sufficient sample size is needed to overcome the effects of variation i.e. have sufficient power).
- Controlling variables and blocking: for minimizing the effects of random variation (eg. control the temperature, or the thrower of the planes, or the skill of the measurer by training).
- Randomization: for avoiding systematic confounding
- Methods for enabling generalisation to conditions beyond the experiment

Methods for ensuring external validity

- Sampling appropriately from the real environment with respect to experimental material and conditions.
- Replicating a broad range of conditions: enabling generalisation to a broader set of circumstances (This is also an added advantage of using blocking as described above).

II.E An example experiment

You are involved in the design of paper planes to sell to students. You have two designs that you are interested in and two potential materials that you wish to test. How should we design an experiment to identify the best design and the best material? Go through the planning phase in groups — see diagram. After coming up with a design, we will run the experiment and analyze it.

	Paper	Cardboard
Design1		
Design 2		

II.F Summary

In this chapter we have:

- outlined the experimental process;
- 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;
- discussed the importance of internal and external validity for an experiment.