Vocational Aqualabs

TRAINING

Vocational Aqualabs – Vocational Generic Skills For Researchers

Experimental Design

Unit 3 – Fundamentals of designing experiments

Trevor Telfer/James Bron
Institute of Aquaculture
University of Stirling

Experimental design

A good experiment requires **Randomisation**, **Replication** and **Controls**.

This follows from the enormous variability of biological material and systems and from the conscious or unconscious **bias** (we shall come onto this later) of the experimenter. Any given **treatment** must be applied to several (**replicate**) experimental units and must be **randomly** allocated to the experimental units. The probability of any given unit receiving any given treatment must be equal. In order to detect changes attributable to treatments, suitable **controls** must be employed.

Experimental variables

In any experiment there are three kinds of variables (**independent**, **dependent** and **extraneous**). Where:

- **Independent variables** are manipulated or chosen by the experimenter (e.g. water flow rate, treatment concentrations, times intervals for growth measurement etc).
- **Dependent variables** are those responding to the independent variables i.e. those measured by the experimenter (e.g. growth rate of fish over time, treatment response etc).
- **Extraneous variables** are those unselected variables affecting the response (e.g. amount of sunlight each day). Extraneous variables that are systematically related to the independent variable (e.g. nitrate concentration of the water flowing into the pond / tanks etc) are termed "**confounding**" variables.

Randomisation

Randomisation is an important factor in experimental design as it minimises the chance of a biased result through uncontrollable variability in the experimental equipment or materials. However, randomisation of an experiment is not always as straight forward as it may appear. Suppose we have two treatments, "Treatment A" and "Treatment B" and we have 20 tanks of fish divided into two groups of 10. The following gives two methods or randomising these for use in the two experimental treatments

Method 1

This is a simple method which uses counters or objects which differentiate treatments. For randomising the two treatment experiment over 20 tanks you will have an experimental set up as shown in Figure 1. Label the tanks in order from 1 to 20.

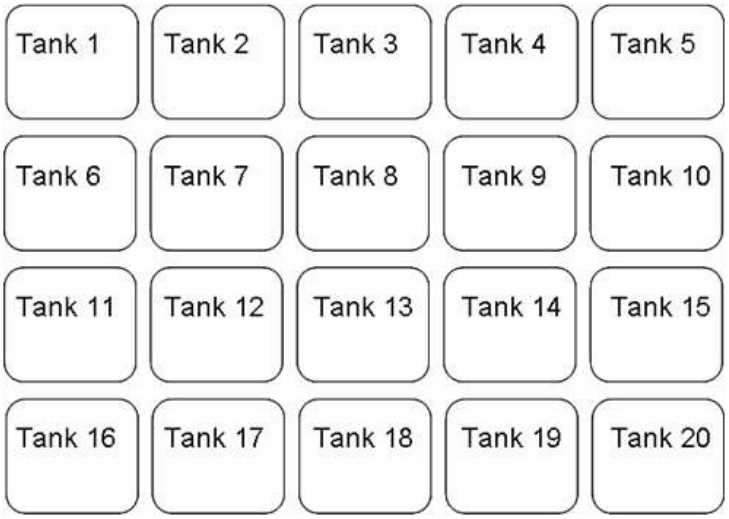


Figure 1: Twenty tanks waiting to have their treatments randomly allocated to them.

Take 10 counters of one colour, say blue (these will represent "Treatment A"), and 10 counters of another colour, say yellow (these will represent "Treatment B"). Now place all 20 counters in a container and mix them up. Without looking remove one counter from the container, this will be the treatment for Tank 1, now remove a second counter, this will be the treatment for Tank 2. Keep on repeating this until all the counters have been used up. This will give a treatment distribution over the tanks something like that shown in Figure 2.

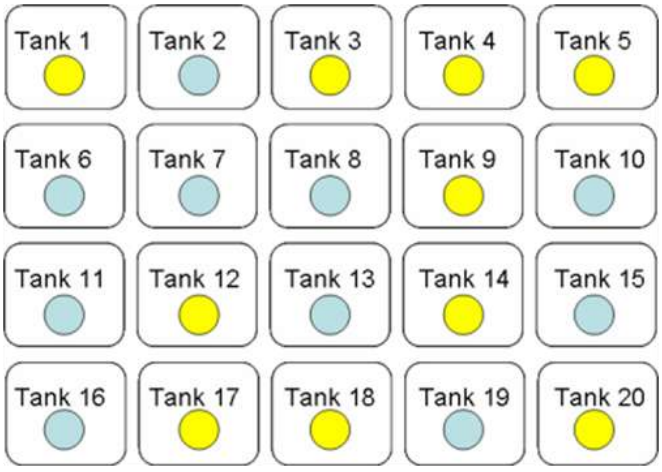


Figure 2: Random allocation of the two treatments "A" (blue) and "B" (yellow) to the twenty experimental tanks.

Method 2

Though Method 1 is a very simple and quick method of randomising treatments within an experiment it is more usual to use electronically generated random numbers. This can be done using computer software (such as spreadsheets) or a calculator. An example is to again randomly allocate the two treatments to the twenty tanks but this time we are going to use a random number generator on a calculator. The calculator produced the following numbers:

10 / 27 / 53 / 96 / 23 / 71 / 50 / 54 / 36 / 23 / 54 / 51 / 50 / 14 / 28 / 02 / 12 / 29

We use these numbers to allocate the treatment to the tank but if the number is larger than the tank number, then we substrate the number of tanks from the number until we can allocate it to a tank. Starting with "Treatment A" (blue dots) the series of numbers can be used as follows:

10	**Assign to Tank 10**
27	**Too large, so 27 – 20 = 7 (Tank 7)**
53	**Too large, so 53 – 20 – 20 = 13 (Tank 13)**
96	**Too large, so 96 – 20 – 20 – 20 – 20 = 16 (Tank 16)**
23	**Too large, so 23 – 20 = 3 (Tank 3)**
71	**Too large, so 71 – 20 – 20 – 20 = 11 (Tank 11)**
50	**Too large, so 50 – 20 – 20 = 10 (Tank 10 already used)**
54	**Too large, so 54 – 20 – 20 = 14 (Tank 14)**
36	**Too large, so 36 – 20 = 16 (Tank 16 already used)**
23	**Number already appeared in the number series**
54	**Number already appeared in the number series**
51	**Too large, 51 – 20 – 20 = 11 (Tank 11 already used)**
50	**Number already used in the number series**
14	**Assign to Tank 14 but already used**
28	**Too large, so 28 – 20 = 8 (Tank 8)**
02	**Assign to Tank 2**
12	**Assign to Tank 12**
29	**Too large, so 29 – 20 = 9 (Tank 9)**

So the allocation of the treatments between the tanks using the random number generator should look like the following in Figure 3:

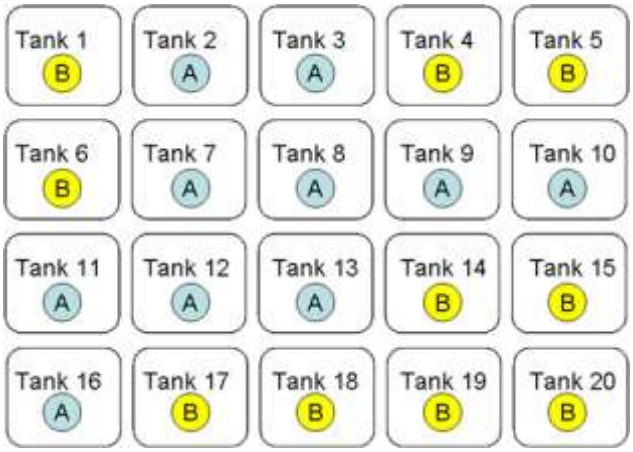


Figure 3: The random allocation of treatments to tanks using "Method 2 – random number generator".

Replicates

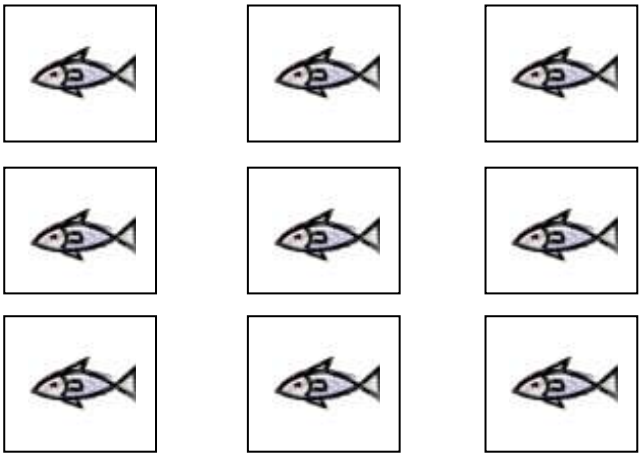
Replicates (individual experimental units subject to the same experimental variable/test condition) are essential to good experimental design especially given the enormous variability of biological material and the immense possibilities of unexpected and inexplicable changes in any experiment conducted in the real world-world. Replicates are even more critical for experiments carried out in the field (since fewer variables may be controlled).

Though the number of replicates required for a sufficiently rigorous design varies considerably, as a guide, it is suggested **at least** 5 replicate groups for any given treatment are used with a bare minimum of 3 replicates. Ideally though there should be as many replicates as possible, though this is often limited by feasibility and budgetary constraints. It is better to have more replicate groups and fewer individuals in each replicate group than to have the reverse i.e. fewer replicate groups with more individuals in each. For example, it is better to have 5 replicates of 100 fish in each replicate than to have 2 replicates with 250 fish in each replicate.

It is important that replication is consistently and completely independent throughout the experiment. A common error in experimental design is called "pseudo-replication". This is defined as "the use of inferential statistics to test for treatment effects with data from experiments where either treatments are not [truly] replicated (though samples may be) or replicates are not statistically independent" (Hurlbert, 1984). Hurlbert (1984) found that 48% of 176 studies reviewed exhibited pseudo-replication.

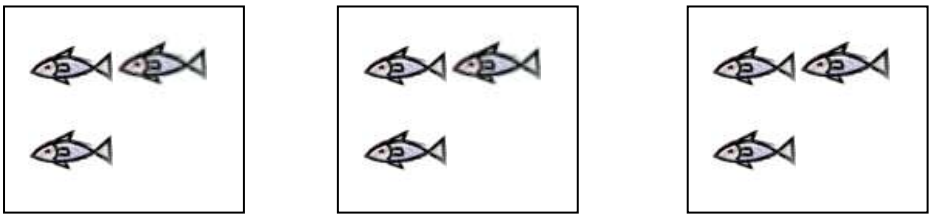
An example of a true replicate and pseudo-replicate design is given below. Consider a growth experiment for fish fed different diets A, B and C. Experiment 1 is truly replicated (3 replicate tanks per treatment each independently fed). Experiment 2 exhibits pseudo-replication with a single tank per treatment containing 3 fish, but not independently fed.

Experiment 1



Treat A **Treat B** **Treat C**

Experiment 2



Treat A **Treat B** **Treat C**

Experiment 2 has a potential problem with confounding variables. There is less potential for randomisation of the replicates (though not shown it would be easy to randomise the design for experiment 1). In experiment 2, imagine something happened to the tank for treatment A (feeder accidentally fed smaller ration), the fish in this tank may all become smaller than the others leading to the conclusion that this is the poorest feed formulation, though actually this results is chance. If a feeder malfunctioned within experiment 1, comparison between replicates would distinguish this fact.

Controls

Controls provide a baseline which allows assessment of changes due to experimental treatments. It is essential that your controls are treated identically in every respect to your test population save for the lack of the test condition itself.

For example: If you coat medicated feed pellets in gelatine containing a test drug, then you must also coat unmedicated pellets in gelatine only as a control (uncoated pellets do not provide a suitable control). Similarly for experiments such as clinical trials non-infected control fish should be subject to sham injection (i.e. imitate an actual injection) – sham injected fish should be injected with, for example, phosphate buffered saline or the carrier of the active ingredient – to simulate the same stress conditions as the test fish, which will be injected with the carrier of the active ingredient and the active ingredient.

Sample size

Sample size is another important feature when designing and experiment. The sample size determines:

- The ability of your experiment to resolve differences between population samples
- The likelihood that a sub-sample will accurately reflect the population it was taken from
- The statistical tests which may be employed to analyse your results, and the statistical power of the methods used.

As a "rule of thumb" a large (ideal) sample comprises 50 or more individuals although a sample of 30 is good. Only drop below 10 where it is unavoidable. Though sample size can be determined specifically as part of the experimental design.

Sample size determination is the act of choosing the number of observations to include in a statistical sample, within the design of an experiment. In practice, however, the sample size used in a study is often determined on the basis of the expense of data collection, and the need to have sufficient statistical power. In complicated studies there may be several different sample sizes involved in the study; for example, in a survey involving stratified sampling there would be different sample sizes for each population. In experimental design, where a study may be divided into different treatment groups, there may also be different sample sizes for each group. Sample sizes may be chosen in several different ways:

- Expedience - For example, include those items readily available or convenient to collect. A choice of small sample sizes, though sometimes necessary, can result in wide confidence intervals or risks of errors in statistical hypothesis testing.
- Using a target variance for an estimate to be derived from the sample eventually obtained
- Using a target for the power (see later) of a statistical test to be applied once the sample is collected.

Larger sample sizes generally lead to increased precision when estimating unknown parameters. For example, if we wish to know the proportion of a certain species of fish that is infected with a pathogen, generally a more accurate estimate of this proportion would be obtained if 200, rather than 100, fish were sampled or observed. Several fundamental facts of mathematical statistics describe this phenomenon, including a better precision to the mean (law of large numbers[1]) and the higher probability of normally distributed data (central limit theorem[2]). However, in some situations, the increase in accuracy for larger sample sizes is minimal, or even non-existent. This can result from the presence of systematic errors or strong dependence in the data, or if the data follow a heavy-tailed distribution.

Sample sizes are judged based on the quality of the resulting estimates. For example, if a proportion is being estimated it may be desirable to have the 95% confidence interval be less than a certain number of units wide. Alternatively, sample size may be assessed based on the power of a hypothesis test. For example, if comparing the preference of a group of fish for a particular formulated feed there may be a requirement to have 80% power to detect a preference of a certain number of measurement units.

Good experimental practice

Using the principles mentioned in the sections above, a checklist for good experimental practical can be developed. This uses a mixture of previously discussed concepts and common sense factors which decrease the likelihood bias (or preconceptions) within the results. The checklist should include:

- Select the sample group carefully to avoid confounding variables
- Randomly assign a treatment, see above
- Try and make studies single or double blind, where:
    - First – Experimenter does not know from which individuals get which treatment
    - Second – Analyst does not know from which group a sample derives
- Crossover design – for two treatments A and B some subjects are randomly assigned to treatment A initially and then move to treatment B and some *vice versa*.

[1] the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed

[2] where conditions under which the mean of a sufficiently large number of independent random variables, each with finite mean and variance, will be approximately normally distributed

Alternative Designs and considerations

Power and pre-sampling

Statistical or experimental power has been mentioned several times previously. What is it? The "power" of a given experimental design is its probability of allowing one to correctly reject a false null hypothesis or accept a valid hypothesis. This depends upon the sample size employed, the actual difference between the populations being studied, the variability of the populations and the level of statistical significance that has been chosen. In practice the easiest way to increase the power of an experimental design is to increase the sample size (see above).

It is possible to calculate the sample size required to detect a given difference between populations. The difference may be estimated by pre-sampling or by using past studies (from the literature) as a guide.

Randomised Block design

Let us think about a wholly randomised design of four treatments assigned randomly across 16 tanks (Figure 4). Again we can allocate treatments to tanks as we have already done above. Unfortunately, if you look at Figure 4, you will see that all the Treatment A's are "randomly" clustered together in the first column of tanks. While this does not matter in an environment where all the other variables having an effect on the tanks are equal, in the real world this does matter. For example, what if there was a gradient of light coming from the left, this would bias the treatments with A's replicates consistently receiving more light than any of the other treatments.

A	B	B	C
A	C	D	B
A	C	D	D
A	D	B	C

Figure 4: A wholly randomised design – four treatments allocated randomly between 16 tanks.

But what about the random allocation of tanks to blocks of tanks? Here, we can use a randomised complete block design (see Figure 5), each block receives one of each treatment randomly allocated. But, how could this block design be improved?

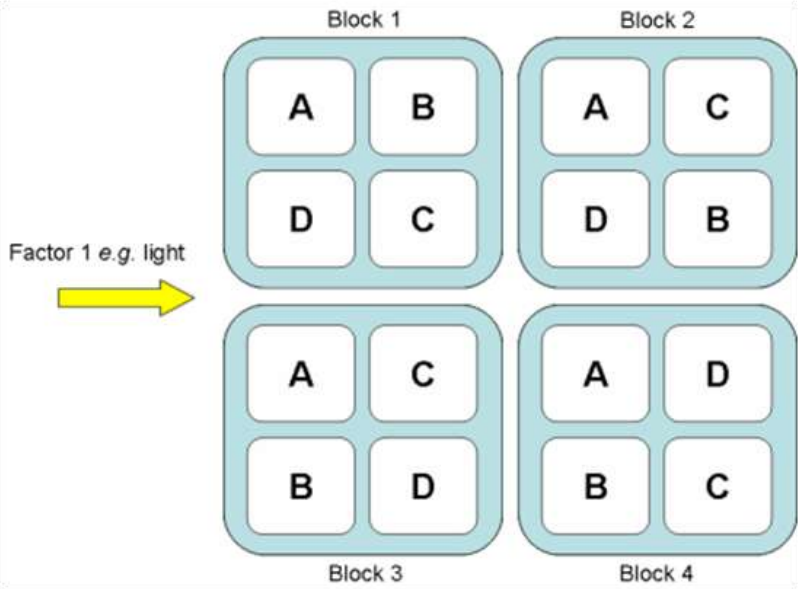


Figure 5: Randomised complete block design. Four treatments are randomly allocated between the four blocks of tanks.

Latin Square Design

As you can see from Figure 6, the random allocation of treatments between tanks in a block design can be improved. If you look at the Figure 6, you will see that each block has one treatment each BUT each treatment has also been randomly allocated in each row and in each column.

This means that each treatment in each column is affected equally by "Factor 1 – Light". Also each treatment in each row is affected equally by "Factor 2 – Wind direction". If you now look back at Figure 5, you can see that there are two tanks receiving Treatment A and any differences that do occur between Treatment A and Treatments B, C and D may also be due to the effects of "Factor 1 – Light" upon them. We do not know, so it is better to use a better design and eliminate this potential source of error in the design of our experiment.

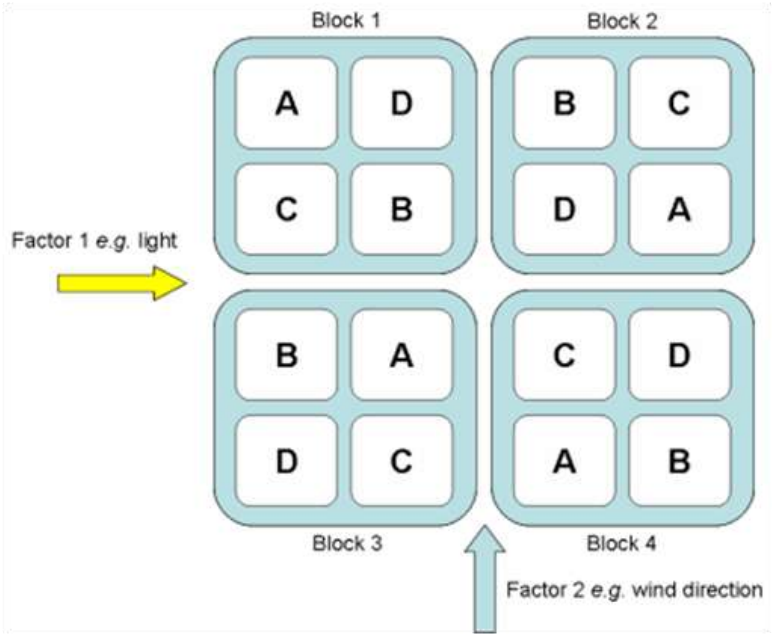


Figure 6: A Latin Square Design for the randomly allocation of treatments.

Similarly, in Figure 16, you can see that there are two tanks receiving Treatment B in the bottom row. These tanks could be affected by the influences of "Factor 2 – Wind direction". So again, it is better to use a proper random design to eliminate this potential source of confounding error. This is an important consideration that only in the allocation of treatments to tanks or ponds in the field but also e.g. the front or back of an incubator / fridge for Petri dishes (uneven circulation, lighting etc).

Factorial design

A factorial design is used to examine the effects and interaction of two or more factors. For example let us consider the example given in Table 1 – food composition where we are looking at the inclusion of different levels of lipid and protein.

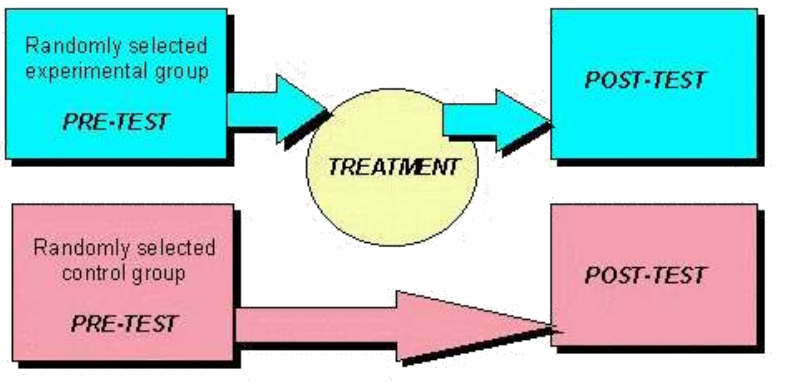
The example given in Table 1 is a "3 x 4 factorial" with 5 replicates. The advantages of this type of design are that you only need a single experiment and the hidden replication examines the interaction.

Table 1: A factorial design to investigate the inclusion of different levels of lipid and protein.

Protein	**Lipid**		
	Low	**Medium**	**High**
Low	5 reps	5 reps	5 reps
Medium	5 reps	5 reps	5 reps
High	5 reps	5 reps	5 reps
Very high	5 reps	5 reps	5 reps

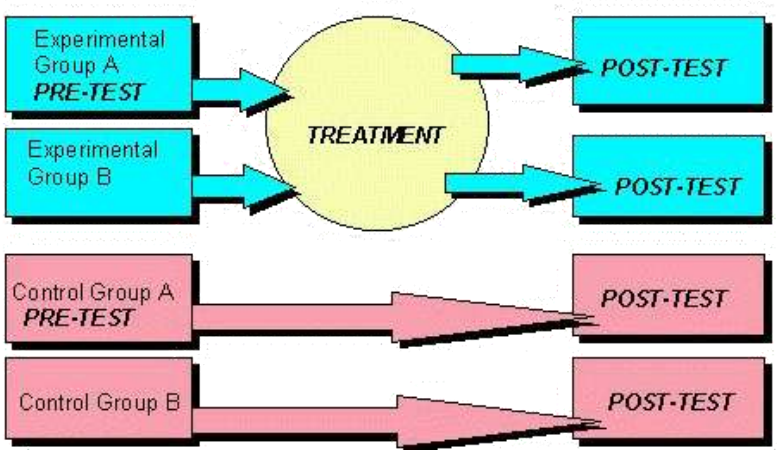
Pre-Post Randomised Group

This design uses a standard type randomised group design method but introduces a pre-test check on the degree of comparability of the control and experimental groups before the treatment is given. This design allows pre-treatment behaviour and the relationship to post-treatment behaviour of the treatments and controls to be compared and taken into account within the results highlighting threats to internal validity of the experiment. It does not however, allow for errors within the experimental procedure which could provide statistical errors in the results.



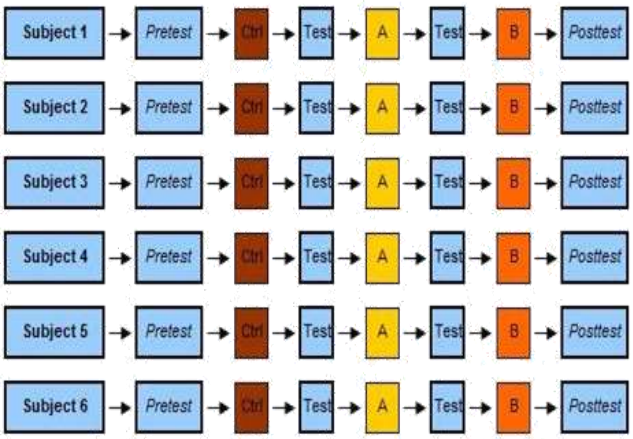
Solomon Four Group

This is similar to the pre-post randomised design but attempts to control for the possible sensitising effects of pre-test or measurements by adding two groups who have not been part o the pre-test or pre-measurement. This can be a complex design and is mostly used in behavioural and educational research.



One-shot Repeated Measures

This design is used to assess the effects of a treatment within the same group or the same individual over a period of time. A measure, or observation, is made more than once to assess the effects of the treatment. This design or variations on it must be used if experiments are assessing, for example, the chronic effect of treatment over time on the same group (e.g. growth, reproductive activity etc) if the measurements at each time point are being directly compared. It is often used if an experimental design needs to incorporate continued treatments one the same group of test organisms throughout the experiment.



Randomised Repeated Measures

This is a variant on the previous design. Here two more experimental methods are compared and repeatedly measured or observed.

A simple guide to the use of the above designs in medical and behavioural applications can be found in "The Basics of Experimental Design – A quick and non-technical guide" by Sytsma – included in the source materials.

.

Vocational Aqualabs

TRAINING

Many thanks for your attention today

Experimental Design
Unit 3 – Fundamentals of designing experiments

For further information
please contact:

Trevor Telfer/James Bron
Institute of Aquaculture
University of Stirling

AQUATT

UNIVERSITY OF ABERDEEN

UNIVERSITY OF STIRLING

INNOVAMAR

Aquark

Education and Culture DG
Lifelong Learning Programme

This project has been funded with support from the European Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained herein.