Design Issues

A *completely randomized design* is one in which each experimental unit (eg, person or animal) is assigned randomly to one of several competing treatments. For example, a study is proposed to compare 4 treatments (eg, a control and 3 distinct active treatments or a control and 3 different doses of the same treatment), and a sample of 20 animals are randomized to the 4 treatments. The randomization could be implemented by 1 of a number of possible techniques ranging from a simple randomization (in which a single sequence of the numbers 1 through 4 is produced, and animals are assigned according to the sequence) to a more involved randomization that uses stratification or permuted blocks. The permuted blocks strategy is used to ensure balance in the randomization process such that equal numbers of animals are assigned to each treatment. This strategy can be designed to ensure balance at specified enrollment points, for example, balance among the 4 treatments after randomization of 8 (2 per treatment) or 12 (3 per treatment) experimental units. This strategy is generally used when enrollment into a study occurs over time. With the permuted blocks strategy, 5 animals would be randomly assigned to each treatment in the present example.

The goal of the analysis is to compare responses among the 4 treatments. If the dependent or outcome variable is continuous, this test is performed with ANOVA. If the outcome variable is categorical, this test is performed with a χ^2 test. These tests are based on the assumption that the measurements within and across treatments are independent or unrelated. If the experimental units are unrelated (ie, not family members or littermates), and 1 measurement has been made per unit, then this assumption is reasonable.

In contrast, in a *repeated measures design*, multiple measurements are taken on each experimental unit. Consider again the application described above in which the goal of the analysis is to compare the 4 competing treatments. A repeated measures design could involve 5 animals, each measured 4 times, once under each experimental condition. The repeated measures design involves a smaller number of animals, which is both efficient and ethically appealing. If 5 animals are measured under each of 4 different experimental conditions, a total of 20 measurements will again be available for analysis. The 20 measurements, however, are not independent but are related within the subjects. Because the measurements might be affected by within subject characteristics (e.g., age or genetic factors), statistical tests that properly account for within-subject correlation are needed. If we assume that measurements taken in the same individual are correlated, the test for a difference in treatments will involve a smaller residual or error variance than that based on a completely randomized design, thereby increasing precision in the analysis.

A *randomized block design* is one in which a set of experimental units are organized into homogeneous groups or blocks on the basis of a characteristic assumed to affect the outcome. The goal is to have *r* replicates of each of *k* treatments in each of *b* blocks, with the total sample size

$n = kbr$. Consider again the study comparing 4 competing treatments ($k = 4$). Suppose the outcome of interest is known to be affected by age. With n = 20 independent experimental units, these might be organized into 5 age groups (e.g. quintiles of age) with 1 replication per group ($k = 4$, $b = 5$, $r = 1$). In a randomized block design, experimental units within each block are randomly assigned to treatments, and this technique reduces variation due to differences in age. The design can be thought of as replications of a completely randomized experiment in which there are as many replications as there are blocks.

Some repeated-measures designs can be viewed as a special case of the randomized block design in which the block is the individual experimental unit (e.g. person or animal). The randomized block design is often used with siblings or littermates. The family unit is the block, and assessments are repeated on each member of the family. The assessments within a family or litter are related. Accounting for the dependencies within the block results in a more precise test of treatment differences.

An experiment design in which experimental units (subjects) are used repeatedly by exposing them to a sequence of different or identical treatments is called a repeated measurements design (for brevity an RM design). Such designs are known by different names in the literature: crossover or changeover designs, (multiple) time series designs, or before-after designs in some special cases.

It may be of interest here to note that an extreme form of an RM design is the one in which the entire experiment is planned on one experimental unit.

An obvious disadvantage of such a design is that if the total time during which a given subject can be under observation is fixed, the number of treatments that can be compared may be severely limited.

The need for these designs can be justified in several ways.

- Due to limitation of the budget, the experimenter has to use each experimental unit for several tests.
- In some experiments the treatments' effects do not have a serious damaging effect on the experimental units and; therefore, these experimental units can be used for the second, third, etc. experiments.
- In some experiments, the experimental units are human beings or animals and often the nature of the experiment is such that it calls for special training over a long period of time. Therefore, due to time limitation, one is forced to use these experimental units for several tests.
- One of the objectives of the experiment is to find out the effect of different sequences as in drug, nutrition or learning experiments.
- Sometimes the experimental units are scarce, therefore the experimental units have to be used repeatedly.

RM designs have application in many branches of scientific inquiry such as: agriculture, animal husbandry, biology, education, food science, market research, medicine, pharmacology, psychology, and social engineering.

Summary

When possible and appropriate, use repeated measures designs. They reduce error-variance, thus increasing power, making it easier to see treatment effects. N can be quite high, even with few participants, thus increasing power. They require fewer participants, thus making recruitment easier, as well as reducing testing hassles. However, time-effects and order-effects must be carefully managed to eliminate confounding.

WHICH MODEL SHOULD WE USE?

The selection of a computational model should be based on our expectation about whether or not the studies share a common effect size and on our goals in performing the analysis.

Fixed effect

It makes sense to use the fixed-effect model if two conditions are met. First, we believe that all the studies included in the analysis are functionally identical. Second, our goal is to compute the common effect size for the identified population, and not to generalize to other populations. For example, suppose that a pharmaceutical company will use a thousand patients to compare a drug versus placebo. Because the staff can work with only 100 patients at a time, the company will run a series of ten trials with 100 patients in each. The studies are identical in the sense that any variables which can have an impact on the outcome are the same across the ten studies. Specifically, the studies draw patients from a common pool, using the same researchers, dose, measure, and so on (we assume that there is no concern about practice effects for the researchers, nor for the different starting times of the various cohorts). All the studies are expected to share a common effect and so the first condition is met. The goal of the analysis is to see if the drug works in the population from which the patients were drawn (and not to extrapolate to other populations), and so the second condition is met, as well. In this example the fixed-effect model is a plausible fit for the data and meets the goal of the researchers. It should be clear, however, that this situation is relatively rare. The vast majority of cases will more closely resemble those discussed immediately below.

Random effects

By contrast, when the researcher is accumulating data from a series of studies that had been performed by researchers operating independently, it would be unlikely that all the studies were functionally equivalent. Typically, the subjects or interventions in these studies would have differed in ways that would have impacted on Fixed-Effect Versus Random-Effects Models the results, and therefore we should not assume a common effect size. Therefore, in these cases the random-effects model is more easily justified than the fixed-effect model. Additionally, the goal of this analysis is usually to generalize to a range of scenarios. Therefore, if one did make the argument that all the studies used an identical, narrowly defined population, then it would not be possible to extrapolate from this population to others, and the utility of the analysis would be severely limited.

Difference between repeated measures design and a simple multivariate design

One should be clear about the difference between a repeated measures design and a simple multivariate design. For both, sample members are measured on several occasions, or trials, but in the repeated measures design, each trial represents the measurement of the same characteristic under a different condition. For example, one can use a repeated measures ANOVA to compare the number of oranges produced by an orange grove at years one, two, and three. The measurement is the number of oranges, and the condition that changes is the year. In contrast, for the multivariate design, each trial represents the measurement of a different characteristic. You should not, for example, use a repeated measures ANOVA to compare the number, weight, and price of oranges produced by a grove of orange trees. The three measurements are number, weight, and price, and these do not represent different conditions, but different qualities. It is generally inappropriate to test for mean differences between such disparate measurements.

Matched pairs design

If we cannot use a repeated measures design it is sometimes possible to match every subject in one group with a very similar person in the other group. In order to get the pairing precise enough, it is common to get one group of participants together and then look round for partners for everyone. Participants can be matched on variables which are considered to be relevant to the experiment in question. For example, pairs of participants might be matched for age, gender and their scores from intelligence or personality tests.

Although this design combines the key benefits of both an independent and repeated measures design, achieving matched pairs of participants is a difficult and time consuming task which may be too costly to undertake. Successful use of a matched pairs design is heavily dependent on the use of reliable and valid procedures for pre-testing participants to obtain matched the pairs.